Developing the Color Temperature Histogram Method for Improving the Content-Based Image Retrieval

P. Phokharatkul, S. Chaisriya, S. Somkuarnpanit, S. Phaiboon, and C. Kimpan

Abstract**—**This paper proposes a new method for image searches and image indexing in databases with a color temperature histogram. The color temperature histogram can be used for performance improvement of content–based image retrieval by using a combination of color temperature and histogram. The color temperature histogram can be represented by a range of 46 colors. That is more than the color histogram and the dominant color temperature. Moreover, with our method the colors that have the same color temperature can be separated while the dominant color temperature can not. The results showed that the color temperature histogram retrieved an accurate image more often than the dominant color temperature method or color histogram method. This also took less time so the color temperature can be used for indexing and searching for images.

Keywords**—**Color temperature histogram, color temperature, an image retrieval and content-based image retrieval.

I. Introduction

WITH the expansion of digital image use, many researchers have been investigating increases in the efficiency of searching and indexing image data. Traditional text–based retrieval is not adequate for visual data. Consequently, the content–based approaches such as shape–based features, texture–based features and color–based features are proposed for searching and for indexing image data. Previous research shows that one of the most popular and easy to extract image features in the content–based approaches are color–based features. There are many methods of color-based extraction, such as color histogram, color correlogram and dominant color temperature. Many of the previous researchers used the color histogram method to represent the color composition of an image [1, 4, 5, 6]. The advantages of the histogram are that it is invariant for translation and rotation of the viewing axis. With this method, the histogram is changed very little when comparing the images taken, with little change of the angle of view [1]. However, histograms represent only primary colors which are red, green and blue. Other colors such as cyan, magenta, yellow can not be represented with a color histogram. When the colors are extracted, they are separated and counted into red, green and blue histograms. As a result, two images may have similar histograms although the perceptive human may feel that they are different colors. Other method, the dominant color temperature method is proposed by Karol Wnukowicz [2, 3]. It is used to index and search for images. The description represents color consistency with human perception.

However, the dominant color temperature can be represented only in 8 color temperatures. The disadvantages are that it cannot represent a color that is higher than 25000^o K, black or gray. Also it cannot separate colors that have the same color temperature so it is not sufficient to represent the visual color.

In this paper, the color temperature histogram is proposed to solve the problem. Because it can represent 46 color temperature ranges from 1667^o–100000^{o}K, a color with the same color temperature value can be separated. The result shows that the color temperature gives a higher percentage of accurately retrieved images than the color histogram or the dominant color temperature.

The paper is organized as follow. First, introduction. In section II, literature reviews are discussed. Section III presents Implementation. Section IV presents our experiment and results. Finally, the conclusions.

P. Phokharatkul is with the Department of Computer Engineering, Faculty of Engineering, Mahidol University, Nakhon Pathom 73170, Thailand (e-mail: egpph@mahidol.ac.th).

S. Chaisirya is with the Master of Science Program in the Technology of Information System Managements, the Faculty of graduate studies,Mahidol University, Nakhon Pathom 73170, Thailand (e-mail: jitty25@hotmail.com).

S. Phaiboon is with the Department of Electronic Engineering, Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bankok, Thailand (e-mail: egspb@mahidol.ac.th).

S. Somkuarnpanit is with the Department of Electronic Engineering, Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bankok, Thailand (e-mail: kssuripo@kmitl.ac.th).

C. Kimpan is with the Department of Informatics Faculty of Information Technology, Rangsit University, Pathum Thani, Thailand.

II. Literature Reviews

A. The Color Histogram [4, 5, 6]

A color histogram can be used to represent the color compositions of an image. The advantages of the histogram are that it is invariant for translation and rotation of the viewing axis. With this method, the histogram is changed very little when comparing the images taken, with little change of the angle of view. However, histograms represent only primary colors which are red, green and blue. Other colors such as cyan, magenta, yellow can not be represented with a color histogram. When the colors are extracted, they are separated and counted into red, green and blue histograms. As a result, two images may have similar histograms although the perceptive human may feel that they are different colors.

B. The Color Temperature Histogram

The color temperature is used to describe a certain property of the light sources [7]. It is calculated from Plank's formula. The formula is based on the relationship between the temperature of a black body radiator and the color of its emitted light after receiving higher temperatures [8]. It is used to show the color appearance of the light by comparison with the color of a black body radiator. When the color of the light matches the black-body radiator, then the light has the same color temperature value as the black–body radiator. However, many radiators are not exactly equal to any of the colors of a black–body radiator. So, a correlated color temperature concept is to propose the color temperature calculation of these radiators. The correlated color temperature was calculated by Robertson's algorithm [7] and used the perceptually homogeneous UV chromaticity diagram (CIE 1960) [9].

Practically, the correlated color temperature is a long expression so many researchers and this research use the term color temperature instead.

C. Estimation Color Temperature

An algorithm for estimation of the color temperature of images is proposed by K. Wnukowicz and W. Skarbek [9]. It has many steps as follows:

1) Linearize image pixels from RGB to sRGB form that suits PC computer monitors.
2) Convert color space from sRGB to XYZ. Because the color temperature calculation must be performed on the perceptually homogeneous UV chromaticity diagram (CIE 1960 : Commission Internationale de I'Eclairages 1960) (Fig1: A,B)
3) Discard black and near black pixels, by checking the Y luminance value. If it is a lower threshold T_i, it is discarded, because black and near black pixels do not significantly impact color temperature perception.
4) Average pixels remain from the discard process. The output of this process is in the form of Xa, Ya and Za. They are used for calculation (x_s, y_s) of a chromaticity value. The (x_s, y_s) chromaticity value is a coordinate on the CIE (1931) chromaticity diagram.
5) Xa, Ya and Za are calculated to (x_s, y_s) a chromaticity value on the CIE (1931) chromaticity diagram
6) Convert (x_s, y_s) to (u_s, v_s) coordinates, from the CIE (1931) chromaticity diagram to CIE (1960) chromaticity diagram, for more accurate differences between values.
7) Find two adjacent isotemperature lines (Fig1: C) that are neighbors to the point (u_s, v_s) on the CIE (1960) chromaticity diagram and calculate the color temperature value by Robertson's method.
8) The output from the previous process is the color temperature in Mega Kelvin (MK^{-1}). To convert a color temperature in the Kelvin scale, 1000000 is divided by the color temperature value into Mega Kelvin.

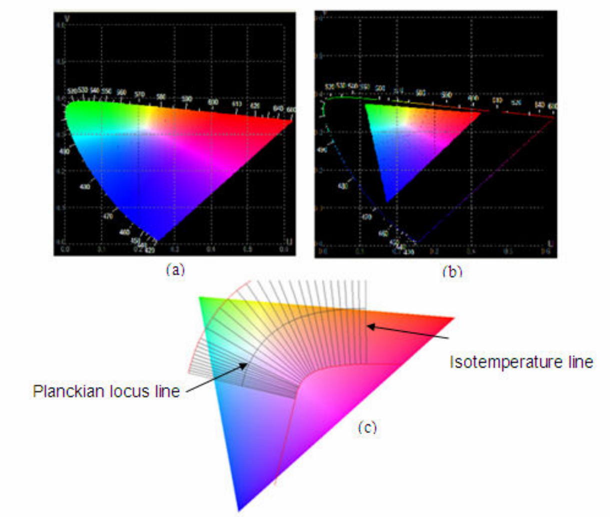

Fig.1 (a) = CIE (1960) diagram, (b) = Show gamut only, (c) = Planckian locus line and isotemperature line

D. Approximate Colors on CIE Chromaticity Diagram

Brand Fortner [10] proposed the boundaries and the names for colors on the CIE chromaticity diagram. It can be adapted to assign the approximated colors shown in Fig. 2.

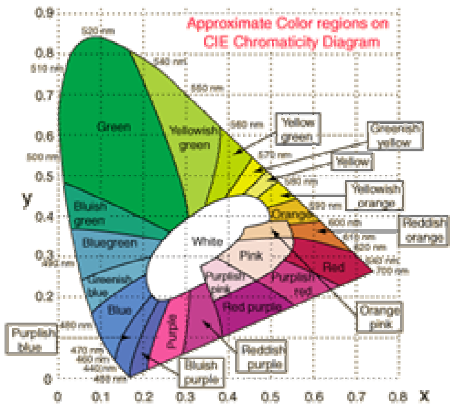

Fig. 2 Approximate colors on CIE chromaticity diagram [10]

From Fig. 2, Photo Research, Inc. proposed RGB values for the color regions of the CIE chromaticity diagram. The RGB values that have middle hue and saturated ranges for each region in the diagram are chosen. The values in each region are shown in Table I.

International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:1, No:8, 2007

TABLE I
RGB VALUES FOR THE COLOR REGIONS OF
CIE CHROMATICITY DIAGRAM [10]

Color name	**Red**	**Green**	**Blue**
Red	191	27	75
Pink	245	220	208
Reddish orange	216	119	51
Orange pink	240	204	162
Orange	228	184	29
Yellowish orange	231	224	0
Yellow	234	231	94
Greenish yellow	235	233	0
Yellow green	185	214	4
Yellowish green	170	209	60
Green	0	163	71
Bluish green	24	162	121
Blue green	95	164	190
Greenish blue	110	175	199
Blue	92	138	202
Purplish blue	88	121	191
Bluish purple	92	102	177
Purple	246	85	158
Reddish purple	196	64	143
Purplish pink	243	208	219
Red purple	175	35	132
Purplish red	209	65	136
White	255	255	255

E. The Color Temperature Browsing Description [2, 3]

The MPEG–7 visual standards specified the dominant color temperature description (DCD). It is a description for representative colors in an image or image region. The color representation is more accurate and compact than color histogram–based descriptions. However, the disadvantage is that it does not enable queries by example searching.

F. The Dominant Color Temperature Distribution [2, 3]

The dominant color temperature distribution was proposed by K. Wnukowiz. It is used as a feature in query by example applications. The description contains up to 8 pairs of values: color temperatures of dominant colors and their percentage contents in an image. However, its description does not sufficiently identify all characterizations of color in an image.

III. IMPLEMENTATION

A. Creation of the Color Temperature Histogram

The color temperature histogram is a combination of the color temperature method and histogram method for descriptive color in an image. Its description approaches the human visual perception. A color temperature histogram consists of 46 bins. Each bin represents an interval of color temperature values. The color temperature values are calculated and derived from approximated color regions [10]. Each bin of the histogram shows a frequency of the same color temperature segments from an image.

TABLE II
INTERVAL OF THE COLOR TEMPERATURE VALUES USED
TO CREATE BINS

Approximate color	**Range**	**Parameter**
Black	β (R, G, B = 0)	3
Red	X= 1667	1
Red – Purplish-red	1667 > X < 1721	0,1
Purplish-red - Red-purple	1721<= X< 1744	0,1
Red-purple - Purple	1744<= X< 1818	0,1
Purple - Reddish-orange	1818<= X< 2089	0,1
Reddish-orange - Orange	2089<= X< 3159	0,1
Orange - Yellowish-orange	3159<= X< 3813	0,1
Yellowish-orange - Greenish-yellow	3813<= X< 3905	0,1
Greenish-yellow - Yellow	3905<= X< 4127	0,1
Yellow - Orange-pink	4127<= X< 4212	0,1
Orange-pink - Yellow-green	4212<= X< 4386	0,1
Yellow-green - Yellowish-green	4386<= X< 4699	0,1
Yellowish-green - Pink	4699<= X< 5239	0,1
pink - Purplish-Pink	5239<= X< 5541	0,1
Purplish-pink - White	5541<= X< 6503	0,1
White	X = 6503	3
Gray	X= 6503	1
Black	X= 6503 (R, G, B != 0)	0
White - Green	6503< X< 6710	0,1
Green - Bluish-green	6710<= X< 8406	0,1
Bluish-green - Purplish-blue	8406<= X< 10287	0,1
Purplish-blue - Greenish-blue	10287<= X< 14434	0,1
Greenish-blue - Blue-green	14434<= X< 16504	0,1
Blue-green - Blue	16504<= X< 100000	0,1
Blue	X= 100000	0

The color temperature histogram is created from the following steps (Fig. 3):

1) Segment an image into blocks because an image has various color temperatures. A block has a size of 4 x 4 pixels. The size gives higher accurate retrieved results and lowers the computation cost as shown in Table III in the experimental and result section.
2) Calculate the color temperature of each segment using color temperature estimation algorithm (section 2.3). This method can not used with a black color image because R, G, B value is 0. So we defined TR_c = β (β value out of 0–600) and parameter =3 for a segment that is black. For a segment with equal R G, and B value, we add a parameter

(γ) to separate the degree of gray, white and black. Parameters are 0=black, 1=gray and 2=white.

3) However, the color temperature is not sufficient to describe color as perceived by humans. All colors are represented by a color temperature value. They are calculated from the same isotemperature line. Then, we add a parameter (χ) that shows the color temperature located over or lower than a Planckian locus line. Parameters are 1=over a Planckian locus line, 0=lower than a Planckian locus line.
4) Create the color temperature histogram by counting the segments that have the same color temperature value and parameter χ then put it in the bin. The range of a bin is shown in Table II.

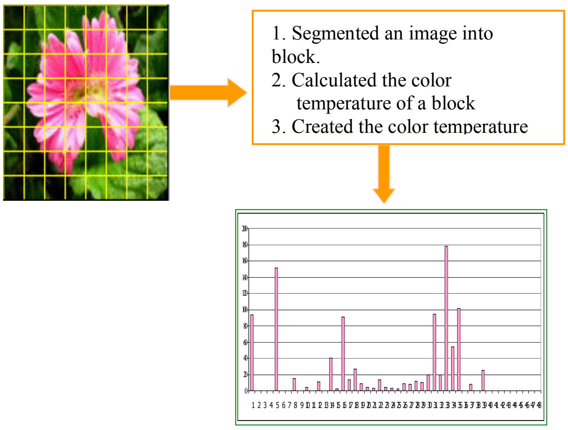

Fig. 3 The process of color temperature creation

B. Clustering of Image

From the database when an image is changed into color temperature histogram vectors, the next step is to use fixed threshold clustering for the images cluster. Fixed threshold clustering is a division of a hierarchical technique for clustering. This method begins with one large cluster and splits into smaller cluster items that are most dissimilar by comparing distance values between the mean of the group and an image. If the distance value is lower than the threshold limit then the image was labeled as a name of mean group. If it is higher than the threshold then the image was assigned as "UnKnown" and waiting to find a suitable group in the next round.

The algorithm of fixed threshold clustering is shown as follows:

1. All images have NameU_i as "UnKnown".
2. The algorithm has to split up the cluster into smaller clusters by

   Loop

   2.1. Set name of cluster (NameC_j)

   2.2. For i=1 to i = n (n = set Ui)

   2.2.1 An image (U_i) is selected for the mean of the cluster (C_j) by a random method.

   2.2.2 Compute distance (D) between mean of C_j and U_i by the City block method.

   2.2.3 (D) <= threshold

   Then

   - NameU_i = NameC_j
   - Compute new mean between mean of C_j and U_i
   - Save NameUi

   2.2.4 If (i=n)

   Then Save mean C_j and NameC_j

   End loop

   when all images do not have NameU_i = "UnKnown"

Where U_i = an image i. , C_j = a cluster j, NameU_i = a cluster name of image i., NameC_j = a name of cluster j and D = distance value between U_i and C_j

A result of the clustering process is a list of mean values of image groups and an identified image group name of each image. Then they will be stored in a database.

C. Searching

To search for images in the system, the sampled images were used to query a set of in-condition images from the database. To search for images from the database the steps below are used:

1) Create a color temperature histogram of the sampling image by the method in section 3.1
2) Find distance between a color temperature histogram of the sampled image and means of color temperature histogram using the cosine method.
3) Sort the distance in ascending in order to choose the one-five means.
4) Each mean value from step 3 will be used to select a group and retrieve images from that group.
5) Find distance between a color temperature histogram of the sampled image and color temperature histogram of images retrieved from step 4 by using the city block distance method.
6) Sort the distance in ascending order.
7) Present the top-twelve images on the monitor as a result.

IV. Experiments and Result

A. Experiments

We used a database of 1752 color JPEG images for all our experiments and used 146 color JPEG images for testing the system. The images are 128 x 128 pixels in size and in many different classes, such as flowers, buildings, natural, etc.

After, we extract a variety of images into color temperature histograms as described in section 3.1. Using the fixed threshold clustering method, we cluster all images into suitable groups and store them in a database as described in section 3.2 and using the method in section 3.3 to search the images in a database.

B. Result of the Experiment for Image Retrieval

The experimental research was concerned with the accuracy of the image retrieval, shown in figures 4, 5, 6 as examples of the result. Figure7 is shown as an example of a result that is not accurately retrieved from our system. The larger image on the left is the query image, and the 12 images on the right are the image results. The suitable size of segments is shown in Table III. In Table IV we show the comparison in percent of accurately retrieved images and the time for extraction per 1752 between method, color histogram and dominant color temperature.

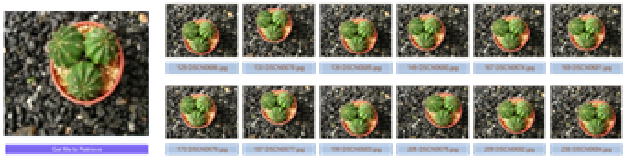

Fig. 4 Accurately retrieval result in our method

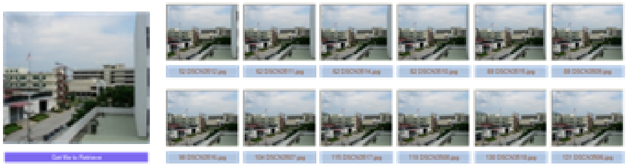

Fig. 5 Accurately retrieval result in our method

Fig. 6 Accurately retrieval result in our method

Fig. 7 Inaccurately retrieval result in our method

TABLE III
SUITABLE SIZE OF SEGMENT IN OUR METHOD

Size	Image error	Time for Extract (for 1752 images)
1 x 1	0.399543%	02:52:54.8593750
2 x 2	0.285388%	00:11:53.7812500
4 x 4	0.114155%	00:03:20.8593750
8 x 8	1.826484%	00:02:04.6875000

TABLE IV
SHOW COMPARISON IN PERCENT OF ACCURATELY RETRIEVED IMAGES AND TIME FOR EXTRACTION BETWEEN METHOD, COLOR HISTOGRAM AND DOMINANT COLOR TEMPERATURE

Method	% accurate	Time for Extract
Color temperature histogram [Proposed method]	99.885845	00:03:20.8593750
Dominant color temperature [2, 3]	72.1274175	00:04.09.8943000
Color histogram	56.93970421	00:03.13.5630000

V. CONCLUSIONS

In this paper the color temperature histogram is proposed for indexing and searching for images in a database. The results show that the percentage of accurately retrieved images by the color temperature histogram method is higher than using color histogram or dominant color temperature, as the color temperature histogram can represent more colors than the color histogram. It can represent 46 color ranges while the color histogram can represent only red, green and blue. Other colors such as pink, yellow, orange, cyan, etc. were separated into red, green, and blue color and counted into red, green and blue histograms. For this reason, two images may have the same histogram although they have different colors. So it is not a good representativeness for images.

The experiment result shows that the color temperature histogram gives a higher percentage of accurately retrieved images than the dominant color temperature as it can represent the color temperature ranges higher than the dominant color temperature. It represents 46 color ranges from 1667-100000^o K and represents black, gray images. Likewise it can separate different colors that have the same color temperature. While the dominant color temperature represents 8 color temperature values from 1667-25000^o K, it cannot represent a color that is higher than 25000^o K, black or gray. Also it cannot separate colors that have the same color temperature. Otherwise, from Table III, experiments and results show the color temperature histogram uses less time for extracting images than the dominant color temperature. So the color temperature histogram is a good feature vector for image retrieval.

REFERENCES

[1] Oge Marques, Borko Furht. Content-based image and video retrieval. U.S.A.: Kluwer Academic Publishers; 2002.
[2] Karol Wnukowicz. Image Indexing by Distributed Color Temperature Descriptions. Fundamenta Informaticae. 2004; 1-10.
[3] K. Wnukowicz. Data dictionary for indexing of dominant color temperature descriptions based on M-tree. Fundamenta Informaticae. 2004; 1-10.
[4] Chuping Liu, Julien Lamoureux, Yuxin Wang, Yunan Xiang. Histogram algorithm for image indexing. http://www.ee.ualberta.ca.html [Accessed 2004 Sep 20].
[5] Ju Han; Kai-Kuang Ma. Fuzzy color histogram and its use in color image retrieval. Image Processing. [online] 2002 Aug; 11(8): 944-952. Abstract from: IEEE Transactions on Image Processing. http://ieeexplore.ieee.org/search/searchresult.jsp

International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:1, No:8, 2007

[6] Seong-O Shim, Tae-Sun Choi. Edge color histogram for image retrieval.Image Processing 2002;May:957-960. Abstract from: Proceedings 2002 International Conference on Image Processing.
[7] Gunther Wyszecki and W. S. Stiles. Color science concepts and method quantitative data and formulae. U.S.A.: A Wiley-Interscience Publication. 2000: 224-229.
[8] Molecular Expressions Microscopy Primer. Light and Color – Color Temperature. http://www.molecularexpressions.com/primer/lightandcolor/colortemperatureintro.html [Accessed 2005 Jan 20].
[9] K. Wnukowicz, W. Skarbek. Color temperature estimation algorithm for digital images properties and convergence. Opto-electronics Review. 2003; 11(3): 193-196.
[10] The C.I.E chromaticity. http://hyperphysics.phy-astr.gsu.edu/hbase/vision/cie.html [Accessed 2005 Jan 20].

Pisit Phokharatkul received the Doctoral of Engineering in Electrical and Computer Engineering (2002) from King Mongkut's Institute of Technology Ladkrabang, M.Eng (Electrical Engineering) in 1990 from King Mongkut's Institute of Technology North Bangkok, M.Eng (Nuclear Technology) in 1985 from Chulalongkorn Univerity, and B.Ed in Physics from Sri Nakharinwirot University in 1981, respectively. Now he is an assistant professor of Electrical and Computer Engineering and Academics Infrastructures Deputy Dean at the Faculty of Engineering, Mahidol University. His interests include pattern recognition, image retrieval, fuzzy logic application, and swarm intelligence.

Somjit Chaisriya received B.Inf.Sc. in information Study, Faculty of Information Technology, Walailak University Thailand. Now she is studying for a Masters Degree in Technology of Information System Management, Faculty of Graduate Studies, Mahidol University, Thailand. Her interests include image processing, image retrieval, efficient intelligence and data warehouse.

Suripon Somkuarnpanit received the Ph.D. degree in Optical Electronics and lazer engineering from Heriot-Walt University UK, M.Sc. in Medical Electronics and Physics from Bartholomew's Hospital Medical College, University of London, UK., and B.Eng. in Electrical Engineering from King Mongkut's Institute of Technology Ladkrabang, Thailand. Now he is an associate professor in Electronics at Faculty of Engineering, King Mongkut's Institute of Technology, Ladkrabang. His interests include opto electronics, lazer and lazer applications, pattern recognition, image retrieval and medical instrumentation. He is a member of the IEEE.

Supachai Phaiboon received the B. Eng. and M. Eng degree in Electrical Engineering from King Mongkut's Institute of Technology North Bangkok, Bangkok, Thailand in 1987 and 1990, respectively. He is an associate professor in electrical at Faculty Engineering Mahidol University. His research interests include radio propagation and measurement, computational intelligence, pattern recognition, speech processing and image retrieval.

Chom Kimpan received D.Eng. in Electrical and Computer Engineering from King Mongkut's Institute of Technology Ladkrabang, M.Sc. in Electrical Engineering from Nihon University, Japan, and B.Eng. in Electrical and Engineering from King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand. Now he is an associate professor at the Department of Informatics, Faculty of Information Technology, Rangsit University, Thailand. His interests are in Pattern recognition, image retrieval, speech recognition and swarm intelligence.