Color Image Segmentation Using Progressive Wavelet Transform 
for Image Retrieval

Din-Yuen Chan, Jin-Der Li and Been-Chian Chien 
Department of Information Engineering 
I - Shou University 
Kaohsiung, Taiwan, Republic of China 
E-mail: {dychan,m873002m,cbc}@csa500.isu.edu.tw

Abstract

Image segmentation is the preliminary process in a content-based image retrieval system. The appropriate image segmentation is the first issue to result in efficient and effective image retrieval. In this paper, we introduce a novel segmentation scheme based on progressive wavelet transform, which is able to segment out the significant objects from an image. The progressive processing eliminates the insignificant signals recursively to smooth the color of each object in an image. By setting a proper threshold, such processing can determine the levels of decomposition via wavelet transform automatically. The experimental results demonstrate that our method can classify the pictures according to their level of complexity and it can extract objects from pictures effectively. Moreover, this method is efficient. It is almost four times as fast as the previous Lin's method in average.

1. Introduction

Image segmentation is one of the fundamental techniques in many applications, such as computer vision, pattern recognition and image retrieval. For retrieving images from large image databases, how to extract objects and useful features from an image continues to be a challenging problem [1]. Recent research on content-based image retrieval integrates different features of images like color, shape and texture to measure the similarity of images. For extracting various features automatically, object recognition is necessary when images are inserted to image databases. However, the automated approach to object recognition is difficult, computationally expensive, and domain specific. For example, the image segmentation algorithm in [9] involving pattern recognition or object’s detailed features generally needs high cost of computation time and superfluous representations. The information obtained from such algorithm is usually too verbose to be the feature of an image since it is inefficient to match such data in large collections of images. To find efficient and effective methods for matching similar images, a fast extracting algorithm and a concise feature notation are required instead of the objects’ detailed contours or their actual representative meanings.

Many researches on feature extraction for image retrieval have been proposed. In 1997, Lin et al. [5] proposed a color segmentation method for image retrieval and Belongie et al. [1] proposed another segmentation method for color and texture in 1998. However, their approaches are still time-consuming. In this paper, we propose an efficient wavelet-based segmentation algorithm to extract objects from images. The proposed segmentation algorithm exploits the concept that pixels of images in neighborhood belong to the same object if their scales of color and luminance are close. The judgement of an isolated object in an image depends on the extent of uniformity of colors and strength of luminance in special locations. The wavelet transform can decompose the specified coefficients of various frequency resolutions by various basic waveforms for different applications [9]. By passing the CIE LUV image pixels through Haar wavelet transforms, the high frequency coefficients represent the differential components resulted from the around pixels. After setting thresholds on high frequency coefficients except for LL subband to filter the insignificant signals out recursively, the colors and gray levels in distinct objects can be easily grouped into their ranges of themselves more clearly. The thresholds are determined by partitioning the values of signals into two parts using the equalization of sub-variances [8]. The lower values of signals under the threshold express non-homogeneity existing in the individual object such that they are forced to zeros. This process indeed supports pixels merging for further clarifying the significant objects with uniform color and
luminance effectively.

In this paper, we will review some related researches in Section 2. The proposed wavelet-based method is presented in Section 3. The experiments and results are demonstrated in Section 4. At last, we make a conclusion in Section 5.

2. Related works

To specify regions of objects by the uniform or smooth colors for image retrieval, the color segmentation algorithm proposed in [9] exploited a simple color classifier to perform adaptive color classification on fine art images. It addresses a region-based approach to promote the image annotation technique. It needs pre-specify or pre-define the desired classes such that well results should depend on a large human knowledge base in general or only be obtained for retrieving special objects. For further outstanding the objects’ appearance, Lin et al. [5] proposed an effective block-based color segmentation method where each basic 8×8 block is initially assigned a representative color. From the basic blocks, the algorithm recursively merges smaller adjacent blocks with close colors into larger regions. During all the merging procedure, the suitable numbers of separated classes can be automatically determined by checking occurrence about the maximum change between the original image and the merged one. In this algorithm, the basic contour for the separated objects can be clearly displayed, but finer and more detailed features will not be obviously exhibited.

Jacobs et al. [4] used the Haar wavelet transform to generate a feature vector for each image. The feature vectors including color, texture and shape can be captured and the user is not required to specify any parameters. However, lacking of the segmentation technique to extract the individual objects, the user can not perform partial queries through the corresponding image databases. To overcome above drawbacks, the WALRUS system [7] employed a novel similarity model, which uses the sliding windows of varying regions for decomposing an image into the rectangular regions, based on the proximity of their signatures. Although this algorithm is effective and regular, the information about shapes of objects are not actually expressed and the significant objects can not be identified due to too many segmented regions. In addition, the representative information of such algorithm is too long to provide image retrieval under a large amount of image collections efficiently. Owing to the fast analysis in wavelet transform, we propose a wavelet-based technique to improve the efficiency and effectiveness of image segmentation. The method is presented in the next section.

3. The proposed segmentation algorithm

For achieving image segmentation effectively and efficiently, a wavelet-based segmentation algorithm is proposed to isolate the significant objects of quite complete contour in this paper. Our method can decompose the image into separated objects that satisfy the visual perception of human automatically. Moreover, the proposed technique applies the fastest Haar wavelet transform to perform the progressive processes of analysis-by-decomposition, so the algorithm is very efficient. The progressive coding technique [6] is very popular in scalable transmission systems. The concept of progressive processing is used to construct the framework of segmentation in this paper; the progressive wavelet transform decomposition can be dynamically terminated according to the complexity of images. The corresponding algorithm is mainly completed by the following three processing steps recursively.

1. Analyze the high frequency signals obtained by Haar wavelet transform.
2. Find an appropriate threshold for the high frequency signals and drop off the signals under the threshold to smooth the colors in each individual object and result in larger color contrast between different objects.
3. Detect whether the threshold converges or not. If it does, the processes are terminated and the significant objects are extracted by the colors in the last LL-subband, or the wavelet transform of next level are performed on the LL-band and go to step 1, otherwise.

3.1 Color analysis based on wavelet transform

For practical implementation, hierarchical image decomposition can be generated stage-by-stage by iterating the convolution of the image and the wavelet bases over a set of progressively reduced resolution levels. The wavelet transform can achieve entropy concentration, de-correlation and feature detection to provide excellent signal analysis. The discrete wavelet transform (DWT) decomposes a discrete time sequence η(n) into a set of scaling functions \( \{ \gamma_{a,b}(n) \mid a, b \} \) by using wavelet basis functions:

\[
\gamma_{a,b}(n) = \sum_{\xi} \eta(n - \xi) \psi_{a,b}(\xi),
\]

∀γ
where the wavelet basis function $\psi_{a,b}(n)$ is obtained by scaling and shifting a single mother wavelet function $\psi(n)$ as

$$\psi_{a,b}(n) = \frac{1}{\sqrt{a}} \psi\left(\frac{n-b}{a}\right).$$

In DWT, the scaling factor $a$ is of the power of two and the shifting factor $b$ is an integer. The fastest wavelet transform uses a simplest mother wavelet named Haar function, which is composed of a pair of rectangular pulses as

$$\psi(t) = \begin{cases} 1 & 0 \leq t < 1/2; \\ -1 & 1/2 \leq t < 1; \\ 0 & \text{otherwise}. \end{cases}$$

In the multiresolution system built on wavelet filter banks, theoretically different wavelet subbands can be analyzed and processed with varying special techniques individually. While the target image being progressively decomposed, we can trace influence and effect from the modification of each individual subband in various levels. In this paper, by examining all the experimental results, without loss of generality, we found that all subbands can be classified into only two parts after each Haar wavelet transform. One just contains the LL-subband coefficients in the current level of wavelet transform and another includes that in the high frequency subbands of all levels. In fact, except the final decomposition, only the signals of the latter part are sent to analyze color differences and perform the color smoothing for all objects in the target image along each decomposing.

Our proposed algorithm can provide extremely fast color segmentation for applications of image retrieval. We choose the LUV representation since the LUV color space fits the visual perception of human. The LUV color space [3] can be transformed from RGB color representation. The transforming method is as follows:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.490 & 0.310 & 0.200 \\ 0.177 & 0.813 & 0.011 \\ 0.000 & 0.100 & 0.990 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix},$$

for $u' = \frac{4X}{X + 15Y + 3Z}$, $v' = \frac{9Y}{X + 15Y + 3Z}$,

$u'_0 = \frac{4X_0}{X_0 + 15Y_0 + 3Z_0}$, $v'_0 = \frac{9Y_0}{X_0 + 15Y_0 + 3Z_0}$,

$L = 116 \times (Y / Y_o)^{1/3}$, $Y / Y_o > 0.01$,

$u = 13 \times L (u - u'_0)$,

$v = 13 \times L (v - v'_0)$,

where the referential color vector $[X_o, Y_o, Z_o]$ is yielded from transforming the vector $[R, G, B]$ of white pixel. We adopt CIE LUV images as the analyzed medium to pass their $L$, $u$ and $v$ values of pixels through the 2D Haar DWT. In our method, the so-called high frequency coefficients are referred as the coefficients of transforming outcomes in the HHI-subband, HL-subband and LH-subband of each level of DWT, and the low frequency coefficients are located in only the LL-subband of the current level. For characteristics of Haar wavelet transform, the differences of pixels’ colors between two adjacent areas can be directly or indirectly represented by high frequency coefficients.

First, we express the strength of difference between neighboring pixels with the mean square roots (MSR) of high frequency coefficients including $L$, $u$, and $v$. After the $i$th decomposing of LL subband, the high frequency signal $s_{hi}$ and the low frequency signal $s_{li}$ are computed by

$$s_{hi} = \sqrt{L_{hi}^2 + u_{hi}^2 + v_{hi}^2},$$

$$s_{li} = \sqrt{L_{li}^2 + u_{li}^2 + v_{li}^2},$$

where $L_{hi}$, $u_{hi}$ and $v_{hi}$ are the high frequency coefficients in the HL, LH and HH subbands of wavelet transform at the $i$th level in the LUV color space, and $L_{li}$, $u_{li}$ and $v_{li}$ are the low frequency coefficients in the LL subband of wavelet transform at the $i$th level.

3.2 The decision of eliminating threshold
Figure 1. The decision of eliminating threshold $\tau_i$.

After above wavelet transform, the signals in high frequency can be partitioned into two regions. The lower part with small values can be considered to be the precise change of colors in each object. Such values may blur the judgement of boundaries between neighboring objects, they should be neglected for clarifying the contours and unifying the colors of objects. Figure 1 shows one of the experimental examples that the high frequency signals belonging to the region $R_0$ will be discarded as above description. The parameter $\tau$ at cutting line is the eliminating threshold, which will partition the signals into two regions $R_0$ and $R_1$. In the following, we will illustrate the method of the progressive wavelet transform and the determination of the eliminating threshold in detail.

For generalization, to eliminate the high frequency signals less than the eliminating threshold can be formulated as a recursive form along the progressive decomposing. As the $i$th decomposing of the LL subband is performed, the corresponding histogram is generated and denoted as $h_i(s_{HI})$. A local eliminating threshold $\tau_i$ is determined for dropping the current high frequency signals and the corresponding $L_i$, $u$ and $v$ coefficients in the $i$th level. Meanwhile, the $i$th global eliminating threshold $T_i$ is proceeded to determine for forcing all the high frequency signals less than it to zeros in the subbands from the lowest level to the $i$th. After above modification, the variable $s_{HI}$ expresses the high frequency signals including the just modified signals and the global high frequency signals $s_{HI(i-1)}$ remained at last time. Specifically, at the time, an appropriate histogram function $g(s_{HI})$ can be formulated by the following recursive equations:

$$
g_0(s_{HI}) = 0; \quad g_i(s_{HI}) = [g_{i-1}(s_{HI(i-1)}) + f(s_{HI}, \tau_i) \cdot h_i(s_{HI})] \cdot f(s_{HI}, T_i), \quad \text{for } i \geq 1,
$$

where variable $s_{HI}$ represents the modified globally high frequency signals, and the hard-limiting function $f(x, y)$ is defined as

$$
f(x, y) = \begin{cases} 
1, & x \geq y; \\
0, & \text{otherwise},
\end{cases}
$$

To mask high frequency signals less than the threshold as zeros. By above recursion, the appropriate histogram and the remained high frequency signals can be quickly obtained from the previous ones.

Before the achievement of above appropriate histogram, it is necessary to decide the eliminating threshold $\tau_i$ from the unmodified histogram $p_i(s_{HI}) = g_{i-1}(s_{HI(i-1)}) + f(s_{HI}, \tau_i) \cdot h_i(s_{HI})$. For achieving the fast algorithm, we proposed a variance equalization technique to derive the eliminating thresholds from dividing above unmodified histogram into two parts of equal variances. In Figure 1, as soon as the ($i$-1)th level LL-band is decomposed by Haar wavelet transform, the histogram $h_i(s_{HI})$ is separated into two regions $R^0_i$ and $R^1_i$ by the center in horizontal axis as the initial boundary, and compute their individual variances $V^0_i$ and $V^1_i$ as

$$
V^0_i = \sum_{s_{HI} \in R^0_i} \left[ \frac{p_i(s_{HI}) \cdot (s_{HI} - E^0_i)^2}{\sum_{s_{HI} \in R^0_i} p_i(s_{HI})} \right], \quad \text{where } E^0_i = \frac{\sum_{s_{HI} \in R^0_i} p_i(s_{HI}) \cdot s_{HI}}{\sum_{s_{HI} \in R^0_i} p_i(s_{HI})},
$$

$$
V^1_i = \sum_{s_{HI} \in R^1_i} \left[ \frac{p_i(s_{HI}) \cdot (s_{HI} - E^1_i)^2}{\sum_{s_{HI} \in R^1_i} p_i(s_{HI})} \right], \quad \text{where } E^1_i = \frac{\sum_{s_{HI} \in R^1_i} p_i(s_{HI}) \cdot s_{HI}}{\sum_{s_{HI} \in R^1_i} p_i(s_{HI})},
$$
where $E_o^R$ and $E_i^R$ are the mean values of regions $R_0^R$ and $R_1^R$, respectively. By continuously moving the boundary to the region of the larger variance, the most proper boundary $\tau$ can be found out by terminating the moving as the difference between $V_0^R$ and $V_1^R$ is minimal or enough small. Therefore, the $\tau$ determination can be formulated as

$$\min_{\tau} \{|V_1^R - V_0^R|\}.$$ 

Being similar to above minimization process for all of high frequency subbands, then the threshold $T_i$ can be yielded. At this moment, the current partition lets the summation of total sub-variances minimum. With the recursive procedure mentioned above, the unimportant colors’ differences are progressively dropped level by level that the population of high frequency zero signals grows up. Thus, the colors of merged neighboring pixels are getting closer and closer along the Haar wavelet transform. Finally, every object uses a homogeneous color for outstanding itself for isolation facilitation.

### 3.3 Criterion of termination and indication of significant objects

In fact, to let the proposed progressive procedure suitably stop is very important because too far decomposing by the wavelet transform will cause the separation of heterogeneous adjacent objects failing. Therefore, we design a termination criterion for terminating the progressive decomposing. At the termination, major individual objects can be successfully separated out according to the visual perception.

Since the small high frequency signals are forced as zero signals, except for zero signals, the distribution about the smaller signals gradually becomes sparser that make the eliminating threshold increasing along each decomposition-elimination process. Simultaneously, the increase of eliminating threshold will lead the variance of the lower part in the appropriate histogram to get lower and lower. But, the decomposition-elimination process will not influence the large frequency signals, i.e., significant differential colors. For above phenomenon, we can easily design the termination criterion by using the local termination threshold as

$$\theta_j = \frac{\max\{S_{ij}\} - \tau_j}{\max\{S_{ij}\}},$$

where $\epsilon$ is an experimental termination threshold. As soon as the ratio $\theta_j \leq \epsilon$, the progressive procedure is terminated immediately and the resulted index $j$ expresses the stopped decomposition level. At the end of segmentation, the remained larger frequency coefficients strengthen the discrimination between the significant objects. And, the $L$, $u$ and $v$ coefficients in the $j$th LL-band are then re-quantized that the number of color vectors is set as that of the isolated objects. Then, these LUV vectors play the key color references that each pixel will be mapped into its own object by searching the nearest key color reference. Through the proposed segmentation, users can exploit the resulted objects to easily make effective queries in the submit-and-return database retrieval.

### 4. Experimental results

For demonstrating the efficiency and effectiveness of our segmentation mechanism, the proposed method was implemented in C++ language on a personal computer with Intel Pentium 233 CPU and 64M RAM. The total number of tested images is about 1000. Each tested image is in size of 256×256. Types of the images include trademark, natural picture and image of texture. The contents of images are various in each type. Some of the pictures are simple but some of them are more complicated. To segment all images appropriately and automatically, the decision of threshold value $\epsilon$ is important in our approach. Since the levels of wavelet process in our method determine the segmented results, proper wavelet processing levels should be found for different images. Here, the value of $\epsilon$ is used to handle the termination of progressive process. The segmentation algorithm will be halted when $\theta$ is either less than or equal to $\epsilon$. For simple pictures, such as trademarks or pictures with few objects, whose values of $\theta$ are small in lower levels, like first level or second level, because most of the large variation of colors occurred only on the boundary lines of objects. These simple images can be decomposed well at lower level. Complex images, contrary to simple images, have larger values of $\theta$ and the decomposed results can not fit visual perception in lower levels. It always needs segmentation of higher levels. Thus, the value of $\epsilon$ is the key parameter for deciding the quality of segmentation. According to our experiments, the segmented results are close to human observation when the value of $\epsilon$ is 0.01 for most images. Consequently, we can precisely analyze all the various kinds of images to deduce their own suitable $\epsilon$ values that may lead the recursion of segmentation algorithm to converge more naturally.

The experimental results are shown in Figure 2 and Figure 3. The pictures shown in this paper are selected from the 1000 tested images randomly. The original images are shown in the left columns of Figure 2 and Figure 3. The names of
original images are given below the images. The middle columns of the figures are the segmented results of our proposed method with $\varepsilon = 0.01$. Under the segmented images, the terminal levels of progressive process and execution times are described. The right columns of the figures are the segmented results of Lin's method proposed in [5] and their execution times.

As Figure 2 shows, both of Mark1 and Sis are trademarks and Mark1 is simpler. They can be decomposed at level 1 and level 2. For simple natural pictures having clear boundaries and few objects, like Rgb004, Rgb006 and Rgb010, the algorithm will be stopped at level 3 and level 4, respectively. The segmented results are better than Lin's method. For complex pictures, such as Cat, Circle and Fruit2, they usually have more objects with distinct colors in them or vague boundaries caused by out of focus or shadows. These pictures will not be decomposed well until level 5 or level 6. Finally, the more complex pictures with complicated background or texture need processing continuously at higher level. However, we suggest that developers had better use other models to capture the features of such pictures, because it is difficult to recognize objects from these pictures correctly even for the human. For example, pictures Tree5 and Flowers in Figure 3 are so complicated as to be unable to be decomposed meaningfully, although the segmented results of our method is more effective and meaningful than Lin's method. For computation time, our method is far less than Lin's method. It takes half of computation time of Lin's method even in the worst case. Most of the pictures can be decomposed in one second successfully.

5. Conclusions

A fast and correct segmentation for images is an important task in many applications. In this paper we develop a method based on progressive Haar wavelet transform to segment objects out of an image. By exploiting the natural characteristics of image’s own, a suitable decomposition level can be determined automatically. At the end of the Haar transform decomposition, the final colors are used to classify objects in the image. After decomposing, the result is close to the visual perception of human. Our experimental results also show that the proposed mechanism indeed improves the efficiency and effectiveness. The proposed segmentation mechanism can be applied to image retrieval as [5]. Nevertheless, more features can be extracted from our results of segmentation, e.g. shapes, since the shape belonging to an object is captured precisely. Applying the algorithm to design a more effective content-based image retrieval system is our future work.

References


<table>
<thead>
<tr>
<th>Original Images</th>
<th>Our Results</th>
<th>Lin's Results [8]</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="Image1.png" alt="Image" /></td>
<td><img src="Image2.png" alt="Image" /></td>
<td><img src="Image3.png" alt="Image" /></td>
</tr>
<tr>
<td>Mark1</td>
<td>Level 1</td>
<td>0.206 sec.</td>
</tr>
<tr>
<td><img src="Image4.png" alt="Image" /></td>
<td><img src="Image5.png" alt="Image" /></td>
<td><img src="Image6.png" alt="Image" /></td>
</tr>
<tr>
<td>Sis</td>
<td>Level 2</td>
<td>0.606 sec.</td>
</tr>
<tr>
<td><img src="Image7.png" alt="Image" /></td>
<td><img src="Image8.png" alt="Image" /></td>
<td><img src="Image9.png" alt="Image" /></td>
</tr>
<tr>
<td>Rgb004</td>
<td>Level 3</td>
<td>0.861 sec.</td>
</tr>
<tr>
<td><img src="Image10.png" alt="Image" /></td>
<td><img src="Image11.png" alt="Image" /></td>
<td><img src="Image12.png" alt="Image" /></td>
</tr>
<tr>
<td>Rgb006</td>
<td>Level 4</td>
<td>1.081 sec.</td>
</tr>
<tr>
<td><img src="Image13.png" alt="Image" /></td>
<td><img src="Image14.png" alt="Image" /></td>
<td><img src="Image15.png" alt="Image" /></td>
</tr>
<tr>
<td>Rgb010</td>
<td>Level 4</td>
<td>1.076 sec.</td>
</tr>
</tbody>
</table>

**Figure 2.** The examples of images and results of segmentation (level 1 to level 4).
<table>
<thead>
<tr>
<th>Original Images</th>
<th>Our Results</th>
<th>Lin’s Results [8]</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Cat" /></td>
<td><img src="image2" alt="Cat" /></td>
<td><img src="image3" alt="Cat" /></td>
</tr>
<tr>
<td>Cat</td>
<td>Level 5</td>
<td>1.375 sec.</td>
</tr>
<tr>
<td><img src="image4" alt="Circle" /></td>
<td><img src="image5" alt="Circle" /></td>
<td><img src="image6" alt="Circle" /></td>
</tr>
<tr>
<td>Circle</td>
<td>Level 5</td>
<td>1.295 sec.</td>
</tr>
<tr>
<td><img src="image7" alt="Fruit2" /></td>
<td><img src="image8" alt="Fruit2" /></td>
<td><img src="image9" alt="Fruit2" /></td>
</tr>
<tr>
<td>Fruit2</td>
<td>Level 6</td>
<td>1.777 sec.</td>
</tr>
<tr>
<td><img src="image10" alt="Tree5" /></td>
<td><img src="image11" alt="Tree5" /></td>
<td><img src="image12" alt="Tree5" /></td>
</tr>
<tr>
<td>Tree5</td>
<td>Level 7</td>
<td>2.021 sec.</td>
</tr>
<tr>
<td><img src="image13" alt="Flowers" /></td>
<td><img src="image14" alt="Flowers" /></td>
<td><img src="image15" alt="Flowers" /></td>
</tr>
<tr>
<td>Flowers</td>
<td>Level 8</td>
<td>2.44 sec.</td>
</tr>
</tbody>
</table>

**Figure 3.** The examples of images and results of segmentation (level 5 to level 8).