Segmentation by color coalition labeling for figure-ground segregation in decoration designs

Mark J. Huiskes
Centre for Mathematics and Computer Science
Kruislaan 413, 1098SJ Amsterdam, The Netherlands
Mark.Huiskes@cwi.nl

Eric J. Pauwels
Centre for Mathematics and Computer Science
Eric.Pauwels@cwi.nl

Abstract

Accurate figure-ground segregation of decoration designs allows for the computation of high-level design features for use in content-based retrieval and indexing systems, and may further serve as a basis for design understanding and various design manipulations. This paper outlines a method for figure-ground segregation that is particularly suitable for color design images generally encountered in the textile industry. We propose a strategy consisting of three main steps: (i) obtain initial candidates for the background by multi-scale detection of color texture regions; (ii) assess the appropriateness of the candidates by a classification algorithm based on visual cues such as relative size, connectedness and massiveness; (iii) integrate the results of the previous steps to produce a hierarchical description of the figure-ground structure of the design. The first step is performed by a novel algorithm for color texture detection based on an analysis of the local color structure of the design. The multi-scale nature of the algorithm allows for the detection of nested backgrounds. Results are presented for two databases of decorative design images, one of specialized type and one of general type.

1 Introduction

Patterns of color and regions of color texture play an important role in the visual structure of decoration designs. Consequently many pixels in design images can be naturally interpreted to take part in various color combinations. As an example consider the design image of Figure 1 (a). The background in this image consists of pixels of two colors: red and black. Rather than viewing such pixels as either red or black it is more natural to view both types of pixels as part of a red-and-black region. Moreover, pixels are often part of a nested sequence of such color combinations. This may be seen by repeating the original design to arrive at the image of Figure 1 (b). The original background pixels are now part of a larger pattern also including the so-called pied-poule motifs, which are yellow. Depending upon the scale at which a design is perceived the red and black pixels may thus also take part in a red-black-yellow combination.

In this study we set out to explore methods for color texture segmentation by direct analysis of the color combinations occurring in an image, i.e. we intend to find the natural color combinations and scales for all image pixels for which this is appropriate. We introduce the method of color coalition labeling and show how an erosion of such a labeling can provide a first selection of useful color combinations.

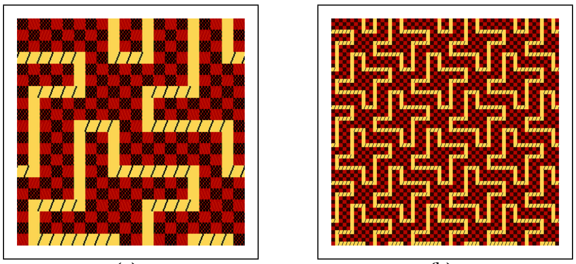

(a) (b)

Figure 1. (a) An example design image. (b) The same design image repeated 9 times and resized to its original size.

We aim to use these techniques to aid in the extraction of regions of interest of the decoration designs, and as a first application we consider the problem of figure-ground segregation.

1.1 Figure-ground segregation of decoration designs

This study was inspired by experience gained in developing a system for content-based image retrieval (CBIR) for decoration designs. See [3] for an overview of this system and a general discussion of decoration designs.

The performance of the CBIR system relies to an important extent on a module for visual design characterization, the so-called feature extraction engine. By means of the engine many types of low level features are computed, among which the visual descriptors defined by the MPEG-7 standard ([2]). Low-level features may characterize color layout, global textural properties, or for instance the dominant directions occurring in the image. However, for meaningful searching and browsing through design collections higher-level characterizations based on the individual elements in the design are required. Only then, if such elements can be identified, it becomes feasible to quantify visual properties such as shape, spatial pattern and organization, and variation in for instance color, shape and orientation among the elements.

In an important subset of the decoration designs the identification of individual design elements can take place by figure-ground segregation. The definition of the figure-ground segregation problem is, however, by no means trivial. Design elements may be arranged in a large variety of ways: they may be overlaid, may fade over into each other, or may form tilings of the image plane. Furthermore the elements may take part in many types of spatial patterns and groupings. Within such arrangements the design elements vary in their level of *salience*, i.e. by the extent to which 'they stand out'. For figure-ground segregation we are interested in those cases where design elements are arranged on a ground, i.e. the case where a number of, usually isolated, salient elements stand out on a non-salient ground. Clearly not all designs possess such ground structure, see for instance Figure 2 (a).

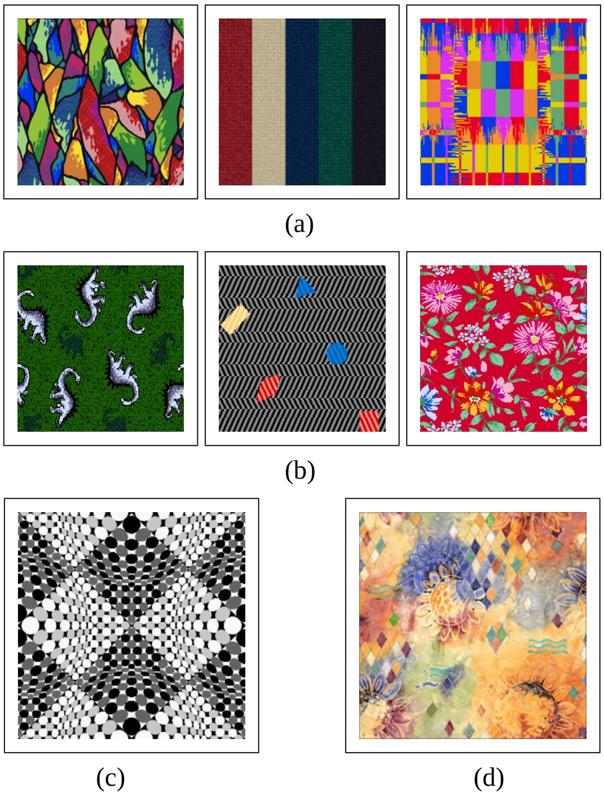


Figure 2. Examples of design images: (a) images without background; (b) images with background; (c) and (d) images for which the background-foreground structure is unclear.

Occurrence of ground structure is often not clear due to the situation that the ordering among the design elements in terms of their salience is not clear. In other cases several figure-ground interpretations are possible simultaneously. An example is shown in Figure 2 (c). The image can be interpreted as black on white, or as white on black. In some cases the occurrence of ground structure is clear, but it is still hard to determine the ground accurately (Figure 2 (d)). The latter two effects are referred to as ground *instability*.

As a final issue we mention the occurrence of nested grounds. An example to this effect is shown in Figure 3 (a). The image can be interpreted to consist of three layers: a plain green layer, a layer of heart motifs and the four angels. The background can thus be either the plain layer, or this layer together with the hearts. A special case of this problem occurs in relation to designs consisting entirely of texture where the entire image may be taken to consist of background. In such cases it is often still useful to analyze the texture additionally in terms of its figure-ground structure, see Figure 3 (b).

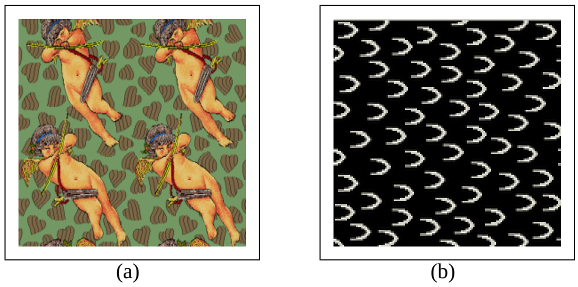


Figure 3. Examples of design images for which the figure-ground structure consists of multiple levels.

1.2 Outline of the paper

In section 2 we introduce the color coalition labeling and describe its application to the detection of regions of color texture at various scales. Next we discuss its application to figure-ground segregation. We propose a strategy consisting of three main steps:

1. Obtain initial candidates for the background by multi-scale detection of color texture regions (section 2).

2. Assess the appropriateness of the individual candidates by an N-nearest neighbor classification algorithm based on visual cues such as relative size, connectedness and massiveness (section 3).

3. Integrate the results of the previous steps to produce a hierarchical description of the figure-ground structure of the design (section 4).

The algorithms are tested by application to images from two decoration design databases. One is a database of tie designs from an Italian designer company. The other database has been provided by a manufacturer of CAD/CAM systems for the textile industry and includes a wide range of decoration design types.

In section 5 we list the results obtained for the two test sets and discuss the general performance of the approach. Conclusions are presented in section 6.

2 Finding color texture regions

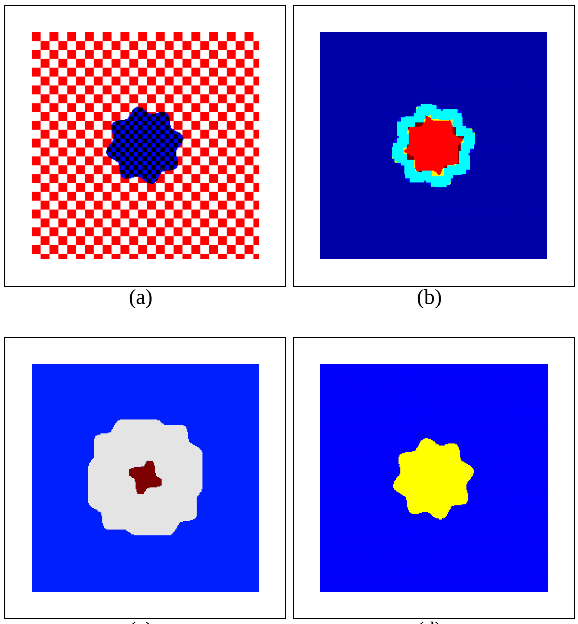
The task of finding color texture regions is closely related to general image segmentation. As any segmentation problem it is about the grouping of pixels that, in some sense, belong together. The approach we take here is based on direct analysis of the color combinations occurring in the image. As the level of homogeneity of a region depends on the scale under consideration, we must investigate the occurrence of color combinations at various scales. For each scale we then define a color coalition labeling that provides each pixel with a label uniquely identifying the colors occurring in a structuring element around the pixel.

We restrict ourselves to generate candidate regions for the image background and will not attempt a full bottom-up segmentation here. Moreover unlike in most segmentation methods we will not demand texture regions to be connected, nor will we attempt to assign every pixel to a segment region.

The algorithm for the construction of color texture regions for a fixed scale is divided in the following main stages:

1. construct color coalition labeling;
2. erode label image and analyze homogeneity of remaining color combinations;
3. grow principal color combinations into color texture regions.

These stages are outlined in Figure 4 and will be further detailed below.



(a) (b) (c) (d)

Figure 4. Main stages in construction of color texture regions: (a) test image of 256 by 256 pixels consisting of two regions of color texture; (b) based on a rectangular window of 13 by 13 pixels structuring element the image has 7 distinct color sets; (c) after erosion and homogeneity checking two color sets remain; white pixels in this image have index 0 and do not correspond to a color combination; (d) growing leads to two regions of color texture.

2.1 Color coalition labeling

In the following we consider indexed images where each pixel has an associated integer value that either refers to a color in a colormap or is equal to zero, indicating that the color of the pixel is to be ignored by the algorithm. More formally, we define an image f as a mapping of a subset $\mathcal{D}_f$ of the discrete space $\mathbb{Z}^2$, called the definition domain of the image, into the set of indices:

$$f : \mathcal{D}_f \subset \mathbb{Z}^2 \to \{0\} \cup \mathcal{C}_f = \{0, 1, \ldots, N\}, \quad (1)$$

where $\mathcal{C}_f = \{1, \ldots, N\}$ is the set of color indices of the image. In practice the definition domain is usually a rectangular frame referred to as the image plane of pixels.

For indexed images we define the *index* or *color set* $\mathrm{cs}_c(f)$ of index c as the set of pixels with index c: $\mathrm{cs}_c(f) = \{x | f(x) = c\}$, or as binary image:

$$[\mathrm{cs}_c(f)](x) = \begin{cases} 1 & \text{if } \ f(x) = c \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

We further define the erosion of an indexed image as the summation of binary erosions performed on the individual color sets while keeping the original indices:

$$\varepsilon_B(f) = \sum_c c\, \varepsilon_B(\mathrm{cs}_c(f)), \quad (3)$$

where B is the structuring element and summation and scalar multiplication are pixel-wise.

For each pixel x we consider the set of colors $\omega_{B_s}(x)$ occurring in a structuring element B_s^x of scale s:

$$\omega_{B_s}(x) = \{c \in \mathcal{C}_f | \exists y \in B_s^x : f(y) = c\}. \quad (4)$$

Each such subset of $\mathcal{C}_f$ is referred to as a *color combination*, and $\omega_{B_s}(x)$ is called the color combination associated with pixel x at scale s. For the structuring element we will usually take a rectangular window with the centre pixel as element origin.

We define the color coalition labeling of f as follows. Let Ω_{B_s} be the set of all color combinations occurring in the image at scale s, then we associate with each combination ω in Ω_{B_s} a label $\lambda_{B_s}(\omega)$ in the order of encounter of such combinations in a forward scan of the image. The color coalition labeling $\Lambda_{B_s}(f)$ of f is then defined by

$$[\Lambda_{B_s}(f)](x) = \lambda_{B_s}(\omega_{B_s}(x)). \quad (5)$$

An example of a color coalition labeling is shown in Figure 4 (b).

2.2 Principal color combination selection

Our aim is to select the principal color combinations of the image, i.e. those color combinations that are most appropriate to extend to full color texture regions for a given scale. To this end we erode each of the index sets of the color coalition labeling under the tentative assumption that the color combinations occurring at the boundaries of regions of color texture are generally thinner than the interiors of such regions. For the erosion we use a structuring element B_t of scale t, i.e. we construct $\varepsilon_{B_t}(\Lambda_{B_s}(f))$. We denote an eroded set associated with ω by $R(\omega)$, i.e. we take

$$R(\omega) = \varepsilon_{B_t}(\mathrm{cs}_{\lambda_{B_s}(\omega)}). \tag{6}$$

As we are interested in finding regions of homogeneous color texture we further investigate homogeneity statistics for color combinations ω for which $R(\omega)$ is nonempty. Note that if statistics are computed based on a structuring element of scale s, taking $t \geq s$ ensures that colors surrounding a region of color texture cannot affect the homogeneity of the statistics in an eroded color set.

So let $S_{B_s}(x)$ be the local statistics at pixel x taken over pixels in a structuring element B_s, and consider a surviving color combination $\omega : R(\omega) \neq \emptyset$. We accept ω as a principal color combination if the following two conditions hold:

1. $R(\omega)$ still contains all colors of the color combination ω.

2. The coefficients of variation of S_{B_s} on $R(\omega)$ are smaller than a given threshold

Both the choice of statistics and of the scale for the erosion structuring element t are subject to a trade-off between the aims of suppression of boundary color combinations and still being able to detect color texture regions that have a relatively large interior[1] scale relative to their exterior scale. We obtained best results by using the erosion as the main mechanism in reducing the number of candidate color combinations (we set $t = 1.5s$), and kept the statistics as simple as possible to allow for maximum detection of color textures. In fact, we take only the relative number of pixels of the dominant color in the structuring element as a statistic. The computation of the coalition labeling and the local statistic can both be implemented taking a single forward scan and a moving histograms approach (see [5]).

2.3 Region growing strategies

Next the color texture regions associated with the principal color combinations are determined by region growing of the eroded color sets. If we denote the final color texture region by $G(\omega) = \mathcal{G}(R(\omega))$ then for a pixel x to be assigned to $G(\omega)$ it should satisfy at least the following conditions:

1. the pixel must have a color index belonging to the color combination: $f(x) \in \omega$.

2. the pixel must have the remaining colors of the color combination in its structuring element: $\omega \subset \omega_{B_s}(x)$.

The pixels in $R(\omega)$ satisfy both conditions; also note that the conditions allow pixels at boundaries of texture regions to have additional colors in their structuring element.

This still leaves the important issue of how to assign pixels for which more than one color combination is feasible. Several strategies are possible such as assigning to the closest or the largest feasible eroded set. In our application we have obtained best results sofar by assigning to the color combination for which the associated eroded region has an average color histogram that is closest to the color histogram of the structuring element of the pixel.

For each scale we thus get a segmentation of the image in regions corresponding to the principal color combinations and a set of pixels with label zero that are not assigned to any color combination.

3 Classification

To determine the background quality of color texture regions, we take a simple yet effective approach based on weighted N-nearest neighbor classification. Based on a number of property variables or features of the region the *ground probability* is estimated that the region is suitable to serve as a background region.

Classification takes place by using a training set of sample regions with features $x_i, i = 1, \ldots, n$ that have been assigned a ground probability $p(x_i)$ by manual annotation. The probability $p(x)$ of a region with features x is determined by taking a weighted average of the probabilities of its N nearest neighbors in feature space, see for instance [1].

The feature variables were chosen by experimentation with the aim of reaching a high level of consistency and a low level of ambiguity:

- Relative area: the region area relative to the total image area.

- Filling coefficient: background regions often possess a complement consisting of components that are not connected to the border and which are removed after filling the background region (see for instance [4] for morphological operations such as hole removal). Let X be the background region, X^c its complement and $\bar{X}$ the background region after hole removal, then the filling coefficient $\mathrm{fc}(X)$ is defined as

$$\mathrm{fc}(X) = \begin{cases} 1 - A([\bar{X}]^c)/A(X^c) & \text{if} \quad X^c \neq \emptyset \\ 1 \quad \text{if} \quad X^c = \emptyset, \end{cases} \tag{7}$$

where $A(X)$ is the area in pixels of region X.

[1] We define the interior scale of a set as the smallest scale at which a set is homogeneous; the exterior scale as the smallest scale at which the erosion of the set is empty

- Spatial reach: measures if the region occurs only in certain parts of the image or all over the image; the image is covered by a grid of boxes and spatial reach is measured by counting the relative number of boxes that are occupied by the region.

- Connectedness: The area of the largest connected component of the region relative to the region area (computed after closing with a small structuring element.)

- Massiveness: the median distance of the region pixels to the region boundary.

The N-nearest neigbor approach allows for straightforward evaluation of inconsistencies and ambiguities. Consistency of the samples can be analyzed by comparing the ground probability of a sample region obtained by classification leaving that example out to the probability obtained by manual annotation. Let p_{-i} be the ground probability obtained by classification using all samples except sample i, then we define the consistency for sample i as $\rho_i = |p_{-i} - p(x_i)|$. In our study we took 350 samples of which only 8 had a consistency smaller than 0.75. It is also simple to assess if examples are sufficiently nearby for reliable classification: for instance by comparing distances of examples to new cases to be classified to the average of such distances occurring for the samples in the sample set. If relatively empty regions or problem cases are encountered additional samples may be added.

4 Synthesis

Using the color coalition labeling approach of section 2 we obtain principal color combinations for a sequence of scales. In this study we took 8 scales that were equally distributed over a range from a smallest window of 3 by 3 pixels to a rectangular window of about 30% of the image size. All resulting regions were classified using the method of section 3. Each region with a ground probability greater than 0.5 is accepted as a potential ground (although ground probabilities are generally found to be either 0.0 or 1.0). If a color combination is found to be feasible for serving as ground at more than one scale, we take two criteria into account to decide on the most appropriate region: (i) the simplicity of the region; (ii) the number of scales at which the particular region, or a region very similar to that region, is found (scale robustness). For the simplicity measure of the region we have taken, rather ad hoc, the sum of the number of connected regions in the background and the foreground (after opening each with a small structuring element; see [4]).

Next we further analyze the determined grounds and their associated color combinations. Every pair of combinations is assigned as either: nested, partially overlapping or disjoint. Large disjoint regions often indicate flipping behavior as in Figure 2 (c). Apart from the analysis of such relations that also includes checking the hypothesis that the entire image consists of a single color texture, for every design a highest quality background is determined using the simplicity and robustness criteria. Based on these results each of the images is automatically assigned to one of four distinct categories or *streams*: I: no figure-ground structure; II: figure-ground structure; III: consists entirely of one color texture, which itself possesses figure-ground structure; IV: consists entirely of one color texture, and does not possess further figure-ground structure. Note that such automatic stream assignment allows for data driven feature computations. For example texture features can be computed for the background regions and full texture images, whereas shape features are computed for foreground elements.

5 Results

Benchmarking figure-ground segregation algorithms for designs is generally difficult for the reasons sketched in the introduction: even for humans identification of figure-ground structure is often not unambiguous. We thus choose to restrict ourselves to cases where we clearly have a ground or we clearly do not, and check if the algorithm output aligns with human perception for the cases where occurrence of figure-ground structure is clear and stable. This approach recognizes the notion that strict bottom-up processing is generally infeasible, unless some sort of context is assumed: in this case we assume that we are dealing with images where a ground is to be identified.

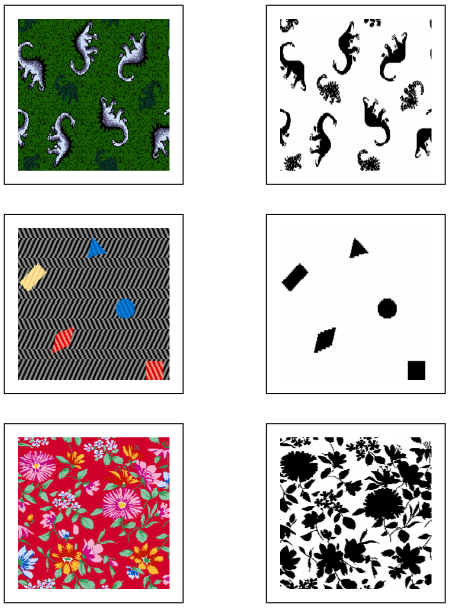


Figure 5. Results of the figure-ground segregation algorithm for the images of Figure 2 (b).

For testing we take three sets of images: (i) Collection 1: 500 images from a specialized database of tie designs, used for training the background detection algorithms; (ii) Collection 2: another 500 images from the same database; (iii) Collection 3: 500 images from a database containing a wide range of decoration designs from the textile industry. As such this database provides a representative test set for images the algorithm is likely to encounter in practice. The images of Collection 2 and 3 have not been used in any way to calibrate the algorithms.

We assigned each of the images in the test sets by manual annotation to either one of the four streams discussed in section 4 or to stream V: occurrence of structure not clear or instable. Rates of correct performance are reported in Table 1.

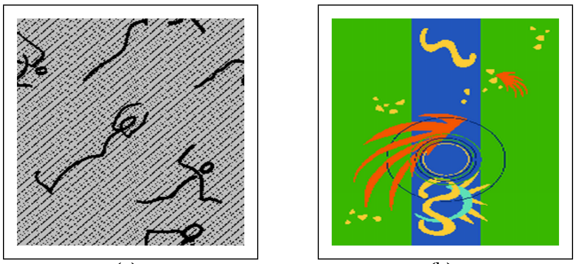
Collection	Stream I	Stream II	Stream III	Stream IV
1	89%	90%	85%	82%
2	92%	91%	77%	90%
3	100%	88%	78%	81%

Table 1. Correct performance rates for images assigned by manual annotation to stream I through IV.

Example images from stream II with results are shown in Figure 5. Errors can largely be attributed to the following types of designs:

- Designs where all colors in the foreground object also occur in the background, and the foreground objects do not disrupt the homogeneity of the background region. An example is shown in Figure 6 (a). Other methods must be used to find such additional structure in the ground.
- Cases where the background consists of a composition of regions, see for instance Figure 6 (b). Currently no combinations of regions are tested for their suitability to serve as ground.
- Cases for which classification is ambiguous, e.g. in images for which the background consists of small isolated patches, that by their shape and layout would rather be expected to be of foreground type. This type of background is hard to detect automatically and generally requires a higher level of design understanding.
- Cases where the choice of simplicity measure leads inappropriate candidates to be accepted. Occurs very rarely.
- Designs with illumination effects, gradients and special types of noise. Main problem here is the occurrence of noise that is not removed by preprocessing and occurs in only part of a color texture.
- Designs where the interior scale of a background region is large in comparison to its exterior scale. Sometimes the region is not found as a candidate since the color combination region disappears by erosion before it is accepted as homogeneous.

Correct performance is directly related to the occurrence of such types of images in the database. For example the mistakes for Collection 3 are mainly of the first type as the set has a relatively high number of binary images with additional fine structure.



(a) (b)

Figure 6. Examples of design images posing difficulties for the algorithm.

6 Conclusions and future work

In decoration designs color textures of different scales of homogeneity often coexist in a single image. Ideally one would like to construct an image where each pixel is assigned a label based on the color combination and scale of the region of which the pixel is a natural part. We set out to do so by independent investigation of the separate scales. In this study we have shown the information obtained in this manner is sufficient for the detection of color texture regions for figure-ground segregation.

In future work we aim to reach a higher level of synthesis in which color coalition labeling images of different scales are combined into a single segmentation. In our current approach color reduction is taken as a preprocessing step. Colors that are close are lumped together, and colors occurring only rarely are removed. We intend to use the color coalition information to perform more meaningful color clustering and envision the color quantization and noise removal to become an integral part of the segmentation process.

Acknowledgments

This work has been partially funded by the European Commission under the IST Programme of the Fifth Framework (Project nr. IST-2000-28427). We thank Sophis Systems NV and Pianezza Srl for supplying the decoration design test collections.

References

[1] R. Duda and P. Hart. *Pattern classification and scene analysis*. John Wiley and Sons, Inc, New York, USA, 1973.

[2] B. Manjunath and T. Sikora. Overview of visual descriptors. In B. Manjunath, P. Salembier, and T. Sikora, editors, *Introduction to MPEG-7 – multimedia content description interface*, pages 231–260. John Wiley and Sons, Ltd., Chichester, England, 2002.

[3] E. Pauwels, M. Huiskes, P. Bernard, K. Noonan, P. Vandenborre, P. Paolo, and M. De Maddelena. Foundit: Searching for decoration designs in digital catalogues. *Proceedings of WIAMIS 2003 (accepted for publication)*, 2003.

[4] P. Soille. *Morphological Image Analysis*. Springer, Berlin, Germany, 1999.

[5] M. Van Droogenbroeck and H. Talbot. Fast computation of morphological operations with arbitrary structuring elements. *Pattern Recognition Letters*, 17(14):1451–1460, 1996.