SPCD - SPATIAL COLOR DISTRIBUTION DESCRIPTOR
A Fuzzy Rule based Compact Composite Descriptor Appropriate for Hand Drawn Color Sketches Retrieval

Savvas A. Chatzichristofis[a], Yiannis S. Boutalis[a,b] and Mathias Lux[c]

[a]*Department of Electrical and Computer Engineering, Democritus University of Thrace, Xanthi 67100, Greece*

[b]*Department of Electrical, Electronic and Communication Engineering, Chair of Automatic Control*
Friedrich-Alexander University of Erlangen-Nuremberg, 91058 Erlangen, Germany

[c]*Institute of Information Technology, Klagenfurt University, Austria*

Keywords: Image retrieval, Compact composite descriptors, Spatial distribution, Hand-drawn sketches.

Abstract: In this paper, a new low level feature suitable for Hand Drawn Color Sketches retrieval is presented. The proposed feature structure combines color and spatial color distribution information. The combination of these two features in one vector classifies the proposed descriptor to the family of Composite Descriptors. In order to extract the color information, a fuzzy system is being used, which is mapping the number of colors that are included in the image into a custom palette of 8 colors. The way by which the vector of the proposed descriptor is being formed, describes the color spatial information contained in images. To be applicable in the design of large image databases, the proposed descriptor is compact, requiring only 48 bytes per image. Experiments demonstrate the effectiveness of the proposed technique.

1 INTRODUCTION

As content based image retrieval (CBIR) is defined any technology, that in principle helps to organize digital image archives by their visual content. By this definition, anything ranging from an image similarity function to a robust image annotation engine falls under the purview of CBIR (Datta et al., 2008).

In CBIR systems, the visual content of the images is mapped into a new space named the feature space. The features that are chosen have to be discriminative and sufficient for the description of the objects. The key to attaining a successful retrieval system is to choose the right features that represent the images as “strong” as possible (Chatzichristofis and Boutalis, 2007). A feature is a set of characteristics of the image, such as color, texture, and shape. In addition, a feature can be enriched with information about the spatial distribution of the characteristic, that it describes.

Regarding CBIR schemes which rely on single features like color and/or color spatial information several schemes have been proposed. The algorithm proposed in (Jacobs et al., 1995) makes use of multiresolution wavelet decompositions of the images. In (Pass et al., 1997), each pixel as coherent or noncoherent based on whether the pixel and its neighbors have similar color. In (Rao et al., 1999)(Cinque et al., 2001)(Lim and Lu, 2003) are presented the Spatial Color Histograms in which, in addition to the statistics in the dimensions of a color space, the distribution state of each single color in the spatial dimension is also taken into account. In (Sun et al., 2006) a color distribution entropy (CDE) method is proposed, which takes account of the correlation of the color spatial distribution in an image. In (Huang, 1997) a color correlograms method is proposed, which collects statistics of the co-occurrence of two colors. A simplification of this feature is the autocorrelogram, which only captures the spatial correlation between identical colors. The MPEG-7 standard (Manjunath et al., 2001) includes the Color Layout Descriptor (Kasutani and Yamada, 2001), which represents the spatial distribution of color of visual signals in a very compact form.

The schemes which include more than one features in a compact vector can be regarded that they belong to the family of Compact Composite Descriptors (CCD). In (Chatzichristofis and Boutalis, 2008a) and (Chatzichristofis and Boutalis, 2008b) 2 descriptors are presented, that contain color and texture information at the same time in a very compact representation.

A. Chatzichristofis S., S. Boutalis Y. and Lux M. (2010).
SPCD - SPATIAL COLOR DISTRIBUTION DESCRIPTOR - A Fuzzy Rule based Compact Composite Descriptor Appropriate for Hand Drawn Color Sketches Retrieval.
In *Proceedings of the 2nd International Conference on Agents and Artificial Intelligence - Artificial Intelligence*, pages 58-63
DOI: 10.5220/0002725800580063

In (Chatzichristofis and Boutalis, 2009) a descriptor is proposed, that includes brightness and texture information in a vector with size of 48 bytes.
In this paper a new CCD is proposed, which combines color and spatial color distribution information. The descriptors of this type can be used for image retrieval by using hand-drawn sketch queries, since this descriptor captures the layout information of color feature. In addition, the descriptors of this structure are considered to be suitable for colored graphics, since such images contain relatively small number of color and less texture regions than the natural color images.
The rest of the paper is organized as follows: Section 2 describes how to extract the color information, which is embedded in the proposed descriptor, while Section 3 describes in details the descriptor's formation. Section 4 contains the experimental results of an image retrieval system that uses either the proposed desciptor or the MPEG-7 CLD descriptor on two benchmarking databases. Finally, the conclusions are given in Section 5.

2 COLOR INFORMATION

An easy way to extract color features from an image is by linking the color space channels. Linking is defined as the combination of more than one histogram to a single one. An example of color linking methods is the Scalable Color Descriptor (Manjunath et al., 2001), which is included in the MPEG-7 standard.
In the literature several methods are mentioned, that perform the linking process by using Fuzzy systems. In (Konstantinidis et al., 2005) the extraction of a fuzzy-linking histogram is presented based on the color space *CIE-L*a*b**. Their 3-input fuzzy system uses the *L**, *a** and *b** values from each pixel in an image to classify that pixel into one of 10 preset colors, transforming the image into a palette of the 10 preset colors. In this method, the defuzzyfication algorithm classifies the input pixel into one and only one output bin (color) of the system (crisp classification). Additionally, the required conversion of an image from the *RGB* color space to *CIEXYZ* and finally to *CIE-L*a*b** color space makes the method noticeably time-consuming. In (Chatzichristofis and Boutalis, 2007), the technique is improved by replacing the color space and the defuzzyfication algorithm. In (Chatzichristofis and Boutalis, 2008a), a second fuzzy system is added, in order to replace the 10 colors palette with a new 24 color palette.
In this paper a new fuzzy linking system is proposed, that maps the colors of the image in a custom 8 colors palette. The system uses the three channels of HSV as inputs, and forms an 8-bin histogram as output. Each bin represents a present color as follows: (0) Black, (1) Red, (2) Orange/Yellow, (3) Green, (4) Cyan, (5) Blue, (6) Magenta and (7) White.
The system's operating principle is described as follows: Each pixel of the image is transformed to the HSV color space. The H, S and V values interacts with the fuzzy system. Depending on the activation value of the membership functions of the 3 system inputs, the pixel is classified by a participation value in one or more of the preset colors, that the system uses.

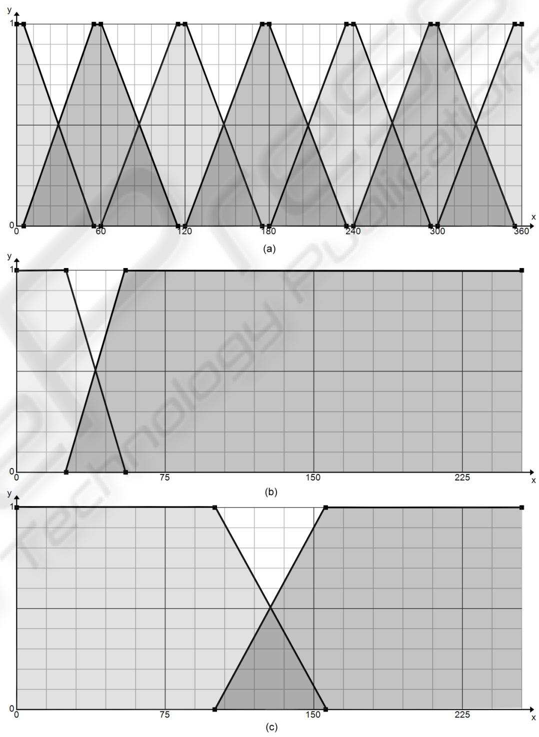

Figure 1: Membership Functions of (a) Hue, (b) Saturation and (c) Value.

For the defuzzification process, a set of 28 TSK-like (Zimmermann, 1987) rules is used, with fuzzy antecedents and crisp consequents.
The Membership Value Limits and the Fuzzy Rules are given in the Appendix.

3 DESCRIPTOR FORMATION

In order to incorporate the spatial distribution information of the color to the proposed descriptor, from each image, 8 Tiles with size of 4x4 pixels are created. Each Tile corresponds to one of the 8 preset colors described in Section 2. The Tiles are described as $T(c)$,$c \in [0,7]$, while the Tiles' pixel values are described as $T(c)_{M,N}$, $M,N \in [0,3]$. The Tiles' creation process is described as follows: A given image J is first sub-divided into 16 sub-images according to figure 2(a). Each sub image is denoted as $\acute{J}_{M,N}$. Each pixel that belongs to $\acute{J}_{M,N}$ is denoted as $P(i)_{\acute{J}_{M,N}}$, $i \in [0,I]$, I Number of pixels in $\acute{J}_{M,N}$. Each $\acute{J}_{M,N}$ contributes to the formation of the $T(c)_{M,N}$ for each c. Each $P(i)_{\acute{J}_{M,N}}$ is transformed to the HSV color space and the values of H, S and V constitute inputs to the Fuzzy-Linking system, which gives a participation value in space $[0,1]$ for each one of the 8 preset colors. The participation value of each color is defined as $MF(c)$, $c \in [0,7]$.

$$\sum_{c=0}^{7} MF(c) = 1 \tag{1}$$

Each $T(c)_{M,N}$ increases by $MF(c)$. The process is repeated for all $P(i)_{\acute{J}_{M,N}}$ until $i = I$. Subsequently, the value of each $T(c)_{M,N}$ for each c is replaced according to the following formula:

$$T(c)_{M,N} = \frac{T(c)_{M,N}}{I} \tag{2}$$

In this way, a normalization on the number I of the pixels, that participate in any $\acute{J}_{M,N}$ is achieved. On the completion of the process, the value of each $T(c)_{M,N}$ is quantized using a linear quantization table to an integer value within the interval $[0,7]$.

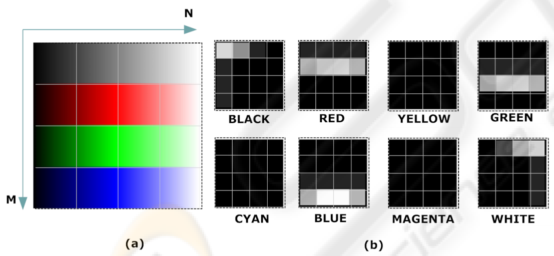

Figure 2: (a) The image is divided into 16 sub images, (b) Result of the 8 Tiles production from the image (a).

The described process is repeated for each (M,N). On completion of the process 8 images of 3 bits are produced, where each one of them describes the distribution (quantitative and spatial) of each color. Figure 2(b) shows the resulting Tiles of Figure 2(a).
Scanning row-by-row each tile $T(c)$, a 16-dimensional vector can be formed, with each vector element requiring 3-bits for its representation. In this vector the element's index determines position (M,N) in the tile $T(c)_{M,N}$. The element's value determines the color quantity at the position (M,N) quantized in space $[0,7]$. By repeating the process for all the Tiles, 8 such vectors can be produced, which if combined, by placing them successively, a 48 bytes descriptor is formed.

$$16(Elements) \times 8(Tiles) \times \frac{3}{8}(bytes) = 48bytes$$

This size compared with the size of other compact composite descriptors is considered satisfactory. The problem occurs in the length of the final vector which is set in 128 elements. During data retrieval from databases, the length of the retrieved information is of great significance. In order to compress the length of the proposed descriptor the following information lossless procedure is applied. The 8 Tiles of 3 bits are combined in order to form a new 24-bit Tile ("color" Tile) of the same size. This Tile is defined as ϕ, while the values of the pixels are described as $\phi_{M,N}$, $M,N \in [0,3]$. In order to define each $\phi_{M,N}$, 3 values are needed. $R(\phi_{M,N})$ describes the amount of red, $G(\phi_{M,N})$ describes the amount of green and $B(\phi_{M,N})$ describes the amount of blue.
The value of each $T(c)_{M,N}$ is expressed in binary form. Given that the value of each $T(c)_{M,N}$ is described by 3 bits, 3 binary digits are needed for its description.
The binary form of $T(c)_{M,N}$ value is defined as a 3 places matrix, $T(c)_{M,N}[B]$, $B \in [0,2]$.
$\phi_{M,N}$ pixel is shaped using the following rules:

$$R(\phi_{M,N}) = \sum_{c=0}^{7} 2^c \times T(c)_{M,N}[0] \tag{3}$$

$$G(\phi_{M,N}) = \sum_{c=0}^{7} 2^c \times T(c)_{M,N}[1] \tag{4}$$

$$B(\phi_{M,N}) = \sum_{c=0}^{7} 2^c \times T(c)_{M,N}[2] \tag{5}$$

The process is repeated for each (M,N).
In order to make clear the production process of ϕ, the following example is given. The $T(c)_{M,N}$ values of the 8 Tiles for a given image J are presented in Table 1. Next to the 3-bit representation of each value, the binary form of the value appears. The first bit represents $T(c)_{M,N}[0]$, the second one represents $T(c)_{M,N}[1]$ and the third bit represents $T(c)_{M,N}[2]$. The value of each $\phi_{M,N}$ channel depends on the combination, using Eq. 3, Eq. 4 or Eq. 5, of all the $T(c)_{M,N}[B]$, $c \in [0,7]$. For example, The Blue channel of $\phi_{M,N}$ depends on the combination of all the $T(c)_{M,N}[2]$ (last column of the table). At the given example, the $B(\phi_{M,N})$ value is equal to $0 \times 2^1 + 1 \times 2^1 + 0 \times 2^2 + 1 \times 2^3 + 1 \times 2^4 + 0 \times 2^5 + 0 \times 2^6 + 1 \times 2^7 = 153$.

Table 1: Combining the 8 $T(c)_{M,N}$ values to create the $\phi_{M,N}$ value.

c	3-Bit Value	$T(c)_{M,N}[0]$	$T(c)_{M,N}[1]$	$T(c)_{M,N}[2]$
0	1	0	0	1
1	0	0	0	0
2	0	0	0	0
3	3	0	1	1
4	1	0	0	1
5	0	0	0	0
6	0	0	0	0
7	1	0	0	1
		$R(\phi_{M,N})$	$G(\phi_{M,N})$	$B(\phi_{M,N})$
		0	8	153

After applying the procedure to each (M,N) the ϕ combined "color" tile is formed.
Scanning row-by-row the ϕ, a 48-dimensional vector is formed.

$$16(Elements) \times 3(Channels) = 48Elements$$

where the values of each element are quantized in the interval $[0,255]$ requiring 1 byte each. Therefore, the storage requirements of the descriptor remained the same (48 bytes), but its length was **losslessly** compressed by 62.5%. This feature allows the faster retrieval from the databases, which it is stored.

4 EXPERIMENTAL RESULTS

The extraction process of the proposed descriptor is applied to all images of the database and an XML file with the entries is created. Given that, the proposed descriptor is an MPEG-7 like descriptor, the schema of the SpCD as an MPEG-7 extension is described as follows.

```
<?xml version ="1.0" encoding="UTF-8"?>
<schema xmlns="http://www.w3.org/2001/XMLSchema"
        xmlns:mpeg7="urn:mpeg:mpeg7:schema :2004"
        xmlns:SpCDNS="SpCDNS" targetNamespace="SpCDNS">
  <import namespace="urn:mpeg:mpeg7:schema:2004"
                              schemaLocation="Mpeg7-2004.xsd"/>
  <complexType name="SpCDType" final="#all">
    <complexContent>
      <extension base="mpeg7:VisualDType">
       <sequence>
         <element name="value">
          <simpleType>
           <restriction>
            <simpleType>
              <listitemType="mpeg7:unsigned3"/>
            </simpleType>
             <length value="48"/>
           </restriction>
          </simpleType>
         </element>
       </sequence>
      </extension>
    </complexContent>
  </complexType>
</schema>
```

During the retrieval process, the user enters a query image in the form of Hand Drawn Color Sketch. From this image the 8 Tiles are exported and the 128-dimensional vector is formed. Next, each descriptor from the XML file is transformed from the 48-dimensional vector to the corresponding 128-dimensional vector, following the exact opposite procedure to that described in Section 3.
For the similarity matching we propose, in accordance to the other Compact Composite Descriptors, the Tanimoto coefficient. The distance T of the vectors X_i and X_j is defined as T_{ij} and is calculated as follows:

$$T_{ij} = t(X_i, X_j) = \frac{X_i^T \times X_j}{X_i^T \times X_i + X_j^T \times X_j - X_i^T \times X_j} \quad (6)$$

Where X^T is the transpose vector of X. In the absolute congruence of the vectors the Tanimoto coefficient takes the value 1, while in the maximum deviation the coefficient tends to zero.

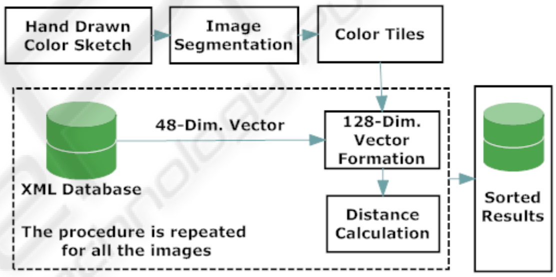

Figure 3: Image Retrieval Procedure.

After repeating the process in all the images contained in the XML file, the images appear to the user sorted according to the distance with the query image. The process is described in Figure 3.
In order to evaluate the performance of the proposed descriptor, experiments were carried out on 2 benchmarking image databases. The first database contains 1030 images, which they depict flags from around the world. The database consists of state flags, flags of unrecognized states, flags of formerly independent states, municipality flags, dependent territory flags and flags of first-level country subdivisions. All the images are taken from Wikipedia. We used 20 Hand Drawn Color Sketches as query images and as ground truth (list of similar images of each query) we considered only the flag that was attempted to be drawn with the Hand Drawn Color Sketch.
The "Paintings" database was also used. This database is incorporated in (Zagoris et al., 2009) and includes 1771 images of paintings by internationally well known artists. Also in this case, as query images were used 20 Hand Drawn Color Sketches and as

ground truth was considered only the painting, which was attempted to be drawn with the Hand Drawn Color Sketch. Figure 4 shows a query sample from each database as well as the retrieval results.

The objective Averaged Normalized Modified Retrieval Rank (ANMRR)(Manjunath et al., 2001) is employed to evaluate the performance of the proposed descriptor.

The average rank AVR(q) for a given query q is:

$$AVR(q) = \sum_{k=1)}^{NG(q} \frac{Rank(k)}{NG(q)} \tag{7}$$

Where

- $NG(q)$ is the number of ground truth images for the query q. In our case $NG(q) = 1$
- K is the top ranked retrievals examined where:
  - $K = min(X \times NG(q), 2GMT)$, $GMT = max(NG(q))$, in our case $K = 2$
  - $NG(q) > 50$ then $X = 2$ else $X = 4$
- $Rank(k)$ is the retrieval rank of the ground truth image. Considering a query assume that as a result of the retrieval, the k^{th} ground truth image for this query q is found at a position R. If this image is in the first K retrievals then $Rank(k) = R$ else $Rank(k) = K + 1$

The modified retrieval rank is:

$$MRR(q) = AVR(q) - 0.5 - 0.5 \times N(q) \tag{8}$$

The normalized modified retrieval rank is computed as follows:

$$NMRR(q) = \frac{MRR(q)}{K + 0.5 - 0.5 \times N(q)} \tag{9}$$

The average of NMRR over all queries defined as:

$$ANMRR(q) = \frac{1}{Q} \sum_{q=1}^{Q} NMRR(q) \tag{10}$$

Table 2 illustrates the ANMRR results in both databases. The results are compared with the corresponding results of the Color Layout Descriptor (CLD). The implementation of the CLD matches the implementation in (Lux and Chatzichristofis, 2008). The reason for this comparison is that only the CLD, from the descriptors that were referred in the introduction, is compact.

Considering the results, it is evident that the proposed descriptor is able to achieve satisfactory retrieval results by using Hand Drawn Color Sketches as queries. The proposed descriptor is implemented in the image retrieval system img(Rummager) (Chatzichristofis et al., 2009) and is available online[1] along with the image databases and the queries.

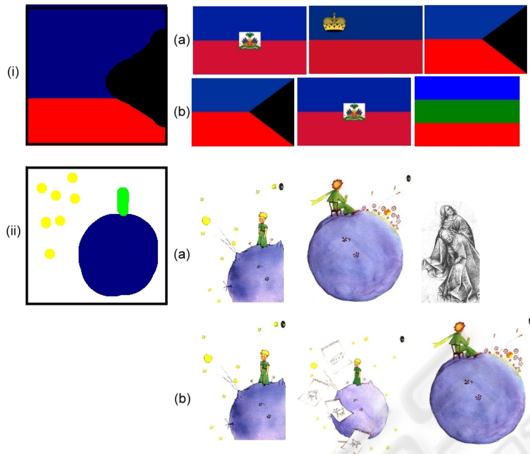

Figure 4: Hand Drawn Color Sketch queries (i) for the Flags database and (ii) for the Paintings database. In each one of the queries (a) shows the first 3 results of the CLD, while (b) shows the first 3 results of the proposed descriptor.

Table 2: ANMRR results.

Database	SpCD ANMRR	CLD ANMRR
Flags	0.225	0.425
Paintings	0.275	0.45

5 CONCLUSIONS

In this paper, a new compact composite descriptor is presented which combines color and spatial color distribution information.Characteristics of the proposed descriptor are its small storage requirements (48 bytes/ per image) and its small length (48 elements). Experiments that were performed at 2 databases with artificial images showed that the proposed descriptor can achieve better retrieval results than the MPEG-7 CLD.

REFERENCES

Chatzichristofis, S. and Boutalis, Y. (2007). A hybrid scheme for fast and accurate image retrieval based on color descriptors. In *IASTED International Conference on Artificial Intelligence and Soft Computing*.

Chatzichristofis, S. and Boutalis, Y. (2008a). Cedd: Color and edge directivity descriptor: A compact descriptor

[1] http://www.img-rummager.com

for image indexing and retrieval. volume 5008, page 312. Springer.

Chatzichristofis, S. and Boutalis, Y. (2008b). Fcth: Fuzzy color and texture histogram-a low level feature for accurate image retrieval. In *Image Analysis for Multimedia Interactive Services, 2008. WIAMIS'08. Ninth International Workshop on*, pages 191–196.

Chatzichristofis, S. and Boutalis, Y. (2009). Content based medical image indexing and retrieval using a fuzzy compact composite descriptor. In *Proceedings of the 6th IASTED International Conference*, volume 643, pages 1–6.

Chatzichristofis, S., Boutalis, Y., and Lux, M. (2009). Img(rummager): An interactive content based image retrieval system. pages 151–153. 2nd International Workshop on Similarity Search and Applications (SISAP).

Cinque, L., Ciocca, G., Levialdi, S., Pellicano, A., and Schettini, R. (2001). Color-based image retrieval using spatial-chromatic histograms. *Image and Vision Computing*, 19(13):979–986.

Datta, R., Joshi, D., Li, J., and Wang, J. (2008). Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys*, 40(2):160.

Huang, J. (1997). Image indexing using color correlograms. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, page 762768.

Jacobs, C., Finkelstein, A., and Salesin, D. (1995). Fast multiresolution image querying. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 277–286. ACM New York, NY, USA.

Kasutani, E. and Yamada, A. (2001). The mpeg-7 color layout descriptor: a compact image featuredescription for high-speed image/video segment retrieval. In *Image Processing, 2001. Proceedings. 2001 International Conference on*, volume 1.

Konstantinidis, K., Gasteratos, A., and Andreadis, I. (2005). Image retrieval based on fuzzy color histogram processing. *Optics Communications*, 248(4-6):375–386.

Lim, S. and Lu, G. (2003). Spatial statistics for content based image retrieval. In *Information Technology: Coding and Computing [Computers and Communications], 2003. Proceedings. ITCC 2003. International Conference on*, pages 155–159.

Lux, M. and Chatzichristofis, S. (2008). Lire: lucene image retrieval: an extensible java cbir library. ACM New York, NY, USA.

Manjunath, B., Ohm, J., Vasudevan, V., and Yamada, A. (2001). Color and texture descriptors. *IEEE Transactions on circuits and systems for video technology*, 11(6):703–715.

Pass, G., Zabih, R., and Miller, J. (1997). Comparing images using color coherence vectors. In *Proceedings of the fourth ACM international conference on Multimedia*, pages 65–73. ACM New York, NY, USA.

Rao, A., Srihari, R., and Zhang, Z. (1999). Spatial color histograms for content-based image retrieval. In *11th IEEE International Conference on Tools with Artificial Intelligence, 1999. Proceedings*, pages 183–186.

Sun, J., Zhang, X., Cui, J., and Zhou, L. (2006). Image retrieval based on color distribution entropy. *Pattern Recognition Letters*, 27(10):1122–1126.

Zagoris, K., Chatzichristofis, S., Papamrkos, N., and Boutalis, Y. (2009). Img(anaktisi): A web content based image retrieval system. pages 154–155. 2nd International Workshop on Similarity Search and Applications (SISAP) .

Zimmermann, H. (1987). *Fuzzy sets, decision making, and expert systems*. Kluwer Academic Pub.

APPENDIX

Table 3: Fuzzy Interface Rules.

If HUE is	And SATURATION is	And VALUE is	Then OUT is
RED 1	GRAY	WHITE	WHITE
RED 1	GRAY	BLACK	BLACK
RED 1	COLOR	WHITE	RED
RED 1	COLOR	BLACK	RED
YELLOW	GRAY	WHITE	WHITE
YELLOW	GRAY	BLACK	BLACK
YELLOW	COLOR	WHITE	YELLOW
YELLOW	COLOR	BLACK	YELLOW
GREEN	GRAY	WHITE	WHITE
GREEN	GRAY	BLACK	BLACK
GREEN	COLOR	WHITE	GREEN
GREEN	COLOR	BLACK	GREEN
CYAN	GRAY	WHITE	WHITE
CYAN	GRAY	BLACK	BLACK
CYAN	COLOR	WHITE	CYAN
CYAN	COLOR	BLACK	CYAN
BLUE	GRAY	WHITE	WHITE
BLUE	GRAY	BLACK	BLACK
BLUE	COLOR	WHITE	BLUE
BLUE	COLOR	BLACK	BLUE
MAGENTA	GRAY	WHITE	WHITE
MAGENTA	GRAY	BLACK	BLACK
MAGENTA	COLOR	WHITE	MAGENTA
MAGENTA	COLOR	BLACK	MAGENTA
RED 2	GRAY	WHITE	WHITE
RED 2	GRAY	BLACK	BLACK
RED 2	COLOR	WHITE	RED
RED 2	COLOR	BLACK	RED

Table 4: Fuzzy 8-bin Color System.

Membership Function	Activation Value			
	0	1	1	0
HUE				
RED 1	0	0	5	55
YELLOW	5	55	60	115
GREEN	60	115	120	175
CYAN	120	175	180	235
BLUE	180	235	240	295
MAGENTA	240	295	300	355
RED 2	300	355	360	360
SATURATION				
GRAY	0	0	25	55
COLOR	25	55	255	255
VALUE				
WHITE	0	0	100	156
BLACK	100	156	255	255