Image Stitching in Smog Weather based on MSR and SURF

Guanghong Li\textsuperscript{a}, Xuande Ji\textsuperscript{b,*}, and Ming Zhang\textsuperscript{c}

\textsuperscript{a}Engineering Training Center, Luoyang Institute of Science and Technology, Luoyang, 471023, China
\textsuperscript{b}School of Electrical Engineering and Automation, Luoyang Institute of Science and Technology, Luoyang, 471023, China
\textsuperscript{c}Luoyang Zhichao Mechanical and Electrical Technology Limited Company, Luoyang, 471000, China

Abstract

Image stitching can enlarge the range of viewing angles and increase different images information, and it is used in many fields such as industry, civil, and military. However, smog weather is an environmental problem in our country, because it can cause serious degradation of images. The loss of characteristic information will have negative impacts on the subsequent stitching process. Firstly, the smog image should be improved. In this paper, the application of the Multi-Scale Retinex (MSR) algorithm and the comparison and objective evaluation between it and the Histogram Equalization (HE) is discussed. Then, after removing the smog, the image is registered using local invariant features and the Speeded-up Robust Features (SURF) algorithm, and the Euclidean distance is adopted to obtain a satisfactory matching. Finally, the image stitching after registration may produce discontinuity of brightness in the overlapping area, and a higher quality stitching image can be achieved more quickly by using the progressive fade-out method. Through experiments and simulations, the smog images could be well stitched after removing the smog.

Keywords: multi-scale retinex; image stitching; speeded-up robust features; image fusion

(Submitted on June 8, 2018; Revised on July 12, 2018; Accepted on August 21, 2018)

© 2018 Totem Publisher, Inc. All rights reserved.

1. Introduction

With development of computer technology, machine vision, and artificial intelligence, digital image processing is playing a prominent role in different fields such as military reconnaissance, safety guard monitoring, intelligent traffic, and monitoring. However, sometimes the images cannot satisfy the requirements of large viewing angles and high resolution due to physical restriction of the cameras. Image stitching can obtain information with rich scenarios [1]. Currently, smog weather frequently occurs in China, and it is an unfavorable factor to puzzle development of the economy and society. Smog will lead to severe deterioration and distortion of the images, so the rich characteristic information in the fuzzy images will be lost fully. This will greatly influence further detection and identification, so it is very important to improve image quality in the smog environment. With the images improved, the subsequent processing can be implemented [2].

Smog elimination includes the technologies of image enhancement and image restoration. The image enhancement is used in this paper. The traditional image enhancement technologies include the histogram equalization, which can uniformly distribute a centralized gray area in the whole gray range, stretch the images in a non-linear manner, reassign the pixel values of the images, and make pixels roughly same within a certain grayscale. Different filters can be used to remove or weaken the noise interferences at the frequency domain, or the high-pass filter method is used to enhance high-frequency signals at the edge, so the fuzzy images become clearer. Retinex is the abbreviation of retina and cortex [3], and it is an image enhancement theory for the human vision system based on scientific experiments and scientific analysis. The basic opinion of the Retinex theory is that the object color depends on the reflection capabilities of the object to the lights with different wavelengths instead of the absolute strength of the reflected light. The object color is not affected by illumination heterogeneity but is consistent.

After removing the smog, the features of the images can be used for subsequent processing. The transformation matrix

* Corresponding author.
E-mail address: 39500448@qq.com
of the two images can be obtained by image registration and can then be projected onto the stitching picture. The different weighted coefficients processing in the overlapped region will achieve a better fusion image.

2. Improvement of the Smog Image

2.1. Application of the Retinex

Smog weather is a realistic problem. With development of industry, growth of urban populations, and increase of motor vehicles, smog weather has become more frequent and brings inconveniences to the shooting equipment and human visual system, including reduced image contrast, non-uniform gray, fuzzy boundaries, and hidden characteristic information.

The basic principal model of this algorithm is the earliest color theory, which was proposed by Edwin Land. An image enhancement method based on the color constancy was applied [4]. One given image \( S(x, y) \) is decomposed into reflected object image \( R(x, y) \) and incident light image \( L(x, y) \) according to the theory. Each point \((x, y)\) in the observed image \( S \) is described as \( S(x, y) = R(x, y)L(x, y) \), and the details of the method are given as follows:

**Step 1** The incident light component and reflected light component are separated by using the logarithm method, as shown in formula (1).

\[
S'(x, y) = r(x, y) + l(x, y) = \log(R(x, y)) + \log(L(x, y))
\]

(1)

**Step 2** The original images are convoluted by using the Gauss template, namely the image \( D(x, y) \) can be obtained after the low-pass filter of the original images. \( F(x, y) \) indicates the Gauss filter function.

\[
D(x, y) = S(x, y) \times F(x, y)
\]

(2)

**Step 3** The original images are subtracted by the low-pass filter images in the log domain to get the high-frequency enhanced image \( G(x, y) \).

\[
G(x, y) = S'(x, y) - \log(D(x, y))
\]

(3)

**Step 4** The inverse logarithm of \( G(x, y) \) is calculated to get the enhanced image \( R(x, y) \).

\[
R(x, y) = \exp(G(x, y))
\]

(4)

**Step 5** The contrast of \( R(x, y) \) is enhanced to get the final result image. Jobson et al. proposed the multiscale Retinex algorithm and its basic Equation, which is described as follows:

\[
R_i(x, y) = \sum_{n=1}^{N} W_n \left[ \log[I_i(x, y)] - \log[F_n(x, y) \times I_i(x, y)] \right]
\]

(5)

Where \( R_i(x, y) \) indicates the output of Retinex, \( i \in R, G, B \) indicates three color bands, \( F(x, y) \) indicates the Gauss filter function, \( W_n \) indicates the weight factor of the scale, and \( N \) indicates the number of the used scales. \( N \approx 3 \) indicates the color image, \( i \in R, G, B \). \( N=1 \) indicates the gray image. Here, we used grayscale images, as shown in Figure 1.
The above original image shows smog shielding and scattering, so the edge of the image scene becomes fuzzy and invisible and the quality of the whole image decreases, severely affecting further match and splicing. The profiles of the buildings in the processed image can be displayed and gray distribution is stretched in the histograms after processing.

2.2. Objective Assessment

To further measure the smog elimination quality of the image, this paper compares the average grayscale, information entropy, and standard deviation by using Retinex and the histogram equalization method [5]. The average grayscale shows the darkness of the images. An image with the smog has a lower image brightness due to smog. After processing, the average brightness is improved. The average grayscale is defined as follows:

\[
AVG_{gray} = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} f(x, y) / (M\times N)
\]  

The standard deviation of the image measures the distribution degree of the image’s average grayscale. A higher standard deviation indicates that most grays of the image are very different from the average. A lower standard deviation indicates that the image gray approximates the average. The calculation Equation is described as follows:

\[
\sigma = \sqrt{\frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} [f(x, y) - AVG_{gray}]^2}
\]

The average gradient shows the change rate of the image gray and can be used to assess the fuzziness of images. If the change rate of the grayscale is higher in one direction in an image, its gradient will become large. Generally, a clear image has a higher average gradient, so the average gradient can be used to measure definition of an image.

\[
AVG_{grad} = \frac{1}{(M-1)(N-1)} \sum_{i=1}^{M-1} \sum_{j=1}^{N-1} [f(x, y) - f(x+1, y)]^2 + [f(x, y) - f(x, y+1)]^2
\]

The information entropy of an image indicates the bit mean of the image gray set and describes the average information entropy of the image information source. The entropy indicates richness of the image details and is defined as follows:

\[
H = \sum_{i=0}^{255} p_i \log p_i
\]

<table>
<thead>
<tr>
<th></th>
<th>(AVG_{gray})</th>
<th>Standard deviation</th>
<th>(AVG_{grad})</th>
<th>Information entropy</th>
<th>(AVG_{grad})</th>
<th>Standard deviation</th>
<th>Information entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original a</td>
<td>157.126</td>
<td>0.174</td>
<td>0.003</td>
<td>6.699</td>
<td>Original b</td>
<td>168.720</td>
<td>0.163</td>
</tr>
<tr>
<td>HE</td>
<td>127.591</td>
<td>0.294</td>
<td>0.004</td>
<td>7.960</td>
<td>HE</td>
<td>127.440</td>
<td>0.292</td>
</tr>
<tr>
<td>Retinex</td>
<td>162.409</td>
<td>0.215</td>
<td>0.054</td>
<td>7.579</td>
<td>Retinex</td>
<td>179.039</td>
<td>0.207</td>
</tr>
</tbody>
</table>

The above table shows that the information entropy will increase after the smog images are enhanced, so it indicates an increase in the image information contents. Compared to the histogram equalization, the mean gradient of the Retinex will increase more significantly. This algorithm can sensitively reflect the image capability to represent minute detail contrast and also show minute detail contract and texture transformation characteristics.

3. Image Registration

The image registration indicates to match two or multiple images obtained at different times, from different equipment, or under different conditions (time, weather, illumination intensity, shooting position, angle, etc.). It can identify the corresponding geometric transformation parameters by establishing the corresponding relation between two images and match one target in two images. Massive research work has been conducted on image registration in numerous fields during recent several years, including pattern recognition, automatic navigation, remote sensing, medical diagnosis, and computer vision [6]. The image match methods are roughly divided into the grayscale-based method and characteristics-based
method. The grayscale-based match indicates to search the gray matrixes of the real-time image window with a certain size and all possible window gray matrixes of actual images pixel by pixel via a similarity measurement method and compare them. Generally, such methods require that the images are of the same source and the quality and computation is complex. The image registration based on these characteristics features stronger resistance to noises, illuminance, view change, repeatability, distinctness, and robustness.

3.1. Blob Detection

The first issue for image registration is to find the stable local invariant feature within the neighborhood of the feature points. The DOG is used to approximate to LoG in the Scale Invariant Feature Transform (SIFT) algorithm proposed by Lowe [7]. The SURF algorithm [8] operation rate is about three times that of SIFT in time. The SURF is highly robust in quality, its recognition rate of feature points is higher than that of SIFT, and its whole performance is higher than that of SIFT in case of change of view, illumination, and scale. Generally, different images have different scales and resolutions, so they shall be searched in different scale spaces. Lindeberg et al. prove that the Gauss kernel is the unique transformation kernel to transform the scale by using some precise mathematical models via different means. The scale space composed of the mathematical forms of different Gauss kernels is standardized and linear and can satisfy translation invariance, non-increasing local extremum, scale invariance, and rotation invariance.

The SURF algorithm identifies the position of the feature points by using the local maximum of the determinant of the approximate Hessian matrix. When the local values of the Hessian determinant reach the maximum, the detected points are the interested points, which are brighter or darker than the surrounding neighborhood. For a given point \((x, y)\) of the image \(f(x, y)\) and a given Hessian matrix \(H(x,\sigma)\) with the scale \(\sigma\), is the convolution of the Gauss Laplace operator and images with the scale \(\sigma\), as shown in Equations (10) and (11). Smard et al. proposed the concept of the box filter and replaces the convolution operation with the simplified Gauss second-order differential template. Viola and Jones simplified the convolution operation to the addition and subtraction operation of the integral images.

\[
H(x,\sigma) = \begin{bmatrix}
L_{xx}(x,\sigma) & L_{xxy}(x,\sigma) \\
L_{xy}(x,\sigma) & L_{xyy}(x,\sigma)
\end{bmatrix}
\tag{10}
\]

\[
det(H) = L_{xx}L_{yy} - L_{xy}^2 \approx D_{xx}D_{yy} - 0.9D_{xy}^2 \tag{11}
\]

The box filter and integral images are used, so the scale space can be expressed by using the indirect method of continuous enlarging of the box filter template instead of image zooming. The filter process will not increase the computation complexity with growth of the template size [9]. The determinant response images of the Hessian matrix can be obtained by using different box filter templates and integral images, and the spot of different scales can be obtained by using 3D non-maximum suppression on the response images.

The most famous descriptor is the SIFT based on the image gradient distribution, but it includes 128 dimensions. It can improve the error tolerance capabilities, and feature extraction and match are time-consuming. Here, the SURF descriptor used has 64 dimensions.

3.2. Identification of Principal Direction

To ensure rotation invariance, first, with the feature points as the center, the Harr wavelet (the edge of the Harr wavelet is 4\(s\)) response of the points within the neighborhood of \(6s\) radius (\(s\) is the scale of the feature points) in the \(x, y\) direction is calculated, and these response values are assigned with the Gauss weight coefficients to increase the contribution of the response close to the feature points, and decrease the contribution of the response remote from the feature points so it can comply with the objective conditions. Secondly, a 60° sector region is rotated with 0.2 radian as the step. The Harr wavelet response value \(dx\) and \(dy\) are added to form a new vector in this sector region, and the whole circular area is traversed to select the direction of the longest vector as the principal director of this feature point. Therefore, the principal direction of each feature point can be obtained by computing candidate feature points one by one, as shown in Equation (12), and the blobs detection is shown in Figure 2.

\[
\begin{align*}
\sum_m &= \sum_x \Delta x + \sum_y \Delta y \\
\theta &= \arctan(\sum_x \Delta x / \sum_y \Delta y) \max \{m_m\}
\end{align*}
\tag{12}
\]
3.3. Description of the Features

To generate the feature descriptor of the feature points, the Harr wavelet response of the images shall be computed. With the feature point as the center in a rectangular region, the 20s×20s images are divided into 4×4 blocks along the principal director. The response is calculated by using 2s Haar wavelet template for each block. Next, the response values are counted to form the feature vector, and the response values are assigned with the weight coefficients to increase the robustness of the geometric transformation [10]. Each block includes 25 sampling pixels. For each region, we shall accumulate dx and dy of 25 sampling pixels as one part of the descriptor. To include the polarity information of the strength change in the descriptor, we shall accumulate the absolute values of dx and dy. Each area can be expressed as

\[ V_{sub}(x, y) = (\sum d_x, \sum |d_x|, \sum d_y, \sum |d_y|) \]

by using a 4-direction feature vector. All 4×4 sub-areas can be combined to form a 64-dimension feature vector.

People have studied different feature descriptors and their corresponding similarity measurement methods. One excellent descriptor can be matched by using the simplest measure. Here, the Euclidean distance of the feature vector of the local area is used as the measure to determine the similarity of the interest point area in two images. For the feature point on the image a, the Euclidean distances from this feature point to all feature points on the image are computed to get a distance set. The distance set is compared and calculated to get the minimal Euclidean and secondary minimal Euclidean distance. Generally, a threshold within 0.6-0.8 is set. When the ratio of the minimal Euclidean distance to the secondary minimal Euclidean distance is less than this threshold, it indicates that the feature point is matched with the feature point of the corresponding minimal Euclidean. Otherwise, no point is matched to this feature point. To reduce this ratio threshold, the matched number will decrease, but the generated match points are more stable. Generally, the initially matched feature point pairs shall be refined. The frequent methods such as mutual guidance match, vote filter method, or binocular stereo geometry constraint are used to remove the outliers. Here the binocular stereo geometry constraint was used, with the outliers were eliminated as shown in Figure 3.

4. Estimation of the Transformation Matrix

After the corresponding point pairs of two images for registration are identified, the parameters of the geometric transformation matrix H between two images are solved with one image as the reference image and another image as the template image by using these matched point pairs. Next, the template images are normalized to the coordinate system of the reference images [11]. Given that the transformation between images is the projection transformation in this paper, the affine transformation model can generally be used for the close scenes.
The above Equation can be expressed with the nonhomogeneous form. The transformation matrix $H$ includes 8 independent coefficients, namely the freedom is 8. $H$ can be estimated by using 4 matched point pairs. A better mosaic effect can be obtained after the preliminary estimated $H$ is refined iteratively. The steps are described as follows:

**Step 1** The corresponding point $(x', y')$ in the image b, the error $e$ between the corresponding points $(e = f'(x', y') - f(x, y))$, and the partial derivative of different components to the error $e$ of $H$ are calculated for each feature point $(x, y)$ in the image a.

**Step 2** The increment function of $H$ is solved to get $\Delta h$, and $H$ is corrected.

**Step 3** The error $e$ is determined. If the error decreases but is not smaller than the set threshold, a new $\Delta h$ is further calculated. Otherwise, it will increase.

**Step 4** When the error $e$ is smaller than the regulated threshold, calculation stops and $H$ is obtained.

5. **Image Fusion**

After the space transformation relation between the image a and b is obtained, to get the composite images, the proper image fusion strategy shall be selected to stitch the images. Image fusion belongs to data information fusion and indicates the information processing to automatically analyze and integrate images. This is obtained under different conditions, according to certain rules by using computer image processing technology and completing the required decision and estimation tasks [12]. The corresponding images can be converted to identify the overlapping regions among images and map the fused images to a new blank image, forming the mosaic image. The brightness changes significantly at the image seam after stitching due to possible differences in brightness, field, and location of different images, as shown in Figure 4.

![Figure 4. Preliminary image fusion](image-url)
Image Stitching in Smog Weather based on MSR and SURF

\[
f(x, y) = \begin{cases} 
  f_a(x, y), & (x, y) \in f_a \\
  d_a f_a(x, y) + d_b f_b(x, y), & (x, y) \in f_a \cap f_b \\
  f_b(x, y), & (x, y) \in f_b 
\end{cases}
\]

(15)

Where \(d_a\) and \(d_b\) indicate the weight values of the images a and b in the overlapping area respectively. Generally they are related to the width of the overlapping area and \(d = \frac{1}{\text{width}}\), the width indicates the weight of the overlapping area, \(d_a + d_b = 1\), \(0 < d_a < 1\), and \(0 < d_b < 1\). \(d_a\) gradually changes from 1 to 0 and \(d_b\) gradually changes from 0 to 1 in the overlapping area, so the image a can gradually change to the image b in the overlapping area, as shown in Figure 5.

Figure 5. Processed image fusion

6. Conclusions

The smog weather severely deteriorates images. First, the smog in the images should be removed. Here, the Retinex algorithm can be used to obtain satisfactory images for further processing, and this is an indispensable and very important step. To find local invariant features in the images without smog, the images can be registered by describing features and selecting proper measures. The RANSAC algorithm can be used to iteratively estimate the transformation matrix of the matched images to achieve a more precise mosaic effect. The gradually emerging weighted fusion method can be used for brightness mutations in overlapping areas, obtaining natural stitching images with smooth transition.

Acknowledgements

This work was funded by Projects 2017YZY29 and 2017YZY30 and supported by the Pre-Research Fund of Luoyang Institute of Science and Technology.

References

5. Y. Han, X. Wang, and Q. Yang, “Research on Defogging Algorithm for Haze Degradation Image,” Science and Technology & Innovation, No. 4, pp. 28-29, 2017
July 2017


