RESEARCH

Color Correction and Repair of Haze Images under Hue-saturation-intensity Color Space and Machine Learning

Haibin Sun and Haiwei Liu*

*Correspondence: 
18769094020@163.com 
College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao, China 
Full list of author information is available at the end of the article

Abstract

To improve the visual effect and quality of haze images after fog removal, a model for color correction and repair of haze images under hue-saturation-intensity (HSI) color space combined with machine learning is proposed. First, the haze image imaging model is constructed according to the atmospheric scattering theory. Second, based on HSI color space, the color enhancement and fog removal of the haze image model is proposed, and a haze image-transmittance gallery is constructed. Third, the visual dictionary of the transmittance graph is obtained by training the k-means clustering algorithm based on density parameter optimization and support vector machine algorithm based on genetic algorithm optimization. Fourth, based on the visual dictionary and the atmospheric scattering model, the haze image is repaired and defogged, and the subjective visual effects and objective evaluation indexes of color enhancement and fog removal of haze images are compared. It is concluded that the algorithm can effectively guarantee the detail and clarity of the image after defogging.

Keywords: Haze image; HIS color space; K-means clustering algorithm; Support vector machine; Visual dictionary defogging

Introduction

Severe weather often occurs in North China, such as fog, haze, and sandstorms [1]. In these inclement weather conditions, suspended particles in the atmosphere reduce the contrast and clarity of images collected outdoors, and may also cause color shifts. [2]. The loss of image details may cause failures in road monitoring systems, military reconnaissance systems, and satellite monitoring systems, which in turn affect people’s lives, social production, and national defense security [3].

In order to solve this problem, experts and scholars in relevant fields at home and abroad have successively proposed degraded image enhancement algorithms to improve the image quality of haze or sand. Dong et al. [4] proposed a defogging algorithm based on the improved multi-scale Retinex algorithm, which has a better defogging effect than the adaptive histogram equalization method and repaired the image degradation problem. In order to solve the problems of uneven distribution of image fog density and color imbalance in haze weather, Mei et al. [5] proposed a multi-layer fusion-based haze image enhancement algorithm and estimated transmission images based on the dark primary color prior theory. Finally, the model was qualitatively and quantitatively verified. Wen et al. [6] based on the
application of the Kuwahara filter refinement algorithm in the process of transmittance refinement, proved that the algorithm can improve the speed of haze image defogging. Chaudhry et al. [7] proposed an image defogging algorithm combining a combination of median filtering, accelerated local Laplacian filtering and image decomposition based on 10 restricted gradients, and qualitative and quantitative verification of outdoor RGB images and remote sensing images.

At present, most researches on image defogging are based on the prior principle of dark primary colors, but there are few studies on applying machine learning to defogging images. Traditional prior defogging method has some disadvantages such as difficulty in capturing some features of haze images and low effect of image defogging. For example, in the process of image defogging, it is susceptible to interference from bright areas and white objects, resulting in image distortion. The existing haze image algorithms are all based on the atmospheric scattering model for image defogging. In the model, the transmittance and atmospheric light are the key variables that affect the defogging algorithm, but these two variables cannot be obtained directly from a single haze image. Therefore, an accurate estimation of the transmittance and atmospheric light of the haze image is the key to improving the dehazing effect. The dark primary color a priori algorithm is to obtain the transmittance and atmospheric light of the image by comparing the original image and the haze image, which has strong limitations. In order to solve this problem, a machine learning algorithm is proposed to capture the features of haze images and estimate the transmittance of haze images more accurately. This article first uses the HIS color space method to construct the haze image-transmittance graph library [8]; Secondly, the machine learning algorithm is constructed using the k-means clustering algorithm optimized by the density parameter method and the GA-SVM algorithm of the support vector machine improved by the genetic algorithm [9–11]. The machine learning algorithm is used to train the haze transmittance image library to obtain visual words with different characteristics, and use these visual words with different characteristics to construct a visual dictionary. Then, when defogging the haze image, first detect the haze image, and then select the best transmittance map according to the characteristics of the haze image; Finally, combined with the atmospheric scattering model, the haze image is defogged. The experimental part verifies the effectiveness of the algorithm from subjective and objective aspects.

**Method**

**Haze image imaging model**

In the haze weather, there will be more suspended particles in the atmosphere than usual, and these suspended particles will generate Mie scattering during natural light propagation [12]. During image acquisition, light from the sky and ground accidentally enters the lens, forming a haze image. [13] Therefore, based on the principle of Mie scattering, the model used to describe the degradation process of the haze image is of great significance for the clear processing of the haze image. The mathematical expression of the haze image imaging process is:

\[
F(x) = Z(x) t(x) + A(1 - t(x))
\]  (1)
In the above formula, $x$ represents the position of pixels in the image, $F$ is the input haze image, $Z$ is the image when there is no fog, $A$ is the atmospheric light value, and $t$ is the transmittance. In the haze imaging model, $t$ and $A$ are unknown quantities. To get a clear image $Z$, we need to estimate the value of image transmittance $t$ and the value of atmospheric light $A$.

When the atmosphere is uniform, the formula for calculating the transmittance $t$ is:

$$t(x, y) = e^{-\beta dx}$$

(2)

In the above formula, $\beta$ is the scattering coefficient of the atmosphere, and $d$ is the distance between the target object and the lens.

Refinement of transmittance based on guided filter

It can be seen from formula (1) that in fog and haze, the collected image contains two models [14]. $Z(x)t(x)$ is the light attenuation model, $A(1 - t(x))$ is the atmospheric light model. However, the formula contains three unknown parameters and is an ill-conditioned equation. It is necessary to introduce a priori condition in the calculation, such as the prior theory of dark primary colors.

The mathematical expression of the dark primary color is:

$$Z^\text{dark}(x) = \min_{y \in \Omega(x)} \left( \min_{c \in \{r, g, b\}} (Z^c(y)) \right) \to 0$$

(3)

In the above formula, $\Omega(x)$ is the partial window of the neutral point $x$, and $Z^c$ is a color channel of $Z$ image in RGB color space.

According to the priori theory of dark primary colors, the deformation of formula (1) is as follows:

$$\min_{y \in \Omega(x)} \left( \min_{c \in \{r, g, b\}} \left( \frac{F^c(y)}{A^c} \right) \right) = \sim t(x) \min_{y \in \Omega(x)} \left( \min_{c \in \{r, g, b\}} \left( \frac{Z^c(y)}{A^c} \right) \right) + (1 - \sim t(x))$$

(4)

In the above formula, $A^c$ is always a positive number, then the coarse transmittance $\sim t(x)$ is:

$$\sim t(x) = 1 - \min_{y \in \Omega(x)} \left( \min_{c \in \{r, g, b\}} \left( \frac{F^c(y)}{A^c} \right) \right)$$

(5)

In real life, when viewing distant objects, there will still be smog. Therefore, to keep a small amount of fog in the distance, a constant $\omega \in (0, 1]$ needs to be substituted into formula (5):

$$\sim t(x) = 1 - \omega \min_{y \in \Omega(x)} \left( \min_{c \in \{r, g, b\}} \left( \frac{F^c(y)}{A^c} \right) \right)$$

(6)

In the above equation, $\sim t(x)$ is the coarse transmittance, and $\omega$ is 0.95.
To solve the halo effect that occurs when coarse transmittance is used for image defogging, a guide filter is introduced for image processing. The formula of the guide filter is:

\[ q_i = a_k F_i + b_k, \quad \forall i \in w_k \]  

(7)

In the above formula, \( q \) is the output image, \( F \) is the guide map, \( w_k \) is the partial window, and \( a_k \) and \( b_k \) are the linear constant coefficients of \( w_k \). To ensure that the error between the input value of the function and the actual values is minimal, the coefficients \( a_k \) and \( b_k \) can be obtained by the least square method:

\[
\begin{align*}
    a_k &= \frac{(1/\lvert w \rvert) \sum_{i \in w_k} F_i p_i - \mu_k \sigma_k^2 + \varepsilon}{\sigma_k^4 + \varepsilon} \\
    b_k &= \bar{p}_k - a_k \mu_k
\end{align*}
\]

(8)

In the above formula, \( p \) is the input image, \( \mu_k \) is the variance of \( F \) in \( w_k \), \( \sigma_k^2 \) is the mean of \( F \) in \( w_k \), \( \lvert w \rvert \) is the total number of pixels in \( w_k \), and \( \bar{p}_k \) is the mean of \( p \) in \( w_k \). Then formula (7) can be updated to:

\[ q_i = \frac{1}{\lvert w \rvert} \sum_{k} \sum_{i \in w_k} (a_k F_i + b_k) \]  

(9)

Image enhancement and defogging based on HSI color space

When the dark channel prior method performs haze image defogging, there will be a deviation in some areas where the dark channel is not 0, resulting in original image distortion and the soft matting algorithm has a high time and space complexity. Therefore, the HIS color space method is proposed to conduct image enhancement and defogging processing for haze images. First, the HIS color space is used to roughly estimate the transmittance and atmospheric light of the haze image, second, a guided filter is used to refine the transmittance, and then the color mapping equation is used to adjust the image. Finally, a clear image is obtained. Through experiments, the HIS haze image enhancement algorithm is proved to have a better dehazing effect.

In the same image, the physical quantities of the HSI and RGB color spaces are not the same, but they can be converted to each other. Compared with the RGB color space [15], the HIS color space is more in line with the human visual perception of color. Therefore, this paper uses HIS color space for image defogging and color enhancement. The mathematical expression for converting RGB to HIS color space is:

\[
\begin{align*}
    H &= \begin{cases} 
    \theta & B \leq G \\
    360 - \theta & B > G 
    \end{cases} \\
    S &= 1 - \frac{3}{\pi} \frac{\min (R, G, B)}{R + G + B} \\
    I &= \frac{1}{3} (R + G + B)
\end{align*}
\]

(10)

\[
\theta = \arccos \left\{ \frac{\frac{1}{2} [(R - G) + (R + B)]}{(R - G)^2 + (R - G) (G - B) + \frac{1}{4} (G - B)^2} \right\}
\]

(11)
Based on formula (9), formula (10), formula (11) and formula (1), the image degradation model that can obtain the HSI color space is:

\[
\begin{align*}
H_F (x) &= H_Z (x) \\
S_F (x) &= t (x) [I_Z (x) / I_F (x)] S_Z (x) \\
I_F (x) &= I_Z (x) t (x) + A_I (1 - t (x)) \\
\end{align*}
\]  

(12)

In the above formula, \( H_F \), \( S_F \) and \( I_F \) are HSI color space models in the haze image, \( H_Z \), \( S_Z \) and \( I_Z \) are HSI color space models in the non-fog image, and \( A_I \) is atmospheric light in the brightness channel.

The mathematical expression of the dark primary colors in the HSI color space is:

\[
\min_{y \in \Omega (x)} \min (1 - S_Z (y), I_Z (y)) \to 0 
\]  

(13)

In the above formula, \( S_Z \) and \( I_Z \) are the variables of the model in the HSI color space.

When performing color enhancement of an image, it is first necessary to estimate atmospheric light and transmittance. The atmospheric light in the brightness channel can be regarded as the average value of the atmospheric light in each RGB channel:

\[
A_I = \overline{A_c}, \quad c \in \{ r, g, b \}
\]  

(14)

The calculation expression of transmittance is:

\[
t (x) = 1 - \omega \left[ \min_{y \in \Omega (x)} \left( \frac{G_F (y)}{A_I} \right) \right]
\]  

(15)

The value of \( \lambda \) is fixed at 0.95.

Based on formula (1), (14), and (15) to obtain the restored image, and then use the color mapping equation to adjust the color in the image. The mathematical expression of the color mapping equation is:

\[
R_d (x) = \frac{0.01 L_{d \text{max}}}{\log_{10} (Z_{c \text{max}} + 1)} \cdot \frac{\log (Z_c (x) + 1)}{\log \left( 2 + \left\{ \frac{Z_c (x)}{Z_{c \text{max}}} \right\}^{\log(\lambda) / \log(0.5)} \right) \cdot 8}
\]  

(16)

In the above formula, \( R_d \) is the image after color adjustment, \( Z_{c \text{max}} \) is the maximum value in the RGB corresponding channel, \( \lambda \) is the offset parameter (value is 0.85), and \( L_{d \text{max}} \) is the maximum brightness (value is 100).

Image dehazing based on machine learning

Firstly, this paper downloads 10000 images containing haze from the Baidu image database with the keyword "haze image"; Second use the HIS color space image enhancement algorithm proposed in 1.3 to generate transmittance libraries for these images; In order to increase the number of samples in the image training set and
make the transmittance library feature richer. Then use the Gaussian pyramid block sampling method to sample these images to obtain the transmittance block corresponding to the haze image. Then the LPB information of the image is extracted to label the image, and the image is simply clustered using the improved k-means clustering algorithm to obtain k initial visual words and the corresponding transmittance map. The traditional k-Means algorithm [16] is usually defined by the Euclidean distance between two samples. But the algorithm has a strong dependence on the selection of the initial cluster center point. So this paper uses density parameters to optimize the k-Means algorithm. The specific improvement steps are: (1) Input data set S, the number of clusters k, and the constant Minpts; (2) Calculate all the density values in the data set S based on the constants Minpts, solve the density mean, and get the density set C. (3) Select the data object corresponding to the maximum value in the set C as the center point C0 of the initial cluster; (4) According to the above method, obtain k initial cluster center points of the high-density area; (5) Use the traditional k-Means clustering step to cluster the samples, and finally output the clustering results. The specific flow chart is shown in Fig. 1.

Then use SVM to cross-validate K visual words of different categories, improve the characteristics and representativeness of each visual word, and obtain multiple visual words and their corresponding transmittance maps. Research shows that the kernel function $g$ and penalty factor $c$ in SVM will affect the performance of the algorithm [17]. To improve the performance of the SVM algorithm, the parameters $g$ and $C$ need to be optimized. The kernel function in the SVM in this paper selects the Gaussian kernel function, and improves the SVM based on GA and obtains the GA-SVM algorithm [18, 19]. The specific optimization process is shown in Fig. 2.

After obtaining the optimal $g$ and $c$ of the SVM, the improved SVM algorithm is used to cross-validate the K initial visual words, so that the obtained visual words are more representative. Finally, a large number of visual words and corresponding transmittance maps are obtained to form a visual dictionary. Suppose the structure of the image is that the upper part is far and the lower part is close, that is, the gray value of the upper part of the transmittance image is 0, and the gray value of the lower part is 1. The upper left corner of the transmittance image is defined as the origin of the image coordinates (0,0), then the gray value of each point in the image is:

$$T(x, y) = \frac{y}{H}$$

In the above formula, $x$ and $y$ are the horizontal and vertical coordinates of the two-dimensional coordinate system, and $H$ is the height of the image.

When defogging the haze image, the Gaussian pyramid block is used to sample the haze image, and the transmittance map corresponding to the visual word similar to the haze image is selected. When selecting a similar transmittance map for the haze image, the corresponding transmittance map may not be found. To solve the problem that the transmittance map of certain areas in the image does not match, after the transmittance map of the haze image is obtained, the guide filter is used to refine the transmittance map of the haze image to improve the quality of image defogging. It has been assumed that the structure of the image is that the upper
part is far and the lower part is close, and the upper area of the picture is generally
the area with the densest smog or the sky. Select the first % of the pixel values from
the first quarter of the image to form a set, and then select the maximum value
of the RGB channel at the corresponding position in the set as the atmospheric
light value Ac of the atmospheric scattering model. Based on the above-obtained
transmittance map and atmospheric light value, the atmospheric scattering model
formula $Z_c(x) = \frac{F_c(x) - A_c}{t(x)} + A_c$ is used to restore the haze image. The haze image
defogging process proposed in this paper is shown in Fig. 3.

Results and Discussion
In the experiment, Intel (R) Core (TM) i5 CPU@ 3.00 GHz and 8.00GB of mem-
ory were used as the hardware simulation environment. The software simulation
environment is QT5.5.0 and Matlab R2012a, Ubuntu operating system.

Defogging experiment based on HIS haze image
In order to verify the haze image enhancement effect based on the HIS color space
proposed in this paper. This paper chooses He algorithm [20], Huang algorithm [13]
and histogram equalization algorithm [17] to compare and analyze the effect of
image enhancement.

It can be seen from Fig. 4 that due to the influence of haze, the scene in the
collected image is almost unclear, and the visual effect is very poor. In Fig. 4(b), the
scene can be observed after processing with the He algorithm, and the overall effect
is better. Fig. 4 (c) and (d) is the use of Huang and histogram equalization algorithm
for image defogging, the scene can be observed in the image, but the overall color of
the image is dark and the color of the distant scene is more saturated serious. Fig.
4(e) shows the haze image enhancement and defogging processing algorithm based
on the HSI color space proposed in this paper. It can be seen that the algorithm can
effectively enhance the color of haze images, and on the whole, the scene is clear
degree and saturation are better than other algorithms.

In order to better evaluate the defogging performance of the HIS color space
method for haze images, five indicators of $G, H, \sigma, e$ and $\tau$ are used for quantitative
evaluation.

| algorithms               | AveGrad | H  | $\sigma_g$ | e   | $\tau$ |
|-------------------------|---------|----|------------|-----|--------|
| He                      | 11.511  | 7.119 | 47.598     | 1.352| 1.738  |
| Huang                   | 15.274  | 7.126 | 50.367     | 0.426| 1.982  |
| Histogram equalization  | 13.886  | 7.003 | 45.289     | 0.229| 1.503  |
| This paper              | 16.893  | 7.379 | 76.326     | 0.908| 2.034  |

It can be seen from Table 1 that the average gradient, information entropy, standard
deviation, and R of the algorithm proposed in this paper are all maximum. It
shows that the algorithm proposed in this paper has obvious advantages over other
algorithms in terms of color enhancement and defogging of haze images. It shows
that the algorithm proposed in this paper can not only eliminate the color shift
phenomenon in the haze image, but also can effectively restore the details of the
image, and at the same time can improve the color saturation, definition, and contrast in the haze image. Therefore, the HSI color space image processing algorithm proposed in this study can enhance and remove the haze in the color of the haze image, which has a very significant improvement effect.

**Testing of improved machine learning algorithms**

This article first tests the performance of the improved machine-learning algorithm to ensure the effectiveness of the machine learning algorithm when defogger the haze image.

**Performance test of improved k-Means algorithm**

To test the superiority of the improved K-Means algorithm in sample clustering, Iris, Seeds, and Wine data sets were selected for clustering algorithm performance testing. Using Average Between-Within Proportion (ABWP) and Average Improved Between-Within Proportion (AIBWP) to evaluate the clustering effect of the improved k-Means algorithm. The Between-Within Proportion formula is Formula 18, and the Improved Between-Within Proportion formula is Formula 19. Then calculate the average of the BWP and IBWP values of all samples. The larger the average values, the better the clustering effect of the clustering algorithm on the samples.

\[
BWP(j, i) = \frac{b(j, i) - w(j, i)}{b(j, i) + w(j, i)} \tag{18}
\]

\[
IBWP(j, i) = \frac{ib(j, i) - iw(j, i)}{ib(j, i) + iw(j, i)} \tag{19}
\]

In the above formula, \( b(j, i) \) is the distance between classes, which is the minimum average distance between the sample \( i \) in the \( j \) class and the samples of other classes; \( w(j, i) \) is the distance within the class, which is the average distance between sample \( i \) in class \( j \) and other samples of this class. \( ib(j, i) \) is the distance between classes, which is the minimum distance between the sample \( i \) in the \( j \) class and the center points of other classes, \( iw(j, i) \) is the distance within the class, which is the average distance between sample \( i \) in class \( j \) and other samples of this class.

The intra-class distance \( w(j, i) \) in BWP can well reflect the degree of connection between the sample and other objects in the same class, and the inter-class distance \( b(j, i) \) can well reflect the relationship between samples and samples in other classes. The larger \( b(j, i) \), the better the separation between classes, and the smaller \( w(j, i) \), the tighter the distribution of samples within the classes. In IBWP, the inter-class distance \( ib(j, i) \) of sample \( i \) can better reflect the relationship between the sample and other classes. If sample \( i \) is far away from the nearest class, it will be farther away from other classes. In addition, If the sample point \( i \) is not classified into class \( j \), it must be classified into the class closest to it. Compared with BWP, IBWP can better evaluate the clustering effectiveness of a single sample.

It can be seen from Fig. 5 that the improved K-means algorithm proposed in this paper is more stable when measuring index values on different data sets. AIBWP value is more stable than the ABWP value, indicating that the improved K-means algorithm clustering result is more stable. At the same time, the value of AIBWP
Table 2  Classification results of GA-based SVM algorithm on different data sets

| data set      | C     | g    | accuracy (%) | time(s) |
|---------------|-------|------|--------------|---------|
| Wine          | 14.620| 0.096| 100.000      | 12.830  |
| Spect Heart   | 1.295 | 6.589| 83.333       | 61.211  |
| Balance Scale | 84.640| 0.037| 93.333       | 113.348 |

is greater than that of ABWP, prove that the improved K-means algorithm in this paper has a high accuracy rate for sample clustering.

SVM classification test based on GA parameter optimization

In the initial algorithm, $C \in [0.1, 100]$, $g \in [0.01, 10]$, the initial population size of the GA-SVM algorithm was set as 60, crossover probability as 0.4, mutation probability as 0.01, and the maximum number of iterations was set as 100. The fitness value was output every five times, and the SVM parameter optimization algorithm constructed in this paper was used for classification experiments on the Wine, Spect Heart, and Balance Scale data sets. The test results are shown in Table II and Fig. 6.

It can be seen from Table I that the GA-SVM algorithm proposed in this paper can output the optimal parameters $c$ and $g$, and the classification accuracy of different data sets is higher than 80%. As can be seen from Fig. 6, the GA-SVM algorithm proposed in this paper after 30, 10, and 20 iterations in the Wine, Spect Heart, and Balance Scale data sets, the average fitness value gradually stabilizes. The results show that the GA-SVM algorithm can achieve intelligent optimization of $c$ and $g$ parameters and has good performance.

Image defogging effect based on machine learning

To verify the effect of the haze image defogging algorithm based on machine-learning proposed in this paper. This paper takes the image containing haze as the research object, then selects the He algorithm [20], Mei algorithm [5], Liu Haibo algorithm [21], and Choi algorithm [22] to compare and analyze the effect of image defogging. As shown in Fig. 7.

As can be seen from Fig. 7, there are no gray areas and bright areas in the input original haze image, and almost all defogging algorithms are related to the dark channel prior. The transmittance calculation method is similar, but the atmospheric light was acquired in different ways. Therefore, similar results are obtained when these algorithms are used to remove fog.

However, the Choi algorithm, He algorithm, Liu algorithm, and Mei algorithm are not obvious in image defogging processing. Besides, Choi’s algorithm showed that distant scenes and leaves were darker in color, indicating that the algorithm lost details during the haze image defogging process. The He algorithm handles the near scene better, while the haze on the far scene does not. The image colors of Liu algorithm and Mei algorithm are different from the real scenes that people are subjectively familiar with, and the phenomenon of image color super-saturation is more serious. Based on (f) image is shown in Fig. 7. The algorithm in this paper can not only protect the detailed information of the image during the defogging process but also prevent the color distortion in the image. Experiments show that compared
with other algorithms, the defogging algorithm based on machine learning training proposed in this paper can effectively complete the defogging of haze images.

To better evaluate the performance of the algorithm in this paper, the following five indexes are used to quantitatively evaluate the images after the fog removal of different algorithms.

1. The average gradient \( G \) is the average value of all points on the image gradient map, reflecting the tiny details and texture changes of the picture, and also reflecting the sharpness of the image. The larger the average gradient value, the sharper the image.

\[
G = \frac{1}{(M-1)(N-1)} \sum_{i=1}^{M-1} \sum_{j=1}^{N-1} \sqrt{\left( F(i,j) - F(i+1,j) \right)^2 + \left( F(i,j) - F(i,j+1) \right)^2 / 2} 
\]

(20)

In the above formula, \( F(i,j) \) is the gray value of \((i,j)\) pixels; \( M \) is the height of the image; \( N \) is the width of the image.

2. Image entropy \( H \) reflects the average signal amount of an image. The larger the value of \( H \), the more detailed information of the image.

\[
H = \sum_{i=0}^{255} \frac{f(i,j)}{M \cdot N} \log \frac{f(i,j)}{M \cdot N} 
\]

(21)

In the above formula, \( f(i,j) \) is the frequency of occurrence of the 2-tuple \((i, j)\).

3. The standard deviation \( \sigma_g \) represents the degree of dispersion between the pixel value and the mean value of the image. The larger the \( \sigma_g \) value, the better the image quality.

\[
\sigma_g = \sqrt{\frac{1}{M \cdot N} \sum_{i=1}^{M} \sum_{j=1}^{N} (p(i,j) - \mu)^2 } 
\]

(22)

In the above formula, \( p(i,j) \) represents the pixel value in row \( i \) and column \( j \), and \( \mu \) represents the mean value.

4. The ratio of the number of visible edges \( e \) is the comparison of the number of visible edges before and after image defogging and represents the change in image richness after image defogging.

\[
e = \frac{n_r - n_0}{n_0} 
\]

(23)

In the above formula, \( n_0 \) is the number of visible edges in the original image; \( n_r \) is the number of visible edges in the image after defogging.

5. The normalized gradient mean of visible edges \( \bar{r} \) is the comparison of the average gradient value of the image before and after defogging, which represents the change of the image clarity before and after the haze image processing.

\[
\bar{r} = \frac{g_r}{g_0} 
\]

(24)
In the above formula, $\overline{g_0}$ is the average gradient value of the original image; $\overline{g_r}$ is the average gradient value of the image after defogging.

Using $G$, $H$, $\sigma_g$, $e$, $\tau$ five indicators to quantitatively analyze the defogging effect of the haze image defogging algorithm, the results are shown in Table II. It can be seen that the values of $G$, $H$, $\sigma_g$, $e$ and $\tau$ are the largest in the image after defogging by the algorithm in this paper. It shows that the algorithm in this paper can not only improve the clarity and color vividness of the image, but also enhance the saturation and contrast of the image in the defogging of the haze image.

![Table 3](image)

**Table 3** Comparison of quantitative indexes of haze image defogging effect of different algorithms

| algorithms   | AveGrad | $H$   | $\sigma_g$ | $e$   | $\tau$ |
|--------------|---------|-------|------------|-------|--------|
| Choi         | 11.308  | 7.630 | 79.078     | 2.151 | 2.026  |
| Mei          | 13.329  | 7.735 | 79.037     | 2.110 | 2.135  |
| He           | 5.667   | 7.336 | 73.668     | 2.138 | 1.863  |
| LiuHaibo     | 7.326   | 7.430 | 66.384     | 2.453 | 2.037  |
| This paper   | 14.289  | 7.778 | 79.312     | 2.039 | 2.519  |

**Conclusions**

To improve the defogging effect of haze images and overcome the disadvantage of traditional prior defogging methods that are difficult to capture certain features of haze images, this paper proposes a color correction and restoration of haze images based on HSI color space and machine learning. This article first obtains a large number of haze images from the Internet to form a haze image library; secondly uses the HIS color space haze image restoration algorithm to correct the smog image to obtain the haze image transmittance library; then based on the density parameter optimization k-means algorithm and GA-SVM algorithm to build a machine learning haze image defogging model, use the image defogging model to train the haze image transmittance library, and obtain many visual words representing different characteristic transmittance maps to form a visual dictionary. When defogging the haze image, first use the Gaussian pyramid block sampling, and then find the visual word corresponding to the haze image block to obtain the coarse transmittance map; then use the guided filter to refine the coarse transmittance map. Obtain the transmittance map of the haze image. Finally, according to the method of obtaining atmospheric light value proposed in this paper, combined with the atmospheric scattering model, a clear image is obtained. Through subjective comparison and objective evaluation, experiments prove that the method proposed in this paper can effectively improve the defogging effect of haze images, and enhance the clarity, detail, and color effects of the image. The haze image restoration and correction algorithm based on his color space and machine learning is effective for haze image defogging. However, this algorithm is aimed at a single haze image, which has great limitations in practical application. In the future, we plan to study real-time haze affirmative restoration and correction based on the theory of this paper.

**Abbreviations**

HSI: hue-saturation-intensity; GA: genetic algorithm; SVM: Support Vector Machine; ABWP: Average Between-Within Proportion; AIBWP: Average Improved Between-Within Proportion

**Availability of data and materials**

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.
Competing interests
The authors declare that they have no competing interests.

Funding
This work was financially supported by the Key R&D project of Shandong Province of China (Grant No. 2019JZZY020326).

Authors’ contributions
Our contributions in this paper were that the first author (Haibin Sun) participated in the designing of the scheme and drafted the manuscript. The second author (Haiwei Liu) conceived of the study, participated in the design, and helped to draft the manuscript. All authors read and approved the final manuscript.

Acknowledgements
Not applicable.

Authors’ information
Haibin Sun received the Ph.D. degree in computer application technology from Jilin University in 2006. He is currently a Professor with the College of Computer Science and Engineering, Shandong University of Science and Technology. He has authored about 10 research papers in various international journals and proceedings of conferences. His research interests include internet of things, embedded system.

Haiwei Liu received the bachelor’s degree in Internet of Things Engineering from Qilu University of Technology in 2018. He is currently pursuing the master’s degree with the College of Computer Science and Engineering, Shandong University of Science and Technology. His research interests include internet of things, sensor technology, artificial intelligence.

Author details
College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao, China.

References
1. Wang, X., M, Z., BJ, H.: Research on traffic flow dynamic control strategy based on mit in bad weather. J. Wuhan Univ. Technol. (2) (2018)
2. Pei, S.C., Chen, Y.A., Chiu, Y.H.: Single image desmoking using haze image model and human visual system. J. Electron. Imaging 28(4), 1 (2019)
3. Kuo, Y.T., Chen, W.T., Chen, P.Y., Li, C.H.: VLSI implementation for an adaptive haze removal method. IEEE Access PP, 1–1 (2019)
4. Dong, N.: Research and application of image enhancement technology in traffic scene of fog and haze. J. Interdiscip. Math. 21(5), 1297–1301 (2018)
5. Mei, Y., Ning, Y.: Multilayer fusion and chunk-based transmittance estimation for natural hazy image enhancement. IEEE Access (2019)
6. Wen, L., Ju, Y., CL, Z., HF, W.: Algorithm of fog removal of image based on improved kawahara filter. Control. Eng. China 026(005), 997–1002 (2019)
7. Chaudhry, A.M., Riaz, M.M., Ghafoor, A.: A framework for outdoor rgb image enhancement and dehazing. IEEE Geoscience. Remote. Sensing. Lett. 15(6), 932–936 (2018)
8. Zhang, Q., Xiang-Yang, D.U., Wang, F.: Vehicle plate recognition system based on labview and hsi color space and mathematical morphology. Control Instruments in Chemical Industry (2019)
9. Trzepieciński, T., Lemu, H.G.: Improving prediction of springback in sheet metal forming using multilayer perceptron-based genetic algorithm. Materials 13(14), 3129 (2020)
10. Zyurt, F., Avc, E., Sert, E.: Uc-merced image classification with cnn feature reduction using wavelet entropy optimized with genetic algorithm. Traitement Du Signal 37(3), 347–353 (2020)
11. Zhao, C., Zhao, H., Wang, G., Chen, H.: Improvement svm classification performance of hyperspectral image using chaotic sequences in artificial bee colony. IEEE Access PP(99), 1–1 (2020)
12. Cha, Y., Malav, R., Pandey, G., Kim, A.: Dehazegran: Underwater haze image restoration using unpaired image-to-image translation. IFAC-PapersOnLine 52(21), 82–85 (2019)
13. Huang, S.C., Chen, B.H., Cheng, Y.J.: An efficient visibility enhancement algorithm for road scenes captured by intelligent transportation systems. IEEE Trans. Intell. Transp. Syst. 15(5), 2321–2332 (2014)
14. Yang, X., Li, H., Fan, Y.L., Chen, R.: Single image haze removal via region detection network. IEEE Trans. Multimedia., 1–1 (2019)
15. Abdullah, A.S.S., Abed, M.A., Al, B.I.: Improving face recognition by elman neural network using curvelet transform and hsi color space. Periodicals. Eng. Nat. Sci. 7(2), 430–437 (2019)
16. Yu, Shyr-Shen, Chu, Shao-Wei, Wang, Chiu-Mu, Chan, Yung-Kuan, Chang, Ting-Cheng: Two improved k-means algorithms. Appl. Soft Comput. (2018)
17. Karimi, F., Sultana, S., Babakan, A.S., Suthaharan, S.: An enhanced support vector machine model for urban expansion prediction. Comput. Environ. Urban Syst. 75(MAY), 61–75 (2019)
18. Devaraj, N.: Feature selection using genetic algorithm to improve svm classifier (2019)
19. Xiang, H., Han, Y., School, M.: Research on bbt travel time algorithm based on svm model based on genetic algorithm optimization. Logistics Sci-Tech (2019)
20. He, K., Sun, J., Tang, X.: Guided image filtering. IEEE Trans. Pattern Anal. Mach. Intell. 35(6), 1397–1409 (2012)
21. Liu, H., Yang, J., Wu, Z., Zhang, Q., Deng, Y.: A fast single image dehazing method based on dark channel prior and retinex theory. Acta Autom. Sin. 041(007), 1264–1273 (2015)
22. Choi, L.K., You, J., Bovik, A.C.: Referenceless prediction of perceptual fog density and perceptual image defogging. IEEE Trans. Image Process. (2015)

Figures
Figure 1 Optimization flow chart of k-Means algorithm based on density parameters.

Figure 2 Flow chart of SVM parameter optimization based on GA algorithm

Figure 3 Construction and training flow chart of haze image database

Figure 4 Comparison of color correction effects of haze images with different algorithms

Figure 5 Improved k-means algorithm index value detection results

Figure 6 Classification accuracy of GA-SVM algorithm on different data sets

Figure 7 Comparison of the effects of different algorithms on haze image dehazing