Image Retrieval Algorithm Fusion of GLCM Features and Tamura Texture Features

Kan Ni[1,a,*], Haohao Zhang[2,b], Xiongwen Jiang[3,c], Taiga Kuroiwa[3,d], Takato Yoshida[3,e], Haolan Zhang[3,f], and Seiji Hashimoto[3,g]

[1]School of Science and Technology, Gunma University, Japan

[2]Ryomosystems, Japan

[3]School of Science and Technology, Gunma University, Japan

* Corresponding author

[a]<t202d003@gunma-u.ac.jp>, [b]<zhang@ryomo.co.jp>, [c]<t212d601@gunma-u.ac.jp>, [d]<t190d509@gunma-u.ac.jp>, [e]<t200d133@gunma-u.ac.jp>, [f]<t221d602@gunma-u.ac.jp>, [g]<hashimotos@gunma-u.ac.jp>

Keywords: image retrieval, GLCM, Tamura features, multi-feature fusion

Abstract. This paper combines color features and texture features, and uses Euclidean distance to calculate the similarity of two images for image retrieval. First, in the HSV space, color features are extracted and normalized. Then, the eigenvalues of GLCM are extracted and combined with Tamura features to form richer texture features. Finally, the color and texture similarity of the image to be retrieved and the image in the image library are calculated respectively, and the color and texture features are fused under different weights to obtain the final similarity. Matlab experiments show that different kinds of images have different precisions when assigning different weights to color and texture. Adjusting the feature weights of an image can improve precision.

1. Background

With the development of digital image technology, people are faced with a variety of images, and it is a problem worth exploring to retrieve matching images from a large number of images according to needs. Image retrieval technology is divided into text-based and content-based retrieval. In the field of intelligent applications, especially for robots equipped with cameras, there is still a huge room for improvement in image processing and identification. This paper aims to solve the problem of judgment and identification after acquiring images in outdoor robot actions.

Content Based Image Retrieval (CBIR) is to extract the features of the image in the image library and the image to be retrieved to compare the similarity, so as to draw a conclusion. Content-based image retrieval features include two aspects: one is low-level visual features, such as color, texture, shape, etc. the other is high-level semantic features, that is, the semantic description of image content, various physical features [1] and Logical relationship. This paper extracts the first type of features, and retrieves images similar to the images to be retrieved from the image library composed of different types of images. A single color-based similarity calculation cannot fully express the image content, so the precision of this retrieval algorithm is very low [2, 3]. In addition, common content-based image retrieval features are texture features, which are used to capture the granularity and recurring patterns of the image surface. Commonly used texture features are features based on Gray Level Co-occurrence Matrix (GLCM) [4]. Literature uses generalized image gray level co-occurrence matrix for image retrieval, but the improvement of precision and recall is not effective significantly [5]. Another commonly used texture feature is the Tamura feature. Reference uses the improved Tamura texture feature to improve the image retrieval performance [6]. Compared with the Tamura feature, there is a certain improvement, but the improvement is limited. This paper combines color features and fused texture features (GLCM and Tamura feature fusion) for image retrieval, and gives the weight coefficients of color features and texture features under the best precision rate, which improves the precision rate.

2. Image description method

There are four types of texture description methods: model method, structural method, spectral method, and statistical method. The model method mainly uses model coefficients to identify texture features, which is difficult to solve; the structural method extracts features by describing texture primitives. Since the wood texture has no obvious texture primitives, the structural method is not applicable; the spectral method is mostly used for standard Or regular texture images, the background of wood texture images is complex and the pixels in a certain texture area are not similar everywhere, the spectral method is not suitable; although the statistical method is computationally complex and time-consuming, it is out of touch with the human visual model and lacks global information. It is difficult to describe the inheritance or dependence of pixels between different scales of texture, but the method is simple and easy to implement [7].

Tamura et al. proposed a texture feature extraction method based on visual psychology, extracting six features of coarseness, contrast, directionality, line-likeness, regularity, and roughness to describe texture, which overcomes the disjointed texture features based on statistics and human visual models. Disadvantages: The disadvantage is that the global information of the texture image is used to a certain extent, and the feature extraction is time-consuming. Among the image feature statistics, the three basic statistics describing texture are gray mean, variance and entropy. Although these three features are all features that describe the global information of texture, they do not describe the relationship between pixels. If the two methods are combined, they can complement each other, describe the texture features better, and improve the classification accuracy.

3. Feature extraction and feature fusion

3.1 Tamura texture features

Tamura et al. based on the research of visual psychology, extracted six texture feature quantities, namely coarseness, contrast, squareness, linearity, regularity, and roughness. These features have good application value in texture synthesis and image recognition.

3.1.1 Coarseness

Larger texture primitive sizes or smaller primitive repetitions can feel rougher. The calculation is as follows:

① Calculate the average brightness of each pixel in the active window with a size of $2^k \times 2^k$ square in the image, as shown in formula (1). where (x, y) is the position of the selected image area in the whole image, g(i, j) represents the pixel brightness value corresponding to the point (i, j) in the selected area, and the pixel is determined by k range,

$$A_k(x,y) = \sum_{i=x-2^{k-1}}^{x+2^{k-1}-1} \sum_{j=y-2^{k-1}}^{y+2^{k-1}-1} g(i,j)/2^{2k} \tag{1}$$

② Calculate the average intensity difference between the non-overlapping active windows in the horizontal and vertical directions for each pixel. As shown in formula (2):

$$\begin{cases} E_{k,h} = |A_k(x + 2^{k-1}, y) - A_k(x - 2^{k-1}, y)| \\ E_{k,v} = | A_k(x, y + 2^{k-1} - A_k(x, y - 2^{k-1} | \end{cases} \tag{2}$$

where $E_{k,h}$ is the average intensity difference in the horizontal direction, $E_{k,v}$ is the average intensity difference in the vertical direction. Determine the optimal size $S_{\text{best}}(i,j) = 2^k$ for the k that maximizes the value of E, and calculate the average value of S_{best} in the whole image to obtain the coarseness,

$$F_{crs} = \frac{1}{m \times n} \sum_{i=1}^{m} \sum_{j=1}^{n} S_{\text{best}}(i,j) \tag{3}$$

where m and n are the length and width of the image, respectively.

3.1.2 Contrast and Directionality

Contrast is obtained by statistics on the distribution of pixel gray values. Defined by $a_4 = \mu_4/\sigma^4$, where μ_4 is the fourth moment and σ^2 is the variance.

Contrast is calculated as follows:

$$F_{con} = \frac{\sigma}{a_4^{1/4}} \tag{4}$$

Orientation is an important feature of an image. Some texture images are obviously directional, while some texture images are more directional.

Tamura uses directionality to measure the apparent directionality of an image. Calculate the gradient vector at each pixel point, the magnitude and direction of the vector are calculated as formula (5):

$$\begin{cases} |\Delta G| = (|\Delta_H| + |\Delta_V| \\ \theta = \tan^{-1}(\Delta_V/\Delta_H) + \pi/2 \end{cases} \tag{5}$$

where Δ_H is the convolution between the image and the first 3 × 3 operator below, which represents the change of the gradient vector in the horizontal direction, and Δ_V is the convolution between the image and the second 3 × 3 operator below, which represents the gradient vector in the vertical direction. The amount of change in direction.

$$\begin{matrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{matrix} \quad \begin{matrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{matrix}$$

Use formula (6) to construct the histogram of θ:

$$H_D(k) = N_\theta(k) / \sum_{i=0}^{n-1} N_\theta(i) \tag{6}$$

where n is the quantization level of the direction angle and t is the threshold. $N_\theta(k)$ refers to the number of pixels under the condition $|\Delta G| \geqslant t,\ (2k-1)\pi/2n \leqslant \theta \leqslant (2k+1)\pi/2n.$

For images with indistinct directions, the histogram H_D appears relatively flat, but for images with obvious directions, the histogram H_D has a more obvious peak. Calculation of Orientation such as formula (7):

$$F_{dir} = \sum_p^{n_p} \sum_{\phi \in w_p} \left(\phi - \phi_p\right)^2 H_D(\phi) \tag{7}$$

Among them, n_p is the number of peaks in the histogram, p is the peak value in the histogram H_D, for a thousand peaks a peak p, w_p is the quantization value range contained in the peak, ϕ_p is the quantization value in the largest histogram value in w_p.

3.1.3 Linearity, Regularity and Roughness

Linearity is calculated as follows:

$$F_{lin} = \frac{\sum_i^n \sum_j^n P_{Dd}(i,j) \cos\left[(i-j)\frac{2\pi}{n}\right]}{\sum_i^n \sum_j^n P_{Dd}(i,j)} \tag{8}$$

where P_{Dd} is the distance point of the n × n local directional co-occurrence matrix.

The texture features of the image are irregular, so partition sub-images are used and the variance of each sub-image is calculated. Here, four characteristics of the sub-image are integrated to measure the regularity of the texture.

$$F_{reg} = 1 - r(\sigma_{crs} + \sigma_{con} + \sigma_{dir} + \sigma_{lin}) \tag{9}$$

r represents the normalization factor $\boldsymbol{\sigma}_{crs}$, $\boldsymbol{\sigma}_{con}$, $\boldsymbol{\sigma}_{dir}$, and $\boldsymbol{\sigma}_{lin}$, are the standard deviations of F_{crs}, F_{con}, F_{dir}, and F_{lin}, respectively.

Based on human visual perception of texture, Tamura et al. Psychological research defines roughness as follows:

$$F_{rgh} = F_{crs} + F_{con} \tag{10}$$

3.2 Implementation of the algorithm in Matlab

The experimental images are all 512×512 pixels. The following experiments use 200 texture images as test samples, including 70 ruled images, 60 parabolic images, and 70 random images. First, the six texture features in the Tamura method are extracted from the grayscale image reduced to 65×65 pixels.

4. Texture features based-on Gray Level Co-occurrence Matrix (GLCM)

4.1 HSV color space quantization and feature extraction

Compared with other visual features, color features are not sensitive to changes in image size, orientation, viewing angle, etc. At the same time, color features have good robustness to changes in image quality and noise. Therefore, the successful extraction of color features plays an important role in image retrieval.

Compared with RGB space, HSV (hue, saturation, brightness) space expresses color in the same way as people's visual habits. Therefore, the RGB space is converted into the HSV space, and in order to improve the efficiency and accuracy of the retrieval, the values in the HSV space are non-uniformly quantized. In this paper, H quantization is 16 levels, S quantization is 4 levels, and V quantization is 4 levels.

After quantizing the HSV space, the three components of H, S, and V are constructed into a one-dimensional color feature vector: HSV=H*16+4*S*4+V.

4.2 Extraction and fusion of texture features

4.2.1 Gray-scale co-occurrence matrix

In this paper, the statistic of GLCM is calculated as a part of texture information. GLCM can reflect the distribution of gray level of gray level image in spatial position, and is used to describe the correlation of gray level space. It is defined as: the joint probability distribution of any two gray pixels with the distance d and the direction θ in the image $I(i,j)$ appearing at the same time, denoted as $p(i,j,d,\theta)$, by these probability values A grayscale co-occurrence matrix $P=[p(i,j,d,\theta)]_{L\times L}$ is formed, where θ generally takes 0°, 45°, 90°, and 135°; L is the number of gray levels. Usually, before calculating GLCM, because the gray level is too large, the amount of calculation is large and time-consuming, so the gray level should be compressed first.

The gray level co-occurrence matrix cannot directly represent the texture features of the image. On the basis of GLCM, some statistics are used to describe texture features. In this experiment, five statistics of contrast, correlation, entropy, stationarity, and energy are used.

a) Contrast

$$\mathrm{CON}=\sum_i \sum_j (i-j)^2 P(i,j) \tag{11}$$

The larger the value of the element far from the diagonal in the grayscale co-occurrence matrix, the greater the contrast and the clearer the texture visually.

b) Correlation

$$\mathrm{COR}=\frac{\sum\sum(i-\bar{x})(j-\bar{y})P(i,j)}{\sigma_x\sigma_y} \tag{12}$$

The larger the COR value, the more uniform and equal the matrix element values are, and the smaller the COR value, the greater the difference between the matrix element values. COR describes how similar the element values of the gray-scale co-occurrence matrix are in the row or column direction.

c) Entropy

$$\mathrm{ENT}=-\sum_i \sum_j P(i,j)\lg P(i,j) \tag{13}$$

When the entropy is larger, it indicates that the image is more complex, and vice versa. Entropy reflects the amount of random information carried by the image and represents the complexity of the image.

d) Energy

$$\mathrm{ASM}=\sum_i \sum_j P(i,j) \tag{14}$$

Energy reflects the stability of image grayscale changes. When the matrix elements are centrally distributed, the ASM value is large at this time, indicating that the image is a regularly changing and stable texture.

e) Contrast

$$\mathrm{IDM} = \sum_i \sum_j \frac{1}{1+(i-j)^2} P(i,j) \tag{15}$$

The inverse moment reflects the roughness of the image texture, the inverse moment of the coarse texture is larger, and the inverse moment of the fine texture is smaller.

4.2.2 Fusion of GLCM texture features and Tamura texture features

The features of contrast, correlation, entropy, energy and inverse moment are extracted by GLCM, and each aspect has eigenvalues in 4 directions. In this experiment, the average value of each direction feature is taken. Therefore, the GLCM feature of each image is a 5-dimensional feature vector; the Tamura feature is a 3-dimensional feature vector. In this paper, the extracted GLCM features and Tamura features are combined into an 8-dimensional comprehensive texture feature. Because the physical meaning and value range of each feature component are different, it is necessary to perform internal normalization processing on each component in the image feature vector.

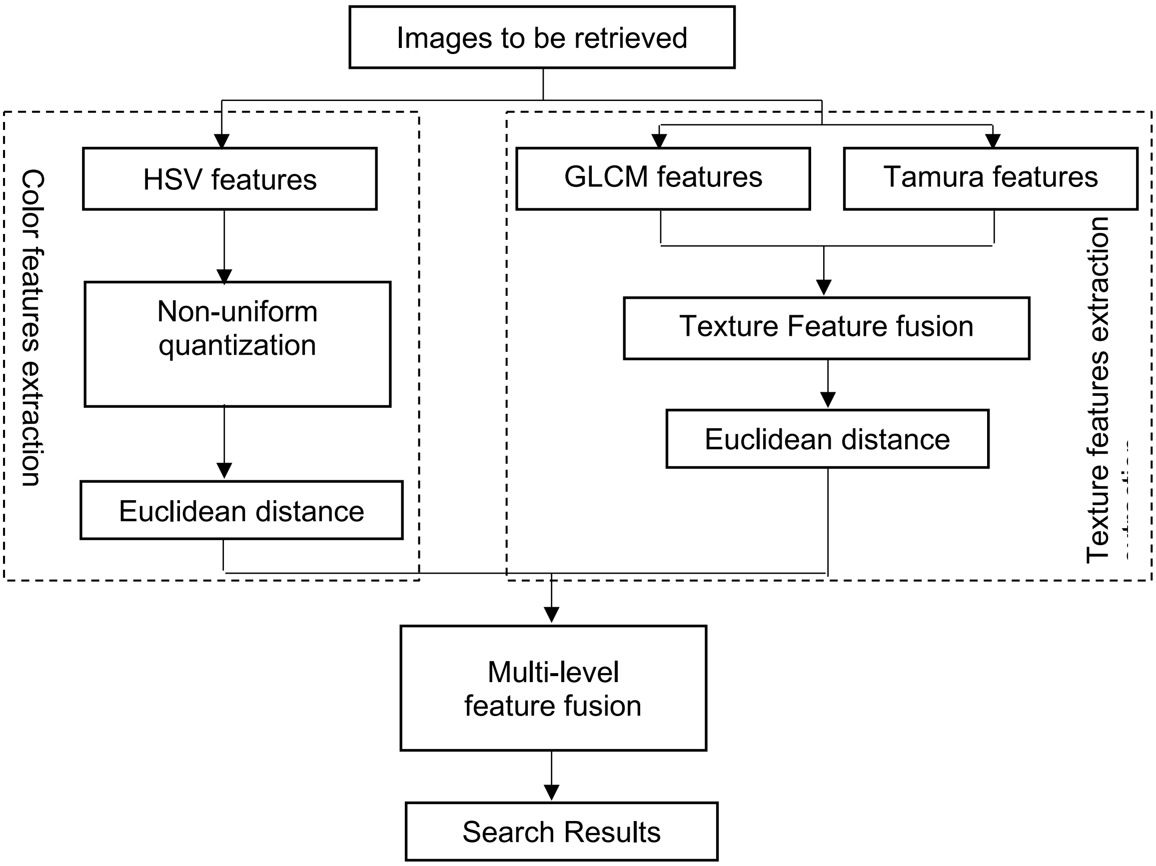


Fig. 1. Algorithm process.

4.3 Image Retrieval Algorithms

Image retrieval is fuzzy search, and Euclidean distance is often used as a similarity calculation formula, so this method is also used for similarity matching in this paper. Then the similarity calculation formula is:

$$D(P,Q) = s_1|C^P - C^Q| + s_2 \sum_{\mathrm{i}=1}^{8} \left|T_{\mathrm{i}}^P - T_{\mathrm{i}}^Q\right| \tag{16}$$

Among them, P and Q are the images to be retrieved and the images in the image library, S_1 and S_2 are the weights of color features and texture features, and C and T represent color features and texture features respectively. In this algorithm, the texture features are the fusion of GLCM features and Tamura features. 8-dimensional feature vector.

Color focuses on describing the overall features of the image, while texture focuses on describing the local features of the image. Any single feature cannot fully extract the image information. Based on the color features, this paper extracts the GLCM and Tamura texture features to make the texture feature information more comprehensive. Compared with previous papers, the accuracy of image retrieval using color features and single texture features is higher. The process is shown in Figure 1.

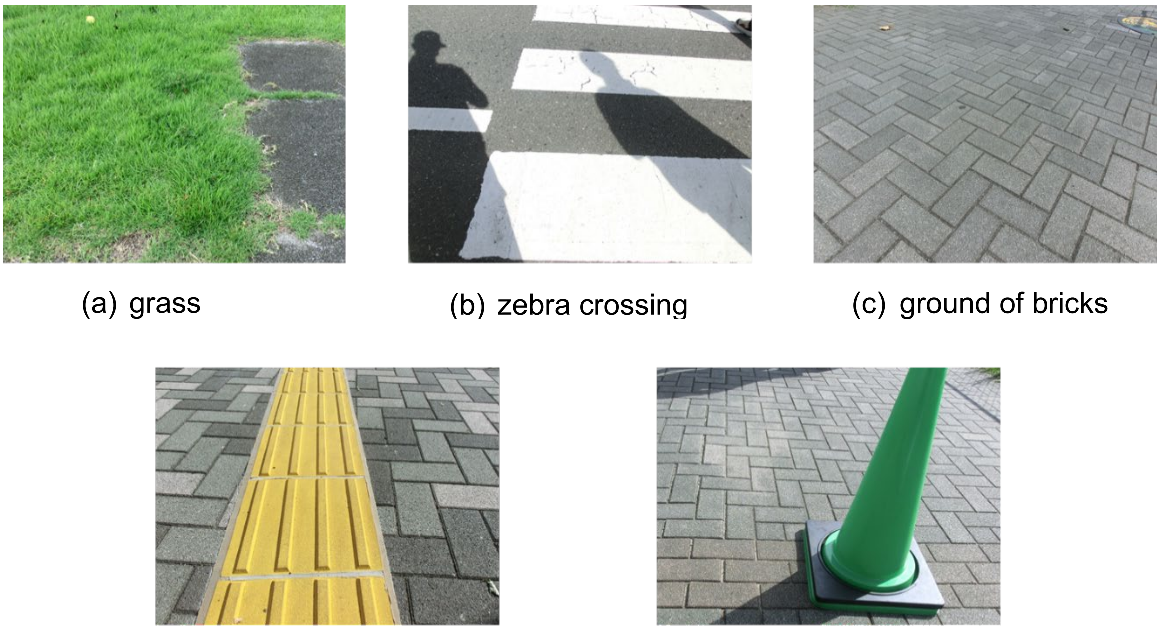
5. Analysis of experimental results

5.1 Experimental setup

The hardware configuration of this experiment is: Intel(R) Core(TM) i7-9700 CPU @ 3.00GHz and RAM is 16 GB; the software configuration is 64-bit Windows11 operating system, MATLAB R2021b. In order to test the performance of the algorithm, the experiment randomly selected a part of the images in the Corel image library (five categories: ground of bricks, zebra crossing, road cones, blind roads, grass, 80 images each) as the image library.

5.2 Experimental Analysis

According to the algorithm proposed above, the similarity distance is sorted in a library, the number of images of the same type as the target image (the image to be retrieved) is screened in 80 images (that is, the total number of images of the target class), and the number of images in the same class is calculated and find the average precision.



(a) grass (b) zebra crossing (c) ground of bricks

(d) blind roads (e) road cones

Fig. 2. Image samples.

$$query\ accuracy = \frac{The\ number\ of\ images\ of\ the\ same\ type\ as\ the\ target\ image\ in\ the\ retrieval\ library}{Total\ number\ of\ target\ class\ images} \times 100\%$$

An example of the 5 types of images is shown in Figure 2, and the experimental results are shown in Table 1.

It can be concluded from Table 1 that when the color and texture feature weights are different, the precision rate is also different, but the precision rate of each image basically conforms to the normal distribution. According to the experimental results, the grass and the zebra crossing have different color and texture weight ratios. The precision rate is the highest when it is 0.6∶0.4. When the weight ratio of color and texture is 0.5:0.5, the precision rate of road cones is the highest, and when the weight ratio of color and texture is 0.4:0.6, the precision rate of blind roads and bricks is the highest.

Table. 1. Experimental results.

Weights / Figures / Proportion	Precision of grass (%)	Precision of zebra crossing (%)	Precision of bricks (%)	Precision of blind roads (%)	Precision of road cones (%)
0.1:0.9	83.94	84.25	63.01	69.01	81.25
0.2:0.8	85.81	89.56	63.45	70.86	82.94
0.3:0.7	85.37	90.25	64.26	71.57	83.81
0.4:0.6	90.63	91.12	64.58	72.59	85.13
0.5:0.5	91.31	91.46	65.95	72.37	85.00
0.6:0.4	91.44	92.50	65.20	71.19	82.38
0.7:0.3	90.86	91.75	64.15	69.66	81.25
0.8:0.2	88.55	86.45	62.46	64.25	80.88
0.9:0.1	87.19	85.06	61.03	62.30	79.25
Average precision	88.34	89.15	63.79	69.31	82.43

Since the color of the zebra crossing image is relatively simple and the color features are obvious, when the color weight is relatively large (0.6:0.4), the precision rate is relatively high (92.50%). The color features of the grass image are also more obvious and the texture itself is different from other types of images, so the precision rate is also relatively high (91.44%). Although the background colors of blind roads and road cones are similar, the texture features are significantly different, so when the weight of the texture is increased to 0.4:0.6, the precision is improved. The background color of the bricks is similar to the background color of blind roads and road cones, and the texture tendency is not very obvious. Therefore, when the weight of the color feature and the texture feature is 0.5:0.5, the precision rate of the bricks is the highest, that is, the bricks have the highest precision. For images, the extracted color and texture features play a similar role in image retrieval. But compared with the other four images, the average precision of bricks is 63.79%, which is relatively low. The poor performance is due to the limited amount of data available, which may require training a deep learning model, resulting in average performance. The overall accuracy obtained by the classifier on GLCM-based images is unstable and low compared to grayscale images. The reason for the poor performance is the limited amount of data available, resulting in average performance.

From the perspective of sample individual pictures, and compared with a separate GLCM algorithm, and using PSNR/SSIM/FSIM to objectively evaluate before and after segmentation, the experimental results show that the improvement is effective in Figure 3.

Image Quality Assessment (IQA) is considered a fundamental characteristic of an image, allowing us to measure the degradation of perceived images. Typically, this degradation is determined by comparing the image to an ideal reference image.

For objective image quality assessment, there exist various techniques and metrics, which can be categorized based on the availability of a reference image.

Among these metrics, the mean squared error (MSE) stands out as the most widely used and simplest full-reference metric. It involves calculating the squared intensity differences between the distorted and reference image pixels and averaging them, along with the peak signal-to-noise ratio (PSNR) of the corresponding quantity.

Metrics like MSE and PSNR are frequently employed in image quality assessment due to their simplicity in computation, clear physical interpretations, and ease of mathematical implementation in optimization contexts. However, they may not always align well with perceived visual quality and lack representation normalization. To address these limitations, researchers have developed two normalized reference methods that account for structural and feature similarities. The Structured Similarity Indexing Method (SSIM) provides a normalized mean value of the structural similarity between two images, while the Feature Similarity Indexing Method (FSIM) provides a normalized mean value of the feature similarity between the two images. All these metrics fall under the category of full-reference image quality measurement. [8]

In this paper, we perform a comparison of the PSNR, SSIM and FSIM values obtained from denoising experiments conducted with different noise concentrations.

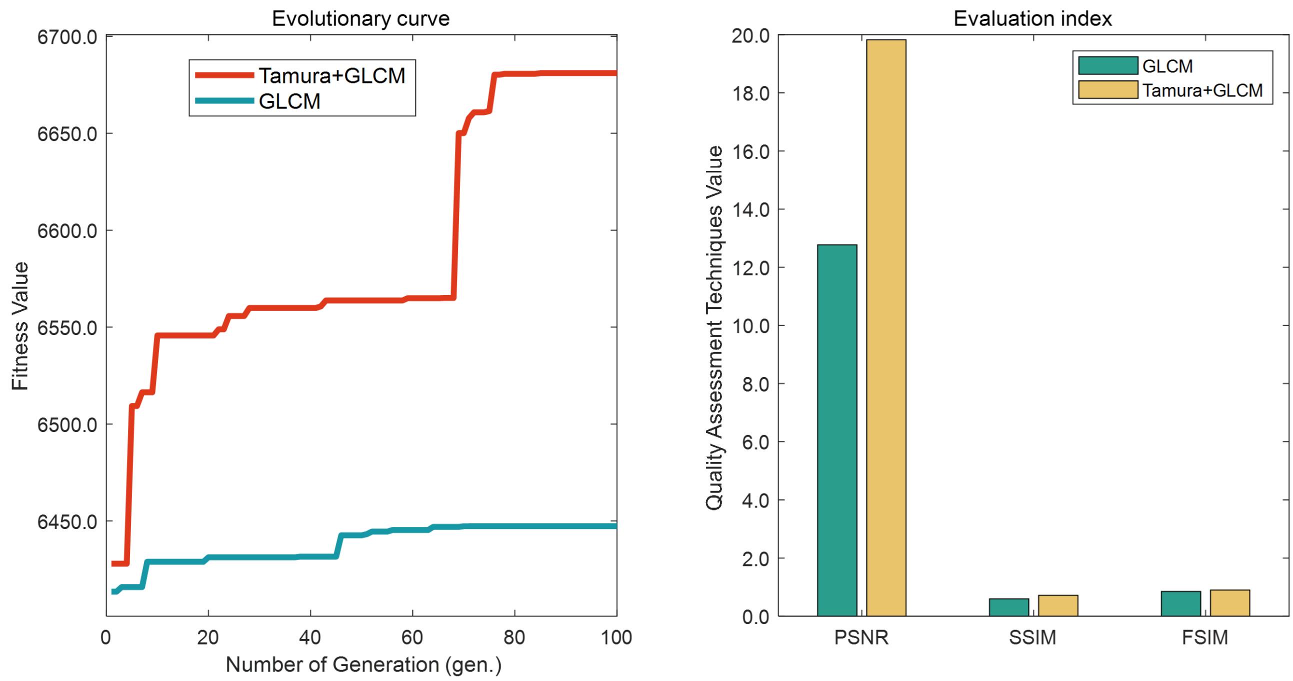


Fig. 3. Evolutionary curve and Evaluation index compare.

It can be seen from the fitness curve and the value of the evaluation index that the result obtained by the improved algorithm is better than that of the single GLCM algorithm, and the improvement is effective. As can be seen from the figure on the right, under the PSNR evaluation standard, the result of the hybrid algorithm has obvious advantages. The mixed algorithm results also have a small improvement under the SSIM and FSIM standards.

6. Conclusion

In this paper there are different textures and shape feature extraction techniques have been discussed. The main objective of content based image retrieval is to develop is an efficient image retrieval scheme. Through the fusion of color features and texture features, the algorithm proposed in this paper can better retrieve similar images and similar images. Experiments show that the precision is closely related to the weights s1 and s2 of color and texture. Different weights of color features and texture features of the same image lead to large differences in precision. Through many experiments, the optimal weight ratio of color features and texture features is obtained when different types of images are retrieved, which improves the retrieval accuracy to a certain extent. In the future, in order to improve the precision, the optimal weight ratio when GLCM features, and Tamura features are fused can be further obtained. A comparison is performed between different textural and shapes features are combination of GLCM textural properties and Tamura textural properties. The combination of tamura textural feature vectors perform better than combination of GLCM Textural and shape moment invariant. The result can be further improved by using KNN K-Nearest Neighbor classifier at preprocessing step which gives better result at the retrieval time. Support Vector Machine, Soft computing techniques like neural network can be also applied as a classifier to improve the retrieval time. Evolutionary algorithms can also be used to optimize the result for the better feature selection process.

References

[1] A. Aqthobilrobbany, A. N. Handayani, D. Lestari, Muladi, R. A. Asmara and O. Fukuda, “HSV Based Robot Boat Navigation System”, ***Proceeding of 2020 International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM)*** (Surabaya, Indonesia), November 2020.

[2] C. -H. Su, H. -S. Chiu and T. -M. Hsieh, "An efficient image retrieval based on HSV color space", ***2011 International Conference on Electrical and Control Engineering*** (Yichang, China), September 2011.

[3] D. S. Y. Kartika and D. Herumurti, "Koi fish classification based on HSV color space", ***2016 International Conference on Information & Communication Technology and Systems (ICTS)*** (Surabaya, Indonesia), October 2016.

[4] D. Indra, H. M. Fadlillah, Kasman and L. B. Ilmawan, "Rice Texture Analysis Using GLCM Features", ***2021 International Conference on Electrical, Computer and Energy Technologies (ICECET)*** (Cape Town, South Africa), December 2021.

[5] M. Umaselvi, S. S. Kumar and M. Athithya, "Color based Urban and Agricultural Land classification by GLCM Texture Features", ***IET Chennai 3rd International on Sustainable Energy and Intelligent Systems (SEISCON 2012)*** (Tiruchengode, India), December 2012.

[6] H. Tamura S. Mori and T. Yamawaki "Textural Features Corresponding to Visual Perception", ***IEEE Transactions on Systems Man and Cybernetics***, Vol. 8, No. 6, pp. 460-473, 1978.

[7] J. Chen et al., "WLD: A Robust Local Image Descriptor", ***IEEE Transactions on Pattern Analysis and Machine Intelligence***, Vol. 32, No. 9, pp. 1705-1720, 2009.

[8] U. Sara, M. Akter, and M. Uddin, "Image Quality Assessment through FSIM, SSIM, MSE and PSNR—A Comparative Study", ***Journal of Computer and Communications***, Vol. 7, pp. 8-18, 2019.