Available online at www.ijpe-online.com

International Journal of Performability Engineering

Vol. 13, No. 6, October 2017, pp. 823-831
DOI: 10.23940/ijpe.17.06.p4.823831

Retrieval of Vehicle Images based on Color Space Fuzzy Quantification in Criminal Investigation

Mingdi Hu[a,*], Mengbin Zhang[a], Yilun Lou[b]

[a]*School of Communications and Information Engineering, Xi'an University of Posts & Telecommunications, Shaanxi, Xi'an ,710119, China*
[b]*School of Mathmatics and Information Science, Shaanxi Normal University, Shaanxi, Xi'an ,710119, China*

Abstract

The color and shape features provide more contributions for vehicle images in criminal investigation generally. A criminal investigation library that was used in studying criminal tools was set up, in which the number of the car images is more than one thousand. In this paper, the color of the car pictures was quantized by the non-uniform quantization, the triangular and the trapezoidal fuzzy membership degree functions respectively at the first step, and then the Euclidean distance and the weighted distance similarity measures methods are used at the second step, the last step is that the car images were retrieved by the above six algorithms in the criminal investigation library. The results show that the weighted distance similarity measure algorithm based trapezoidal membership degree is better than others; meanwhile the precision and recall are significantly higher than other else. The triangular fuzzy quantization algorithm is not very different from the non-uniform quantization algorithm, and even the precision and recall ratio is smaller than the latter. Under the same quantization condition, the weighted distance algorithm is slightly better than the Euclidean distance, but the difference is not great.

Keyworlds: criminal investigation; fuzzy membership degree function; euclidean distance; weighted distance; vehicle images retrieval

(Submitted on July 25, 2017; Revised on August 30, 2017; Accepted on September 15, 2017)

(This paper was presented at the Third International Symposium on System and Software Reliability.)

1. Introduction

Image information becomes the most important in a variety of information resources because of huge image database and visual [7]. Everyone can get more and more image information every day with the rapid development of communication and network technology. The approaches of capture useful images from in the huge image database become serious, so it is urgent to find a new way to retrieve images accurately [13]. Nowadays images can be retrieved in three ways: text-based, content-based, and semantic-based [3,9,11]. The Text-based retrieval approach has two drawbacks: one is time-consuming because that images need text annotation manually; another is that results of retrieval are inaccurate because of the man-made understanding of query images [5].

The content based image retrieval (CBIR) is to retrieve images through low-level features like color, texture and shape that can represent an image. The method is to extract the low-level features on the example image firstly, to compute the similarity between the query image and the images in image database secondly, to sort images by similarity measure lastly, and then the top images will be displayed. CBIR is more effective and subjective than text-based apparently, but fails to describe semantic concepts elaborately, so many researches proposed some methods in order to reduce the semantic gap between content-based and semantic-based [4], one of these approaches is fuzzy quantization based that is closer to human perception [2,8,14]. The fuzzy theory is the basis of artificial intelligence, so color space can be quantized by fuzzy membership degree function, which is more coordinately with human perception. A color-based non-uniform quantization of the Euclidean distance algorithm is proposed [14]. In [8], a sub-fuzzy color histogram construction method is proposed. It aggregates the color images by the fuzzy C-means clustering algorithm and constructs a approximate triangle membership function. In [2], a novel image retrieval method is proposed by trapezoid membership degree function which extract the color features by quantizing the edge in HSV (hue,saturation and value). All above algorithms have been designed to extract

* Corresponding author.
E-mail address: humingdi2005@xupt.edu.cn

features for image retrieval and have their advantages respectively as well.

In this paper, a criminal investigation library that was used in studying criminal tools was set up, in which the number of the car images is more than one thousand. In this paper, the color of the car pictures was quantized by the triangular and the trapezoidal fuzzy membership degree functions respectively, and then the Euclidean distance and the weighted distance similarity measures methods are used to computer the similarity degree underlying with membership function, the last step is that the car images were retrieved by the above six algorithms in the criminal investigation library. The results show that the weighted distance similarity measure algorithm based trapezoidal membership degree is better than others; meanwhile the precision and recall are significantly higher than other else. The triangular fuzzy quantization algorithm is not very different from the non-uniform quantization algorithm, and even the precision and recall ratio is smaller than the latter. Under the same quantization condition, the weighted distance algorithm is slightly better than the Euclidean distance, but the difference is not great. The innovation of this paper lies that the criminal investigation car image library have been set up and comparison and analysis among all six algorithms are clarified.

In the rest sections, we give the basic knowledge firstly. In the second section, criminal investigation car image library is introduced briefly. Then, we describe six algorithms by collocating non-uniform quantification, triangular membership, trapezoidal membership with the Euclidean distance and the weighted distance. The analysis of several simulation results are shown in the, last section is the conclusion of the paper.

2. Basic knowledge

2.1. HSV model [6]

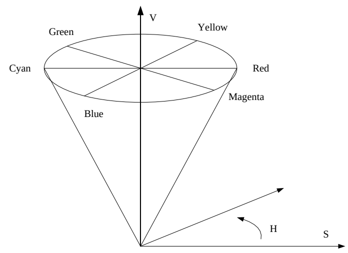

Figure 1. HSV model

The HSV color model can be represented by an inverted cone. The long axis represents the brightness, the distance from the long axis to original point represents saturation, the angle around the long axis is defined as hue, that is, the HSV model is shown in Figure 1.

HSV color space is widely used in color feature extracting. In this space, hue is used to distinguish color, saturation is the percentage of white light add to a pure color and value refers to the perceived light intensity. The advantage of HSV color space is that it is closer to human conceptual understanding of colors.

2.2. Image similarity calculation [10]

Assume that the histogram feature vector of example image is $(q_0, q_1, ..., q_L)$, and other histogram feature vector in the image database is $(s_0, s_1, ..., s_L)$.

(1) Euclidean distance

$$D(Q,S) = \sqrt{\sum_{i=0}^{L-1} (q_i - s_i)^2} \quad (1)$$

Where L is the dimension of the histogram vector.

(2) Weighted distance

$$D(Q,S) = \sqrt{\sum_{i=0}^{L-1} \sigma (q_i - s_i)^2}, \quad \sigma = \begin{cases} q_i & if \quad q_i, s_i > 0 \\ 1 & if \quad q_i \cup s_i = 0 \end{cases} \quad (2)$$

Where σ is the weighted factor.

2.3. Fuzzy sets

Definition2.3.1[12] Let X be the universe and μ_A be a function from X to [0,1] . The function μ_A defined by:

$$\begin{gathered}\mu_A : X \rightarrow [0,1] \\ \forall x \in X,\ x \rightarrow \mu_A(x)\end{gathered} \tag{3}$$

is called a fuzzy subset. Where, a real number $\mu_A(x) \in [0,1]$ corresponding to any $x \in X$ represents the membership degree of x belong to X , μ_A is called the membership degree function of *A*.

3. The Establishment of Image Database of Criminal Investigation

Figure 2. Picture preprocessing

There are Volkswagen, Toyota, Ford, modern and other 20 kinds images of automobile brand in the criminal investigation vehicle image library, and there are 50 vehicles each brand car. According to the different color among all cars, they are divided into 8 categories. Each category includes color lattices 125 images with 96*64 type JPG.

The process of the car image library establishment:

(1) Image acquisition. In the car home website, we collect the car pictures and sort these images in accordance with the car's brand and car color.

(2) To reduce the image pixels. The image pixels in its home website is too high, which would lead to longer time retrieval in one experiment. On the other hands, the retrieval results are still effectively although the pixels of them are reduced appropriately.

(3) Image segmentation. Every car is located in the middle of the whole image, so the background has undermined the retrieval effectuation. Before retrieval, we need to pretreat the car images, meaning to cut the picture boundary according to a certain percentage, as shown in Figure 2.

4. Non-uniform quantization image retrieval

In general, the acquired images are described in RGB space, so the image needs to be transformed from the RGB space model to the HSV space model in practical experience. After converting to HSV color space, the range of HSV components: h∈[0, 360], s∈[0, 1], v∈[0, 1]. In order not to be confused, the h, s, and v are defined as the continuous values of the RGB space color converted to the HSV space, while H, S, V are the discrete value of h, s, v after quantization. The three components of H, S, and V are quantized by unequal interval according to the human's colors perception:

(1) According to the human visual resolving power, in this paper, chroma H is quantified into 8 parts while saturation S and the luminance V are quantized into 3 parts.

(2) Quantification based on different ranges colors and subjective color perception:

$$H = \begin{cases} 0 & if \quad h \in (315,20] \\ 1 & if \quad h \in (20,40] \\ 2 & if \quad h \in (40,75] \\ 3 & if \quad h \in (75,155] \\ 4 & if \quad h \in (155,190] \\ 5 & if \quad h \in (190,270] \\ 6 & if \quad h \in (270,295] \\ 7 & if \quad h \in (295,315] \end{cases} \qquad S = \begin{cases} 0 & if \quad s \in [0,0.2] \\ 1 & if \quad s \in [0.2,0.7] \\ 2 & if \quad s \in [0.7,1] \end{cases} \qquad V = \begin{cases} 0 & if \quad v \in [0,0.2] \\ 1 & if \quad v \in [0.2,0.7] \\ 2 & if \quad v \in [0.7,1] \end{cases} \tag{4}$$

According to the above quantization level, the color components are combined into one-dimensional feature vector:

$$G = HQ_s Q_v + SQ_v + V \tag{5}$$

Among them, Q_s and Q_v are the quantization series of components S and V, $Q_s = 3$, $Q_v = 3$.
Equation (5) is actually:

$$G = 9H + 3S + V \tag{6}$$

Thus, the three components of *H*, *S*, and *V* are distributed over a one-dimensional vector. The range of values for *G* is [0,1,..., 71], and calculating *G* to obtain 72 bin one-dimensional histogram.

Suppose $(g_0, g_1 ..., g_i ..., g_{71})$ is the normalized histogram of the car image to be retrieved, and $(g_0^Q, g_1^Q \ldots g_i^Q \ldots g_{71}^Q)$ is the normalized histogram of the image Q in the image library.

(1) Using Euclidean distance calculates similarity

$$D_E = \sqrt{\sum_{i=0}^{71} (g_i - g_i^Q)^2} \tag{7}$$

Three different colors of the cars were retrieved experiments under the method, according to (7) calculating the smallest q of the first 20 car pictures, the retrieve results shown in Figure 3:

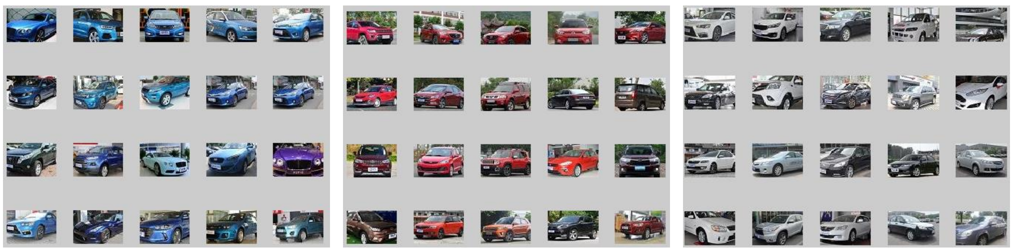

Figure 3.Three colors of cars in non - uniform quantitative European distance retrieval

(2) Using weighted distance calculates Similarity

$$D_W = \sqrt{\sum_{i=0}^{71} \omega (g_i - g_i^Q)^2} \qquad \omega = \begin{cases} g_i & if \quad g_i, g_i^Q > 0 \\ 1 & if \quad g_i \cup g_i^Q = 0 \end{cases} \tag{8}$$

As a result of the normalized histogram, $\omega \le 1$, the weighted distance is less than the Euclidean distance.

Three different colors of the cars were retrieved experiments under the method, according to (7) calculating the top 20 car pictures with the smallest of D, the retrieve results shown in Figure 4:

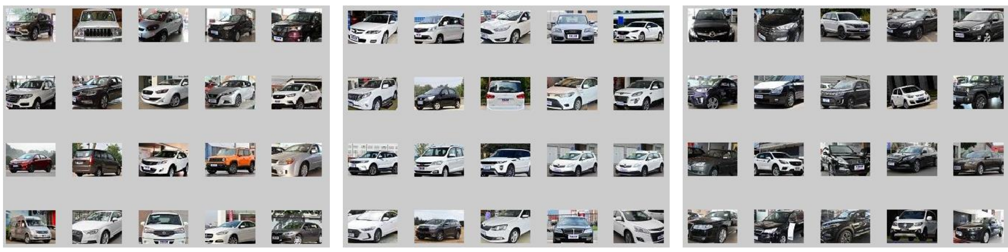

Figure 4. Three color vehicles in non - uniform quantization weighted distance histogram retrieval

5. Triangle membership degress algorithm

In general, the chroma H in the HSV color space is the most important distinguishing visual feature, and the colors difference is mainly described by the difference in color chromaticity. Therefore, this article only quantifies the h component more finely, while s and v quantification method are unchanged.

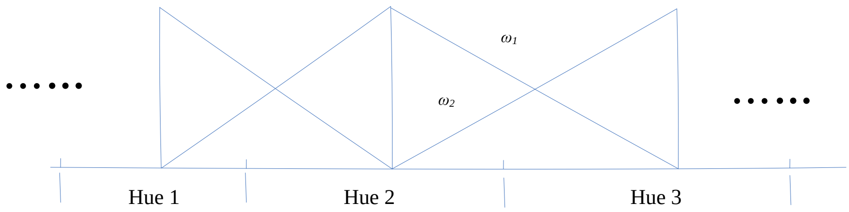

Figure 5. Triangle membership function model of H

We take the red as an example. Before the quantification, only one color really belongs to the red. In Figure 5, the maximum value of Hue 2 is red. The farther away from this point, belonging to the red membership will be smaller and becomes 0 until the transition to the maximum value of the next Hue. It can be seen that the pixels between the maximum of Hue 1 and the maximum of Hue 2 are only 0 in the two chromaticity spaces and 0 in the other six chromaticities. In this quantification method, we define it as the triangular membership algorithm for H.

The simulation steps are as follows:

Step 1. Image chroma H non-uniform quantification. The specific method is shown above.

Step 2. Let h_1 , h_2 , h_3 , h_4 , h_5 , h_6 , h_7 , h_8 are the quantization borders in the non-uniform quantization method (corresponding to (4): $h_1 = 315$, $h_2 = 20$, $h_3 = 40$, $h_4 = 75$, $h_5 = 155$, $h_6 = 190$, $h_7 = 270$, $h_8 = 295$), and the middle value of each chromaticity interval is the maximum value of this chromaticity space, ie $(h_2 - h_1)/2$, $(h_3 - h_2)/2$ ......,The degree of chromaticity membership between the maximum values of each chromaticity space is the intersection of two triangles (Figure 5), ie ω_1, ω_2 . The chromaticity in this interval is expressed as:

$$H = \omega_1 \bullet 0 + \omega_2 \bullet 7 \tag{9}$$

Where ω_1, ω_2 are weights, which indicate the degree to which the chromaticity of the two chromaticityes contribute to the chromaticity of the neighborhood. The weight is the slope of the triangle as shown in Figure 5.

Step 3. According to the formula (6) *H, S, V* three components in a one-dimensional vector distribution, calculate the fuzzy histogram.

Step 4. the image library all the pictures are in accordance with the above three steps to calculate the fuzzy histogram, stored in the mat file.

(1) Using Euclidean distance calculates similarity

Using the Euclidean distance algorithm to search the three kinds of car criminal investigation pictures randomly selected, according to (7) calculating the smallest of the top 20 car pictures, the search results as shown in Figure 6:

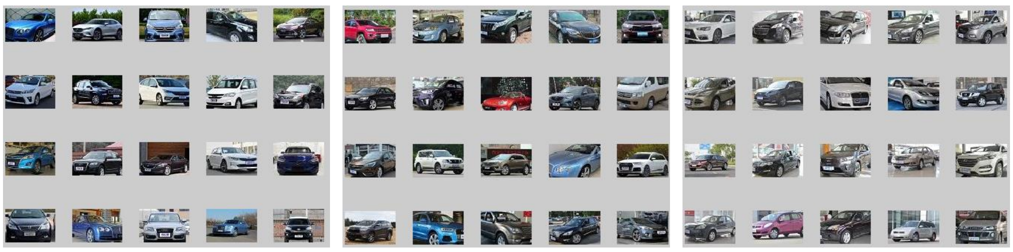

Figure 6. Retrieval of Euclidean Distance of Triangular Membership Function

(2) Using weighted distance calculates Similarity

Using the weighted distance algorithm to randomly select the three kinds of car criminal investigation pictures to retrieve the experiment, according to (8) calculating the smallest distance of the first 20 car pictures, the search results shown in Figure 7:

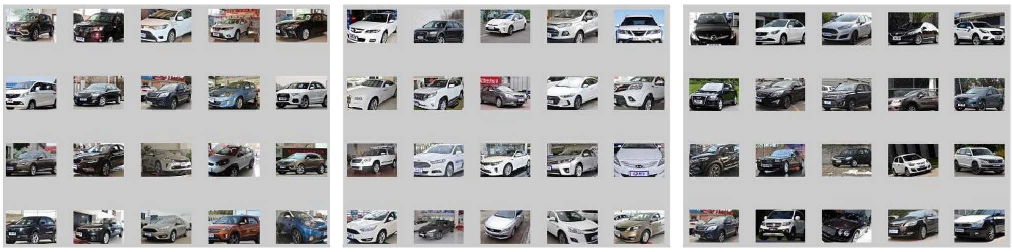

Figure 7. Retrieval of Weighted Distance of Triangular Membership Function

6. Trapezoidal membership degree algorithm

The hue is usually not truncated at the junction of the two hue intervals, but there is a small tonal transition interval. From the visual characteristics of the human eye, the transition from one hue to another in this interval is gradual, that is, the transition process is a weakening of the hue effect and the gradual enhancement of the other hue effects The For example, in the transitional range of orange and yellow tones, the effect of orange on the orange side is greater than that of yellow, and the yellow influence on the yellow side is definitely greater than orange. Thus, this small neighborhood, which exists at the junction of every two hue intervals, is a region that needs to be quantitatively quantified. For this model (Figure 8), we define it as the trapezoidal membership function model for H.

The simulation steps are as follows:

Step 1. Image H non-uniform quantification. The specific method is similar to the above.

Step 2. In Figure 9, suppose that $h_1, h_2, h_3, h_4, h_5, h_6, h_7, h_8$ are the quantization borders in the non-uniform quantization method, respectively (in the corresponding formula (4): $h_1 = 315$, $h_2 = 20$, $h_3 = 40$, $h_4 = 75$, $h_5 = 155$, $h_6 = 190$, $h_7 = 270$, $h_8 = 295$), $\varepsilon_1, \varepsilon_8$ are the small neighbors area. In the experimental project found that when the size of each neighborhood is equal, the retrieval effect is not ideal. In this article, $\varepsilon_1 = (h_2 - h_1)/k$. Similarly, $\varepsilon_8 = (h_1 - h_7)/k$, k is a constant. Because h is non-uniform quantization, so in general $\varepsilon_1 \neq \varepsilon_8$.

When $h \in (h_1 + \varepsilon_1, h_2 - \varepsilon_1]$, h does not fall on the quantization boundary, quantify the same, $H = 0$;

When $h \in (h_1 - \varepsilon_7, h_1 + \varepsilon_1]$, h falls within a certain neighborhood of the two tones,

$$H = \omega_1 \bullet 0 + \omega_2 \bullet 7 \tag{10}$$

Where ω_1, ω_2 are weights, indicating the degree of contribution of the two tones at the junction of the hue to the neighborhood tones. The weight size is shown in Figure 8 as the slope equation on both sides of the trapezoid.

Step 3. According to the formula (6) *H, S, V* three components in a one-dimensional vector distribution, calculate the fuzzy histogram.

Step 4. the image library all the pictures are in accordance with the above three steps to calculate the fuzzy histogram, stored in the mat file.

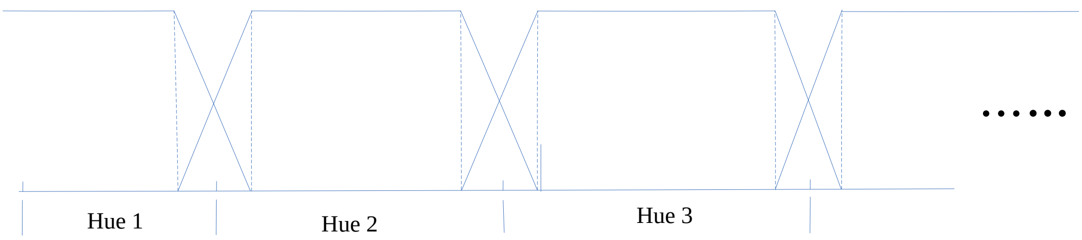

Figure 8. Trapezoidal membership function model of H

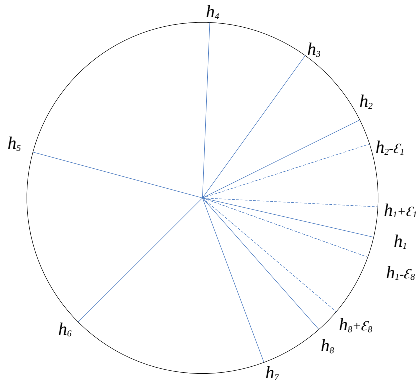

Figure 9. Circular fuzzy quantization

(1) Using Euclidean distance calculates similarity

Using the Euclidean distance algorithm to search the three kinds of car criminal investigation pictures randomly selected, according to (7) calculating the smallest of the top 20 car pictures, the search results shown in Figure 10:

Figure 10. Retrieval of Euclidean Distance of Trapezoidal Membership Function

(2) Using weighted distance calculates Similarity

Using the weighted distance algorithm to randomly select the three kinds of car criminal investigation pictures to carry out the search experiment, according to (8) calculating the smallest of the top 20 car pictures, the search results as shown in Figure 11:

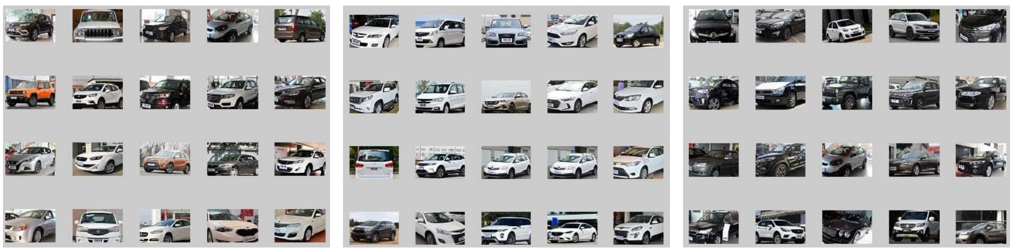

Figure 11. Retrieval of Weighted Distance of Trapezoidal Membership Function

7. Summary of experimental results

In order to effectively validate the search results, in the experiment, the image in the retrieved image is first searched and the image similar to that of the image is found manually,and all images similar to the images to be retrieved are recorded as collections: {relevant}; All the images obtained according to the retrieval result of the algorithm to be retrieved are recorded as a set: {retrieved}. Considering the method of paper [1], we use the following two quantities to express the validity of the algorithm.recall: according to the image retrieval can retrieve relevant images of recall; precision: indicates the precision of the search result, defined as follows:

$$\begin{cases} recall = \dfrac{\{retrieved\} \cap \{relevant\}}{\{relevent\}} \\ precision = \dfrac{\{retrieved\} \cap \{relevant\}}{\{retrieved\}} \end{cases} \tag{11}$$

Judging from the retrieval effect, the trapezoidal membership algorithm is better than the non-uniform quantization algorithm, and the retrieval effect of the triangle membership algorithm is much worse than that of the non-uniform quantization method. Next we look at the results through precision and recall.

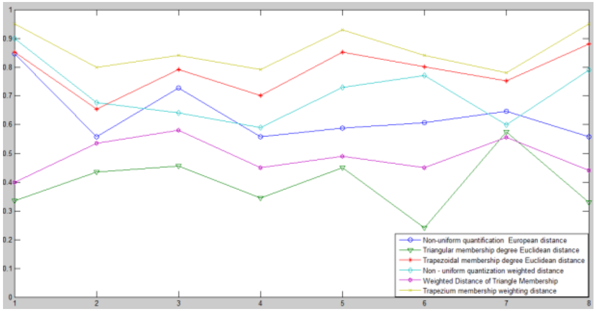

Figure 12. Average precision

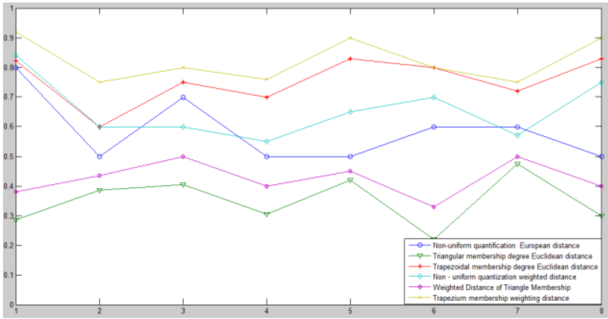

Figure 13. Average recall

From Figure 12, Figure 13 we can get the following conclusions:

(1) The trapezoidal membership algorithm is superior to the non-uniform quantization and the triangular membership function algorithm, whether it is using the European distance or the weighted distance.

(2) Using triangle membership degree algorithm to retrieve image, the effect is worse, even less than the effect of non-uniform quantitative retrieval.

(3) When using Euclidean distance and weighted distance to calculate the image similarity, the weighted distance detection of the results of the picture is generally better than the European distance search, but the difference is not particularly obvious.

For the reason, firstly, the value of chroma H is [0,360], the color change is non-linear, but we defined the maximum value of each hue space as the middle value of each hue, which makes a big error in the selection of the whole membership; Secondly, the color of the car is simple, in the same color field, there is a linear change due to the existence of membership function, so it will increase the difference between the two shades. The trapezoidal membership algorithm guarantees that the primary chromaticity of each hue remains the same, only changes the edge chromaticity. But two similar colors can easily be divided into two different color spaces during the color of the edge. While in this color space, the fuzzy theory algorithm increases the computational complexity, but increases the accuracy and reduces the disparity between the two similar images.

8. Conclusions

For the research of car criminal image processing, domestic and foreign research is generally still in the early stage, but with the rapid development of multimedia, this topic has become a hot spot. In this paper, The experimental comparison of the car criminal investigation image database is carried out by six kinds of image retrieval methods, and some meaningful results are obtained. In terms of retrieval effect, both the Euclidean distance and the weighted distance are used to compute similarity, the trapezoidal membership function quantification algorithm shows obvious advantages. The results show that the triangular membership function search results show a relatively poor effect, which may be related to the color of the car image in this experiment is relatively simple. Weighted distance used to compute image similarity is better than Euclidean distance retrieval. These results have a significant reference value to the selection of the car criminal investigation image retrieval and optimization algorithm, and lay the foundation for the car image retrieval with multiple features in the following chapters.

References

1. A. Baeza, and R. Yates, “Ribeiro-Neto B.Modern Information Retrieval,” *Addison Wesley*,1999.
2. H. He, Y. L. Yu, “An Integrated Fuzzy Histogram Method for Image Retrieval,” *Journal of Image and Graphics*, vol. 6, no. 7, pp. 694-698, 2001. (in Chinese)
3. K. Iqbal, M. Odetayo, and A. JAMES, “Content-Based Image Retrieval Approach for Biometric Security Using Colour, Texture and Shape Features Controlled by Fuzzy Heuristics,” *Journal of Computer and System Sciences*, vol. 78, no. 4, pp. 1258-1277,2012.
4. B. Jyothi, Y. Madhaveelatha, PGK. Mohan, and VSK. Reddy, “Steerable Texture Descriptor for an Effective Content-Based Medical Image Retrieval System Using PCA,” *Advances in Intelligent Systems and Computing*, vol. 239, no. 2, pp. 289-298,2016.
5. J. Li, and H. Y. Guo, “Multi-sample Prototype Selection and Active Learning Strategy in Text Retrieval,” *Computer application*, vol. 32, no. 10, pp. 2899-2903, 2012. (in Chinese)
6. G. H. Liu, and J. Y. Yang, “Content-based Image Retrieval Using Color Difference Histogram”, *Pattern Recognition* , vol. 46, no. 1, pp. 188-198,2013.
7. S. Murala, and QMJ. Wu, “Spherical Symmetric 3D Local Ternary Patterns for Natural, Texture and Biomedical Image Indexing and Retrieval”, *Neurocomputing*, vol. 149, no. 8, pp. 1520-1514,2015.
8. X. Su, “Image Retrieval Based on Region Fuzzy Histogram”, *Journal of Jiangsu University of Science and Technology (Natural Science Edition)*, vol. 32, no. 4, pp. 234-239,2007. (in Chinese)
9. W. W. Wang, G. P. Zhang, and D. Qiu, “Research on Medical Image Retrieval System Based on DICOM Text and Content”, *Computer Engineering and Design*, vol. 33, no. 3, pp. 1014-1018,2011. (in Chinese)
10. B. Z. Wei, J. Wei, and H. Ying, “Optimization and Simulation of Efficient Retrieval Algorithm for Massive Multimedia Image Information”, *Computer Simulation*, vol. 33, no. 11, pp. 280-283,2016. (in Chinese)
11. Y. Yang, “Research on Content Retrieval Enhancement Method Based on Object Semantics”, Nanjing: *Nanjing University*, 2015. (in Chinese)
12. L. A. Zadeh, “Fuzzy Sets”, *Information and Control*, vol. 8, no. 3, pp. 338-353,1965.
13. T. S. Zeng, “Image Retrieval Technology Based on L * a * b Color Space and Gabor Wavelet Transform”, *Journal of Southwest China Normal University (Natural Science Edition)*, vol. 33, no. 6, pp. 124-129, 2011. (in Chinese)
14. Y. K. Zhang, Y. F. Li, and J. G. Sun, “Image Retrieval Based on Multi - feature and High Efficiency Index”, *Computer Engineering and Applications*, vol. 52, no. 7, pp. 181-185,2016. (in Chinese)

Mingdi Hu, female, born in Ankang, shaanxi province, received Ph.D. from Shaanxi Normal University. Now interests in multi-valued logic, fuzzy information processing, and image retrieval.

Mengbin Zhang is a master student from the School of Communications and Information Engineering,Xi’an University of Posts ＆ Telecommunications.His research interests include criminal investigation image processing and fuzzy information processing.

Yilun Lou is a undergraduate student from the School of Mathmatics and Information Science, Shaanxi Normal University. His research interests include criminal investigation image processing and fuzzy information processing.