



Article Target Search for Joint Local and High-Level Semantic Information Based on Image Preprocessing Enhancement in Indoor Low-Light Environments

Huapeng Tang ^{1,2}, Danyang Qin ^{1,2,*}, Jiaqiang Yang ^{1,2}, Haoze Bie ^{1,2}, Yue Li ¹, Yong Zhu ¹ and Lin Ma ³

- ¹ Department of Electronic and Communication Engineering, Heilongjiang University, Harbin 150080, China; 2211781@s.hlju.edu.cn (H.T.); 2211778@s.hlju.edu.cn (J.Y.); 2221716@s.hlju.edu.cn (H.B.); 2017021@hlju.edu.cn (Y.L.); 1997006@hlju.edu.cn (Y.Z.)
- ² National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China
- ³ Department of Electronics and Information Engineering, Harbin Institute of Technology,
- Harbin 150080, China; malin@hit.edu.cn
- Correspondence: qindanyang@hlju.edu.cn

Abstract: In indoor low-light environments, the lack of light makes the captured images often suffer from quality degradation problems, including missing features in dark areas, noise interference, low brightness, and low contrast. Therefore, the feature extraction algorithms are unable to extract the feature information contained in the images accurately, thereby hindering the subsequent target search task in this environment and making it difficult to determine the location information of the target. Aiming at this problem, a joint local and high-level semantic information (JLHS) target search method is proposed based on joint bilateral filtering and camera response model (JBCRM) image preprocessing enhancement. The JBCRM method improves the image quality by highlighting the dark region features and removing the noise interference in order to solve the problem of the difficult extraction of feature points in low-light images, thus providing better visual data for subsequent target search tasks. The JLHS method increases the feature matching accuracy between the target image and the offline database image by combining local and high-level semantic information to characterize the image content, thereby boosting the accuracy of the target search. Experiments show that, compared with the existing image-enhancement methods, the PSNR of the JBCRM method is increased by 34.24% at the highest and 2.61% at the lowest. The SSIM increased by 63.64% at most and increased by 12.50% at least. The Laplacian operator increased by 54.47% at most and 3.49% at least. When the mainstream feature extraction techniques, SIFT, ORB, AKAZE, and BRISK, are utilized, the number of feature points in the JBCRM-enhanced images are improved by a minimum of 20.51% and a maximum of 303.44% over the original low-light images. Compared with other target search methods, the average search error of the JLHS method is only 9.8 cm, which is 91.90% lower than the histogram-based search method. Meanwhile, the average search error is reduced by 18.33% compared to the VGG16-based target search method. As a result, the method proposed in this paper significantly improves the accuracy of the target search in low-light environments, thus broadening the application scenarios of target search in indoor environments, and providing an effective solution for accurately determining the location of the target in geospatial space.

Keywords: indoor low-light environments; location information; high-level semantic information; missing features in dark regions; noise interference; image preprocessing enhancement; target search; geospatial space

1. Introduction

With the emergence of new technologies, including computer networks and multimedia information processing, digital images have become the main way of obtaining



Citation: Tang, H.; Qin, D.; Yang, J.; Bie, H.; Li, Y.; Zhu, Y.; Ma, L. Target Search for Joint Local and High-Level Semantic Information Based on Image Preprocessing Enhancement in Indoor Low-Light Environments. *ISPRS Int. J. Geo-Inf.* 2023, *12*, 400. https://doi.org/10.3390/ijgi12100400

Academic Editors: Christos Chalkias, Marinos Kavouras, Margarita Kokla, Mara Nikolaidou and Wolfgang Kainz

Received: 22 August 2023 Revised: 24 September 2023 Accepted: 28 September 2023 Published: 30 September 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). information from people's daily lives, work, and studies because of their wide range of contents. Therefore, efficiently searching for the location information of the target image from visually information-rich images has been a focus of attention. However, images acquired in indoor low-light environments often suffer from quality degradation problems, such as missing features in dark areas and noise interference. These issues limit the effectiveness of feature extraction and matching algorithms, preventing them from providing application value in following target search tasks. As a result, how to improve the quality of low-light images in order to complete the subsequent target search task and then determine the target's location information is particularly critical.

In order to solve the quality degradation problem of low-light images, researchers have explored histogram equalization (HE)-based methods [1], Retinex-based methods [2], and deep-learning-based methods [3]. Although histogram equalization approaches can improve the contrast of low-light images, the enhanced images still have varying degrees of blurring, causing the number of extracted feature points to fall short of the requirements of the subsequent target search task.

In contrast to histogram equalization, methods based on Retinex theory divide the image into illumination and reflection components by a priori regularization or specific regularization, and use the estimated reflection component as the enhancement result, thereby reducing the loss of detail information in the original image. However, Retinex theory-based methods often ignore the treatment of noise during the enhancement process, which leads to a decrease in the accuracy of feature matching, thus reducing the accuracy of subsequent target searches.

Compared to the above two methods, the deep-learning-based method has made great progress in terms of accuracy, speed, and enhancement in low-light enhancement tasks. However, most deep-learning methods are still unable to balance noise control and luminance accuracy in the enhancement process, which makes the accuracy of feature extraction and matching degrade, thus reducing the accuracy of the target search. Aiming at the problem that the degradation of low-light image quality leads to the inability of the subsequent target search task, a joint local and high-level semantic information (JLHS) target search method is proposed based on joint bilateral filtering and camera response model (JBCRM) image preprocessing enhancement, as shown in Figure 1.

In the image preprocessing enhancement stage, the JBCRM image-enhancement method is composed of three parts: strengthening local features, denoising, and sharpening, in order to solve the problem of difficult feature extraction, thus providing better visual data for the subsequent target search task. Firstly, the indoor low-light images captured by a monocular vision camera (DJI Pocket 2, Shenzhen, China) are divided into an illumination component and a reflection component using the LECARM method. Secondly, the decomposed illuminance and reflection components are processed using the camera response model, thus obtaining the lighting-enhanced illuminance and reflection components, and these two components are fused to obtain the LECARM-processed image. Then, the OPJB filter is constructed by approximating the optimal parameters σ_d and σ_r through multiple experiments, thus rejecting the noise interference contained in the LECARM-processed image. Finally, the optimal parameter μ is chosen through several experiments to establish the OPUSM sharpening method, thereby accentuating the texture features of the denoised image. In the target search stage, the JLHS target search method consists of two parts: coarse search and fine search. Firstly, the target image and the offline database image are feature extracted using the rough search based on local feature SIFT (RLFS), respectively, and the corresponding feature vectors are generated and stored as corresponding npy files. Secondly, the BBF method is used to match the feature vectors in the offline database with the feature vectors of the target image, and the Euclidean distance is used to sort the matching results in descending order, so as to obtain the top six matching database images as the coarse search images. Then, the last layer of semantic information of each coarse search image and target image are extracted using the VGG16 fine search based on Keras (VFSK), respectively, and the corresponding feature vectors are generated and stored as

h5 files. Finally, the KD-Tree method is used to match the h5 feature vector corresponding to each coarse search image and the h5 feature vector corresponding to the target image, and the cosine similarity is used to sort the matching results in descending order, thereby obtaining the coarse search image that is most similar to the target image. The position information of this coarse search image is the position information of the target image.



Figure 1. JLHS target search method based on JBCRM image preprocessing enhancement.

The main contributions of this paper are as follows:

- (1) To address the issue of the difficulty in extracting feature information from low-light images, an image-enhancement method based on JBCRM is constructed. This method improves the image quality by strengthening the features of the dark region and reducing noise interference, so as to solve the problem of difficult feature information extraction.
- (2) Aiming at the problem that current target search methods are unable to balance accuracy and search time, a target search method based on JLHS is designed. By combining local feature scale-invariant feature transform (SIFT) with high-level semantic features for image description, the method increases the matching accuracy between the target

image and the offline database image, thereby improving the target search accuracy and reducing the target search time.

The remaining structure of this paper is as follows: Section 2 introduces the related work. Section 3 introduces the proposed JBCRM image-enhancement method, and the corresponding simulation experiments are carried out. Section 4 describes the designed JLHS target search method. Section 5 shows the simulation and analysis results of the target search. Section 6 summarizes the conclusions of the proposed method and prospects for future research directions.

2. Related Work

This section introduces two main parts: enhancement methods for low-light images and search methods for targets. (1) Current low-light image-enhancement methods are presented, which solve the problem of difficult feature information extraction by improving the image quality. (2) Existing target search methods describe the image content in different ways to improve the accuracy of feature matching, thus raising the accuracy of the target search.

2.1. Low-Light Image-Enhancement Methods

Currently, the methods to solve the difficulty of feature information extraction by improving the image quality are mainly classified into three categories: histogram equalizationbased methods, Retinex model-based methods, and deep-learning-based methods.

The methods based on histogram equalization improve the brightness and contrast of low-illumination images by expanding the grayscale range. Reference [4] proposes an extended method based on histogram equalization called contrast limited dynamic quadrihistogram equalization (CLDQHE). The method splits the whole histogram into four sub-histograms and performs adaptive histogram cropping, thus overcoming the defects of over-enhancement and over-smoothing in traditional histogram equalization methods. Reference [5] describes an adapted contrast enhancement using modified histogram (ACMHE). This method divides the histogram of the input image into four sub-histograms based on the brightness median. Then, independent histogram equalization is performed on each partition, resulting in natural contrast enhancement and brightness preservation. Although histogram equalization-based approaches increase the contrast and brightness of low-light images to varying degrees, they nevertheless suffer from noise interference and color distortion, which degrade the accuracy of feature extraction and matching, thereby decreasing the accuracy of the target search.

To further improve the quality of low-light images, researchers have successively proposed the multi-scale Retinex (MSR) method [6] and the multi-scale Retinex with color recovery (MSRCR) method [7]. In order to better preserve the structural information of the original low-light image, reference [8] proposes an image-enhancement method that combines Zero—DCE and Retinex. This method first utilizes the Retinex model to decompose the image into an illumination component and a reflection component. Then, the illumination component is enhanced by using deep light curve estimation, and the reflection component of the image is kept unchanged, so as to achieve the purpose of maintaining the structural characteristics of the image. Ref. [9] proposes a global attention-based Retinex network (GARN) for low-light image enhancement by embedding global attention modules in different levels of the network. However, Retinex-based methods produce unnecessary halo artifacts and noise interference, which degrade the accuracy of feature matching, thereby reducing the accuracy of the target search.

With the immense success of convolutional neural networks in various computer vision tasks, deep-learning-based methods have been widely used in the field of image enhancement. Reference [10] proposes a trainable convolutional neural network (CNN) called LightenNet for enhancing low-light images. The method takes the low-light images as the input, outputs their illumination components, and then obtains the enhanced image based on the Retinex model. Retinex-Net [11] improves the image brightness by using an

end-to-end image decomposition model and a continuous low-light enhancement network. Reference [12] presents a stacked sparse denoising autoencoder (SSDA) method to enhance low-light images. This method enhances the image by recognizing signal features in low-light images and adaptively enhancing the brightness of the image without overamplifying the brighter parts of the image. Despite significant progress in low-light imageenhancement tasks using convolutional neural network-based deep-learning methods, there are still issues such as a loss of detail information and noise interference that prevent the number of extracted feature points from being sufficient for subsequent target searches.

To solve these problems, this paper proposes an image-enhancement method based on a joint bilateral filtering and camera response model (JBCRM). This method enhances the quality of the image by highlighting details in the dark areas and removing noise interference, thereby solving the problem of difficult feature extraction and providing better visual data for subsequent target search tasks.

2.2. Target Search Methods

Existing target search methods can be classified into three main categories based on their search principles: text-based image retrieval (TBIR), content-based image retrieval (CBIR), and semantic-based image retrieval (SBIR).

The TBIR approaches mostly employ text annotation to add keywords to images, thereby completing the target search task. Ref. [13] proposes a target search method based on embedding and scene text. The first step of this method involves utilizing the maximally stable extremal region (MSER) algorithm to detect candidate text regions. Then, geometric features and stroke width transformations are used to eliminate unwanted false-positive text regions. Simultaneously, keywords are formed using a neural probabilistic language model, and the detected keywords are used to index and search the text images. Ref. [14] presents a hybrid text–visual correlation-based learning method. The method mines textual relevance from image tags, and then combines textual relevance and visual relevance to accomplish the search task. Although TBIR approaches have increased the target search accuracy to some level, image annotation is required to complete the search operation, which increases this method's manual expenditures.

The CBIR-based target search methods accomplish the search task mainly based on the features of the image content, thus avoiding the process of manually labeling the images. CBIR methods are mainly based on two types of visual features: local features and global features. The former captures underlying features from key points or salient blocks of an image. The latter considers the whole image as a salient region and convolves it, mainly including color [15], texture [16], and shape [17]. Compared to a local featurebased target search, the global feature-based target search method is relatively simple and computationally fast, but it is ambiguous, which means the semantic meanings expressed by images with similar features may be different, thus leading to a lower accuracy of the target search. The common method based on local features is the scale-invariant feature transform (SIFT) method [18], which generates 128-dimensional feature vectors for each key point. Meanwhile, the SIFT feature vectors are invariant to image scaling and rotation, with robustness to affine transformations, noise interference, and luminance transformations. [19] proposes a target search framework based on the VLAD model and speeded-up robust feature (SURF) descriptors. This framework converts 64-dimensional SURF descriptors into 8-dimensional SURF descriptors, and then constructs a codebook using a two-step clustering algorithm. After that, it uses an expandable overlapping segmentation method and a feature-fusion strategy to accomplish target search tasks. Although CBIR-based target search methods improve the accuracy of the target search to a great extent, they also face the problem of a semantic gap between low-level visual features such as color, texture, and shape and high-level abstract attributes such as emotion, filling, and expression in the human mind.

To improve the performance and accuracy of content-based target search methods, semantic gaps need to be reduced. With the advancement of machine learning and deep

learning in recent years, numerous SBIR-based target search approaches have been presented. These methods can reduce the semantic gap between the low-level features of the image and the high-level concepts in the human mind, and improve the accuracy of the target search. Ref. [20] designs a color attention function to describe the importance of different image blocks and combines color with texture to construct candidate regions. Meanwhile, it is input into the deep neural network (DNN) for feature extraction, and a similarity function is designed to calculate the distance between different images, where the top-ranked image is used as the searched image. Ref. [21] proposes a target search method that combines deep-learning semantic feature extraction and regularized Softmax. The method first constructs the convolution depth Boltzmann machine (C-DBM) by combining the deep Boltzmann machine (DBM) and the convolutional neural network (CNN). Then, the Dropout regularized Softmax classifier is used to classify the image features, and the image is searched based on the sorted output. Ref. [22] presents a semantic target search method that fuses the visual saliency model with the bag-of-words model. This method uses a visual saliency-based segmentation method to segment the image into background regions and foreground targets. Then, multiple features, including SIFT features, are extracted and fused from the background region and foreground target, respectively. Meanwhile, the fusion z-score normalized chi-squared distance is used as the similarity measure to complete the target search. Although this method has a better target search performance, the computational complexity of segmentation is still large, and the performance of segmentation has a significant impact on the search performance. Allani et al. [23] propose a target search system that fuses semantic and visual features. The system automatically builds a modular ontology for semantic information and organizes visual features in a graph-based model. These two elements are then combined in a component called "pattern" for subsequent target retrieval. Chen et al. [24] present a method based on deep image search called deep semantic hashing (DSH). This method considers the visual and semantic features of the image based on deep learning and uses the semantic information to generate the hash function of the hash code, thus improving the accuracy of the subsequent target search. Although the target search accuracy is greatly increased by SBIR-based methods, they are still challenging owing to the limitations of present artificial intelligence and related technology. In order to improve the target search accuracy and shorten the search time, this paper constructs a joint local and high-level semantic information (JLHS) target search method. By combining the local feature SIFT with the high-level semantic feature, this method increases the feature matching accuracy of the target image and offline database images, thereby improving the precision and decreasing the search time for the target search.

3. JBCRM Image-Preprocessing-Enhancement Method

This section introduces two main parts: (1) Introducing the proposed imagepreprocessing-enhancement method, JBCRM, in this paper. (2) Demonstrating the simulation and analysis of each mainstream image-enhancement method and the JBCRM method.

3.1. Construction of JBCRM Image-Enhancement Method

Under low-light conditions, the lack of light makes the captured images often suffer from quality degradation problems such as missing dark areas, low brightness, low contrast, noise interference, and color distortion. These problems make it difficult for the feature extraction algorithms to extract feature information from the image, thus failing to meet the number of feature points required for subsequent target searches. To address these issues, this paper improves LECARM [25] by introducing a new denoising model, known as a denoising model based on joint bilateral filtering and unsharp masking (JBUSM), in order to build a low-light image-enhancement method based on a joint bilateral filtering and camera response model (JBCRM). The specific steps of the JBCRM image-preprocessingenhancement method are as follows. According to the Retinex model, the illuminance arriving at the camera is first divided into the illuminance component and the reflection component.

$$G = Z \times F, \tag{1}$$

where *Z* and *F* are the illuminance and reflection components, respectively. *G* is the amount of illumination reaching the camera, which is also known as scene irradiance.

Next, the camera's nonlinear process is described using the camera response function (CRF), which explains the link between image irradiance *G* and low-light image *L*, as shown in Equation (2):

$$L = f(G), \tag{2}$$

where *f* represents the nonlinear function CRF.

According to Equation (2), Equation (1) can be written in the following form:

$$L = f(Z \times F), \tag{3}$$

The irradiance *G* of an image produces a nonlinear transformation in many cases due to the nonlinear processing of the camera. Therefore, the mapping function between different exposure images can also be a nonlinear function, which is called the brightness transform function (BTF). The BTF describes the relationship between the L_0 and L_1 of two images taken under different exposures in the same scene, as shown in Equation (4):

$$L_1 = g(L_0, k), (4)$$

where *g* represents the BTF function and *k* denotes the exposure rate.

The CRF and BTF are the fundamental components of the camera response model, which describe the basic properties of image processing in the camera. Based on the definitions of CRF and BTF, the relationship between two images taken at different exposures of the same scene is represented by Equation (5):

$$f(k \times G) = g(f(G), k), \tag{5}$$

The equation is known as a parametric equation, which describes the relationship between f and g and can be used to convert between the two functions.

As a result, Equation (6) can be used to calculate the enhanced image L^e of a lowilluminance image *L* captured by the same camera in the same scene:

$$\mathcal{L}^e = f(F \times 1), \tag{6}$$

where 1 denotes a matrix whose all elements are 1 [26]. Based on Equations (1) and (5), the relationship between L and L^e can be derived as shown in Equation (7):

$$L^e = f(F) = f(G \times (1 \otimes Z)) = g(f(G), (1 \otimes Z)),$$

$$\tag{7}$$

where \oslash stands for the division of elements. Equation (7) adjusts the exposure of the input image *L* to produce the illumination-enhanced image *L*^{*e*}. As a result, the output image *L*^{*e*} can be written as follows:

$$L^{e} = g(f(G), (1 \oslash Z)) = g(f(G), k_{0}),$$
(8)

where exposure k_0 is a matrix representing the required exposure per pixel.

1

Then, Equation (9) is used to remove the noise from the illumination-enhanced image L^e to obtain the denoised image L_j^e . The pixel value of any point p in image L^e after the filtering process is $L_i^e[p]$.

$$L_{j}^{e}[p] = \frac{\sum_{y \in \Omega} F_{\sigma_{d}}(\parallel p - q \parallel) \times G_{\sigma_{r}}(\parallel D_{p} - D_{q} \parallel) \times L_{q}^{e}}{\sum_{y \in \Omega} F_{\sigma_{d}}(\parallel p - q \parallel) \times G_{\sigma_{r}}(\parallel D_{p} - D_{q} \parallel)},$$
(9)

among them,

$$F_{\sigma_d}(\|p-q\|) = exp\Big(-\Big[(x-u)^2 + (y-v)^2\Big]/2\sigma_d^2\Big),$$
(10)

$$G_{\sigma_r}(\parallel D_p - D_q \parallel) = exp\left(-\left[\left(D_p - D_q\right)^2\right]/2\sigma_r^2\right),\tag{11}$$

where L^e is the input image and Ω is the set of neighborhoods of the center pixel p. The coordinates of point p are (x, y) and the coordinates of point q are (u, v). D_p and D_q are the pixel values corresponding to the guide image position (x, y) and location (u, v), respectively. F and G represent the spatial domain filter centered on (x, y) and the value domain filter centered on (x, y), respectively. σ_d is the standard deviation of the spatial domain, which is used to adjust the weight values of pixels with larger spatial distances. σ_r is the standard deviation of the similarity factor controlling the gray range, which is used to adjust the weight values of pixels with larger pixel differences.

Finally, the denoised image L_j^e is sharpened by using Equation (12), thus resulting in the JBCRM-enhanced image L_{ibl} .

$$L_{jbl}(x, y) = L_{j}^{e}(x, y) + \mu \times \left[L_{j}^{e}(x, y) - Q(x, y) \right],$$
(12)

where μ is the enhancement factor. Q(x, y) is the low-pass template with the expression:

$$Q(x, y) = \frac{1}{M \times N} \sum_{i=x-(M-1)/2}^{x+(M-1)/2} \sum_{j=x-(M-1)/2}^{y+(N-1)/2} L_j^e(x, y),$$
(13)

where $M \times N$ is the size of the template and M = N.

3.2. Design of JBUSM Denoising Model

In order to better remove the noise interference and retain more structural information of the original image, the JBUSM denoising model consists of two parts: the optimal parameter-based joint bilateral filter (OPJB) and the optimal parameter-based USM-sharpening method (OPUSM). The specific contents of the JBUSM denoising model are as follows.

3.2.1. Construction of the OPJB Filter

As can be observed from Equation (9), the denoising effect of the joint bilateral filter depends on the parameters σ_d and σ_r . The larger the parameter σ_d , the better the noise reduction effect. The smaller the parameter σ_r , the better the noise reduction effect. To select the optimal parameters σ_d and σ_r for processing low-illumination images, this paper uses multiple experiments to approximate the optimal parameters σ_d and σ_r , thus building the OPJB filter. The details are as follows: firstly, two images are selected from the mainstream low-light datasets (LIME [27], LOL [28], MEF [29], SICE [30], GladNet [31], and actual scene images), including global low-light images and local low-light images. At the same time, the selected images are combined into a new low-light dataset (NLLD) containing different scenes, as shown in Table 1. Then, LECARM is employed to process the images in the NLLD. Finally, multiple experiments are used to approximate the optimal parameters σ_d and σ_r .



The steps for determining the optimal parameter σ_d are shown below: firstly, σ_r is kept constant. Then, σ_d is taken in the range of 10–100 at intervals of 10, and the corresponding joint bilateral filters are formed. Finally, these filters are utilized to process the LECARM-enhanced image. Meanwhile, the peak signal-to-noise ratio (PSNR) and Laplace operator are introduced as evaluation metrics. For the images processed by different σ_d , the corresponding evaluation indicators are shown in Figure 2.



Figure 2. Evaluation indicators corresponding to different σ_d ; they should be listed as (**a**) PSNR corresponding to different σ_d ; (**b**) Laplace operators corresponding to different σ_d .

As can be seen from Figure 2, the trend of the corresponding evaluation indexes after different σ_d treatments is as follows: the value of the Laplace operator tends to increase between 10 and 30, which means that the clarity of the image is increasing. Between 30 and 100, the value of the Laplacian operator shows a downward trend, indicating that the clarity of the image continues to decline. Through the analysis, the PSNR value shows an increasing trend in the range of 10~100, which indicates that the noise interference in the image is continuously decreasing. Considering these results, this paper selects $\sigma_d = 30$ as the optimal parameter to maintain the high definition of the image.

Similarly, the optimal parameter σ_r is determined as follows: firstly, fix σ_d = 30. Then, σ_r is taken at intervals of 10 in the range of 5–50 to form the corresponding filters. Finally, these filters are utilized to process the LECARM-enhanced image. Meanwhile, the PSNR and Laplace operators are utilized as evaluation metrics. The comparison results of each evaluation metric are shown in Figure 3.

As can be seen from Figure 3, between 5 and 15, the value of the PSNR decreases as σ_r increases, indicating a decrease in the denoising performance of the image. Between 15 and 50, as σ_r increases, the PSNR value remains constant, indicating that the denoising performance reaches a stable state. Between 5 and 10, as σ_r increases, the value of the Laplacian operator continuously increases, indicating an improvement in the clarity of the image. However, between 10 and 50, the Laplace transform tends to become unstable as σ_r increases. Combining the results of the above analysis, $\sigma_r = 5$ is selected as the optimal parameter in this paper for the better removal of noise interference from the image. Concurrently, the OPJB filter is constituted by combining $\sigma_d = 30$, as mentioned above.

Table 1. Composition of the NLLD.



Figure 3. Evaluation metrics corresponding to different σ_r ; they should be listed as (**a**) PSNR corresponding to different σ_r ; (**b**) Laplace operators corresponding to different σ_r .

In order to evaluate the denoising effect of the OPJB filter, the PSNR and structural similarity index metric (SSIM) are introduced as evaluation indexes in this paper. For the NLLD, the average values of the PSNR and SSIM in the images before and after denoising using the OPJB filter are shown in Table 2.

Images before and after Denoising	LECARM-Processed Images	OPJB-Denoised Images		
PSNR	11.56	11.80		
SSIM	0.34	0.37		

Table 2. Average value of each evaluation index before and after denoising of NLLD.

From the comparison results in Table 2, the corresponding PSNR and SSIM values of the LECARM-enhanced images after OPJB denoising are significantly improved by 2.08% and 8.82%, respectively. These evaluation indicators show that the OPJB filter also eliminates the noise interference contained in the LECARM-enhanced image to a certain extent while maintaining the integrity of the original image structure information.

3.2.2. Construction of OPUSM-Sharpening Method

Since the OPJB filter inevitably removes the texture details in the image when removing noise, it is necessary to sharpen the denoised image to highlight the edge and texture details of the image. To address this problem, this paper proposes the OPUSM-sharpening method by highlighting the details of the image, thus further improving the visual effect of the

image. As can be seen from Equation (12), the sharpening effect of the USM method depends on the size of parameter μ . The larger the μ , the better the sharpening effect and the richer the details. In order to select the optimal parameter μ for constituting the OPUSM-sharpening method, this paper uses several experiments to approximate μ . The specific construction process of the OPUSM method is as follows.

Firstly, LECARM is used to process the images in the NLLD to obtain the illuminationenhanced images. Secondly, the illumination-enhanced images are denoised using the OPJB filter. Then, μ is taken at intervals of 10 in the range of 10–50, thus constituting the corresponding USM-sharpening methods. Finally, the OPJB-denoised images are processed separately using these USM-sharpening methods. Meanwhile, the PSNR and Laplace operators are introduced as evaluation metrics. The comparison results of each evaluation index are shown in Figure 4.



Figure 4. Evaluation indicators corresponding to different μ in the range of 10–50; they should be listed as (**a**) PSNR corresponding to different μ ; (**b**) Laplace operators corresponding to different μ .

As can be seen from Figure 4, with the increase in μ , the PSNR value of the image decreases and the Laplace operator value increases, indicating that the noise interference of the image increases and the detail information increases. Therefore, the μ should not be too large when sharpening the image so that the noise is not amplified. Based on the above analysis results, this paper takes the value between 1 and 9 at the interval of 1 near $\mu = 10$, thus forming the corresponding USM-sharpening method to approximate the optimal parameter μ . Then, these USM methods are used to sharpen the image after OPJB denoising. At the same time, the PSNR and Laplace operators are introduced as evaluation indexes. The comparison results of each evaluation index are shown in Figure 5.



Figure 5. Evaluation indicators corresponding to different μ in the range of 1–9; they should be listed as (a) PSNR corresponding to different μ ; (b) Laplace operators corresponding to different μ .

As can be seen from Figure 5, between 1 and 9, the PSNR value of the image stabilizes around 11.80. The corresponding PSNR value is 11.80 for both μ = 6 and μ = 7. However,

 μ = 6 corresponds to a larger Laplace operator as compared to μ = 7. Therefore, μ = 6 is selected as the optimal parameter in this paper, thus constituting the OPUSM-sharpening method. Meanwhile, it is constructed as a JBUSM denoising model together with the OPJB filter mentioned above.

In order to evaluate the clarity of denoised images after sharpening by the OPUSM method, the Laplace operator is introduced as an evaluation index in this paper. For the NLLD, the average values of the Laplace operator in the images before and after sharpening by the OPUSM method are shown in Table 3.

Table 3. Average values of the Laplace operator before and after image sharpening in NLLD.

Images before and after Sharpening	OPJB-Denoised Images	OPUSM-Sharpened Images		
Laplacian operator	1481.22	2078.82		

As shown in Table 3, the average Laplacian value of the OPJB-denoised images increases dramatically after sharpening by the OPUSM method, increasing by 40.35%. This shows that the OPUSM-sharpening method has significantly improved the clarity and contrast of the OPJB-denoised images.

3.3. Simulation and Analysis of JBCRM Image-Enhancement Method

To scientifically analyze the JBCRM image-enhancement method suggested in this paper, the images in the NLLD are processed using MF [32], NPE [33], LIME [34], Al-Ameen [35], Dong [36], and the JBCRM method, respectively. Meanwhile, the PSNR, SSIM, Laplace operator, universal quality index (UQI), and mean square error (MSE) are introduced as evaluation metrics. The image-enhancement effect of each algorithm is shown in Table 4, and the comparison result of each evaluation index is shown in Table 5.

Table 4 shows that the images processed by MF and Al-Ameen exhibit blurred edge information. The images processed by NPE and Dong exhibit a slight halo and noise interference. LIME-processed images have more noise and less detail. The JBCRM image-enhancement approach suggested in this paper produces images with richer features, clearer texture structures, and more "realistic" colors when compared to existing image-enhancement methods.

For the four metrics, PSNR, SSIM, Laplace operator, and UQI, the larger the value, the less image noise, the more similar to the original image structure information, the higher the clarity, and the better the quality. For the MSE evaluation index, the smaller its value, the higher the image contrast. As shown in Table 5, compared with other image-enhancement methods, the PSNR of images processed by the JBCRM increased by 34.24% at the highest and 2.61% at the lowest. The SSIM has increased by 63.64% at most and 12.50% at least. The Laplace operator improved by a maximum of 54.47% and a minimum of 3.49%. The UQI has increased by 43.75% at most and 4.55% at least. The MSE has decreased by 46.66% at most and 0.91% at least. Combining the results of the aforementioned analyses, the JBCRM-enhanced low-light images have a higher quality, higher clarity, less noise, and structural information that is more akin to the original image. When compared to previous image-enhancement techniques, the JBCRM method presented in this paper takes the least amount of time. As a result, the JBCRM image-enhancement method proposed in this paper can significantly improve the quality of low-light images while also shortening the time required for image enhancement, thus providing better visual information and reducing the time required for image preprocessing for subsequent target search tasks.

To assess the impact of the JBCRM method proposed in this paper on feature extraction, four feature extraction algorithms commonly used for target search are used to extract features from the original image and the JBCRM-enhanced low-light image, respectively. These feature extraction algorithms are SIFT, oriented fast and rotated BRIEF (ORB), accelerated-KAZE (AKAZE), and binary robust invariant scalable key points (BRISK). The number of feature points for each feature extraction algorithm are shown in Table 6.

 Table 4. Visualization effect of the NLLD after enhancement by each algorithm.

Image-Enl Algo	hancement rithm	Original Image	MF	NPE	LIME	Al-Ameen	Dong	JBCRM
	Global low-light							
LIME	Local low-light							
	Global low-light					अन्यहरू तो		
LOL	Local low-light							
MEE	Global low-light	a series of the						
WIEF	Local low-light							
SICE	Global low-light							
SICE	Local low-light	YA TOK	Xn Strong	Yn Rom				
	Global low-light							
Gladivet	Local low-light							
Actual	Global low-light	Dist						
scene	Local low-light	1 A						

Table 5. Evaluation metrics for the NLLD enhanced by various algorithms.

Image-Enhancement Algorithm	MF	NPE	LIME	Al-Ameen	Dong	JBCRM
PSNR	11.80	10.89	8.79	9.42	11.50	11.80
SSIM	0.32	0.30	0.22	0.26	0.32	0.36
Laplacian operator	1911.25	1613.16	1756.32	2008.72	1320.16	2078.82
UQI	0.23	0.22	0.16	0.16	0.21	0.23
MSE	4895.78	6132.78	9094.23	8085.09	5068.32	4851.21
Average image processing time/s	0.32	1.37	0.74	0.35	0.77	0.028

Feature Extraction Method	SIFT	ORB	AKAZE	BRISK
Original images	116.42	152.83	48.83	160.67
JBCRM-enhanced images	331.00	184.17	197.00	573.42
Growth rate of feature points	184.32%	20.51%	303.44%	256.89%

Table 6. Number of feature points in the original and JBCRM-enhanced images.

From Table 6, compared to the original images, the number of feature points in the JBCRM-enhanced images increases at the most by 303.44% and at the least by 20.51%. As a result, the low-light images enhanced by the JBCRM have a substantial increase in the number of feature points during feature extraction using the feature extraction algorithm, thus providing a sufficient number of feature points for the subsequent target search.

4. Construction of Target Search Method Based on JLHS

This section presents two main parts: (1) Introducing the process of constructing the JLHS target search method. (2) Simulation and analysis of the existing target search method and JLHS method in the real scenario. The details are as follows.

Currently, global features and local features are the main methods for characterizing image content in CBIR-based target search methods. Compared with the former, the latter can more appropriately characterize the feature information contained in images. The most widely used local features are SIFT local features, which produce 128-dimensional feature vectors for each key point. Compared with speeded-up robust features (SURF) and ORB local features, SIFT feature vectors are resistant to affine transformations, noise interference, and luminance transformations, and are unaffected by image scaling and rotation. As a result, SIFT local features are widely used in the field of target search. However, SIFT feature-based target search methods have a low accuracy and must be combined with other methods to increase target search precision. The success of deep-learning-based methods has provided an effective solution for this. Although deep-learning-based target search methods have a better search accuracy, they take a lot of time and computer resources. To improve the search accuracy and shorten the search time, a joint local and high-level semantic information (JLHS) target search method is proposed in this paper. This method consists of two parts: a rough search based on local feature SIFT (RLFS) and a VGG16 fine search based on Keras (VFSK). The specific construction process of the JLHS search method is shown in Figure 6.



Figure 6. Structure of JLHS target search method.

In indoor low-light environments, the JBCRM image-enhancement method is used to preprocess the acquired image, and then, the JLHS method is used to search the image. The specific steps are as follows:

 In the offline feature database generation stage, firstly, a monocular vision camera (DJI Pocket 2) is used to collect low-light images of the selected experimental site, and the corresponding position coordinates of the images are recorded to form a low-light image database. Then, the images in the low-light database are preprocessed using the JBCRM image-enhancement method, thus obtaining the JBCRM-enhanced image database. Finally, the images in the JBCRM-enhanced image database are feature extracted using the feature extraction algorithm SIFT to form SIFT feature vectors, thus constructing an offline feature database.

- (2) In the query stage, the target image is first taken using the monocular vision camera (DJI Pocket 2) at the same experimental site. Then, the JBCRM image-enhancement method is used to preprocess the target image. Finally, the feature extraction algorithm SIFT is used to extract the features of the JBCRM-enhanced image to form the SIFT feature vector, and the SIFT feature vector is stored.
- (3) For the target search, the offline feature database's SIFT feature vectors and the JBCRM-enhanced images' SIFT feature vectors are first matched using the RLFS coarse search method, thus yielding the coarse search images with the top six numbers of matched points. Then, the last layer of convolutional features of the coarse search images and the last layer of convolutional features of the JBCRM-enhanced images are compared using the VFSK fine search technique. By arranging the results in descending order based on the cosine similarity, the most similar database image to the target image is obtained. This database image's position coordinate is the target image's position coordinate.

4.1. Construction of Offline Feature Database

In order to scientifically evaluate the effectiveness and feasibility of the JLHS target search method, this paper selected indoor corridors during morning and evening hours as the experimental site. In Figure 7, the center of Figure a is taken as the origin of the world coordinates, its horizontal direction is taken as the x-axis of the world coordinate system, and its vertical direction is taken as the y-axis of the world coordinate system. Meanwhile, five acquisition points are selected in the x-axis direction, and fifteen acquisition points are selected in the y-axis direction. Each acquisition point acquires images at 90° intervals in the clockwise direction, and a total of 300 images are acquired. The corresponding position coordinates of each image are recorded, thus forming a low-light image database. Then, the JBCRM image-enhancement method is used to preprocess the low-light database, thus obtaining the JBCRM-enhanced image database. Finally, the feature extraction algorithm SIFT is used to extract the SIFT features of the images in the JBCRM-enhanced image database. Then, the form of npy files, thereby completing the construction of the offline feature database. The process of building the offline feature database is shown in Figure 7.



Figure 7. Construction of offline feature database.

4.2. Constructing RLFS-Based Coarse Search Technology

Aiming at the problem of the long search time of the traditional SIFT target search methods, this paper uses the best bin first (BBF) search method to match the target image with the offline database image. At the same time, this paper uses the Euclidean distance as



the similarity measure of key points in two images to reduce the probability of a mismatch. The specific steps of the RLFS-based coarse search technique are shown in Figure 8.



Firstly, the JBCRM image-enhancement method is used to preprocess the image in the low-light image database and the target image, respectively, thus obtaining the corresponding JBCRM-enhanced image database and JBCRM-enhanced image. Secondly, the SIFT feature extraction algorithm is utilized to extract features from images in the JBCRM-enhanced image database, thereby acquiring the set of SIFT feature points and forming the feature vectors, which are saved as npy files. Simultaneously, the SIFT feature extraction algorithm is used to extract features from the JBCRM-enhanced image, yielding the corresponding SIFT feature points and constructing the feature vector, which is stored as an npy file. Thirdly, the SIFT feature vectors of the JBCRM-enhanced image database are matched with the SIFT feature vector corresponding to the JBCRM image by using the BBF search method, respectively. Then, the Euclidean distance is used to calculate the similarity between the JBCRM-enhanced image and each image in the JBCRM-enhanced image database, and the similarity is sorted in descending order. Finally, the top six similarity images in the JBCRM-enhanced image database are selected as the coarse search images. Among them, the details of the SIFT feature vector construction process are as follows.

4.2.1. Establishment of Scale Space and Detection of Extreme Points

The whole scale space establishment process of the image is as follows: for an image of size $N \times N$, the image is convolved with the Gaussian kernel, thereby obtaining Gaussian spaces of different scales, which are expressed as:

$$L(x, y, \sigma) = G(x, y, \sigma) \times I_D(x, y),$$
(14)

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2 + y^2)/2\sigma^2},$$
(15)

where $G(x, y, \sigma)$ is the Gaussian function with variable parameters; σ is the scale space factor; $L(x, y, \sigma)$ is the spatial function at a specific scale.

To obtain stable Gaussian scale space extreme points, the original image is convolved with *G* with different scale factors. Then, the images of two adjacent Gaussian spaces are subtracted to obtain the difference of Gaussians (DOG), thus eliminating the unstable edge points, whose mathematical expression is:

$$D(x, y, \sigma) = [G(x, y, k\sigma) - G(x, y, \sigma)] \times I_D(x, y) = L(x, y, k\sigma) - L(x, y, \sigma),$$
(16)

To ensure the stability and uniqueness of the SIFT features, each sampling point on the DOG needs to be compared with 8 neighboring points at the same scale, as well as 18 points corresponding to the neighboring scales above and below, which are 26 points in total. If the DOG value of this sampling point is greater than or less than the other 26 points, the point is set as a local extreme point.

4.2.2. Precisely Determine the Location of the Feature Points

In the candidate set of local extremum points of scale space, there are many lowcontrast and unstable edge points, which directly affect the stability and anti-interference ability of matching. Therefore, these edge points need to be removed to improve the accuracy of matching. The specific removal principle is as follows: the principal curvature value is relatively large in the direction of the edge gradient, while the principal curvature value is small along the edge direction. The principal curvature value of the candidate feature points is proportional to the eigenvalue of the 2×2 Hessian matrix. The expression of the Hessian matrix is:

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{yx} & D_{yy} \end{bmatrix},$$
(17)

Let α and β be the eigenvalues of H, and the value of α is greater than β . At the same time, let $\alpha = r\beta$, and then, the trace Tr(H) and determinant Det(H) of H are as follows:

$$Tr(H) = D_{xx} + D_{yy} = \alpha + \beta, \tag{18}$$

$$Det(H) = D_{xx}D_{yy} - (D_{xy})^2,$$
 (19)

$$Ratio = \frac{(Tr(H))^2}{Det(H)} = \frac{(\alpha + \beta)^2}{\alpha\beta} = \frac{(r+1)^2}{r},$$
(20)

If *Ratio* $\leq \frac{(r+1)^2}{r}$, it is retained as the feature point; otherwise, it is discarded.

4.2.3. Direction Distribution of Feature Points

To ensure the rotational invariance of the feature points, it is necessary to assign a principal direction for each feature point based on the magnitude and direction of the gradient. The specific process of determining the main direction of feature points is as follows: firstly, the direction of each feature point is calculated. Then, the gradient information of the pixels around the feature point is counted, and the corresponding gradient histogram is plotted at 45-degree intervals. Finally, the peak of the gradient histogram is selected as the principal direction of the feature point.

For the scale space image $L(x, y, \sigma)$, the size and direction of the gradient at each feature point are as follows:

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2},$$
 (21)

$$\theta(x, y) = tan^{-1} \left(\frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)} \right),$$
(22)

where m(x, y) is the magnitude of the gradient and $\theta(x, y)$ is the direction of the gradient.

4.2.4. Generating SIFT Feature Point Descriptors

Through the above three processes, the position, scale, and direction information of the feature points are obtained successively. In order to improve the probability of the correct matching of feature points, it is necessary to establish corresponding feature descriptors for each feature point. The specific steps are as follows: firstly, rotate the coordinate axis to align with the main direction of the feature points mentioned above. Then, take a 16×16 window centered around the feature point within the same scale domain. Next, divide the window into 4×4 sub-block regions (seed points), as shown in Figure 9. Finally, the gradient histograms of each seed point in eight directions (every 45° is a direction) are counted based on Equations (21) and (22), and each gradient histogram is Gaussian weighted, so as to weaken the influence of the place far away from the feature points on the feature points.



Figure 9. Generation process of SIFT feature descriptor.

In the left figure, its center position is the position of the feature point, and each cell represents a pixel in the scale space where the neighborhood of the feature point is located. The arrow in each small box corresponds to the direction of the gradient at the feature point, the length of the arrow represents the gradient magnitude, and the circle indicates the range of Gaussian weights. In the right image, the gradient histogram of eight directions is drawn in each 4×4 box, and the cumulative value of each gradient direction is calculated, thus forming a seed point. Each feature point consists of 4×4 seed points, each with vector information in eight directions, thereby producing a 16×8 128-dimensional SIFT feature vector.

4.3. Constructing VFSK-Based Fine Search Technology

VGG16 was proposed by the Visual Geometry Group of the University of Oxford in 2014, and its specific structure is shown in Figure 10. VGG16 consists of five convolutional blocks and three fully connected layers. The first two convolutional blocks consist of two convolutional layers and a pooling layer, while the last three convolutional blocks consist of three convolutional layers and a pooling layer. In this paper, conv*i* is used to denote the *i*th convolutional block, conv*i_j* denotes the *j*th convolutional layer of the *i*th convolutional block. The number of filters in the 5 convolutional blocks is 64, 128, 256, 512, and 512, respectively. Compared with the traditional convolutional neural network model, the VGG16 network structure is very simple, which can enhance the richness and hierarchy of the feature representations, thus better capturing the visual features. Therefore, the VGG16 is chosen as the base model for the fine search in this paper.

After using the above RLFS coarse search, there is still the problem of mismatching, which leads to low accuracy in the target search. As a result, this paper builds the VGG16 fine search based on Keras (VFSK) in the fine search stage, thereby increasing the target search accuracy even more, as illustrated in Figure 10.

To improve the search accuracy and efficiency, VFSK fine search technology employs the VGG16 model to extract high-level semantic features from the conv5 layer, thus completing the accurate search of the target images. The details of the VFSK fine search technology are as follows: firstly, the convolutional feature mapping is extracted from the conv5 layer of each coarsely searched image and JBCRM-enhanced image, respectively, using the VGG16 model with a total number of channels of 512, thus constituting the corresponding h5 feature vectors. Then, the cosine similarity is used to calculate the similarity between the h5 feature vector corresponding with each coarse search image and the h5 feature vector corresponding with the JBCRM-enhanced image, and the obtained similarity is sorted using the K—means clustering method, thereby obtaining the coarse search image that has the highest similarity with the JBCRM-enhanced image. The position coordinates corresponding to this coarse search image are the position coordinates of the target image.





Among them, the calculation formula for the cosine similarity between two feature vectors is as follows:

$$\cos(\theta) = \frac{a \times b}{\|a\| \times \|b\|},\tag{23}$$

where *a* and *b* are two different h5 feature vectors. $cos(\theta)$ is the cosine similarity of the two h5 feature vectors, which ranges from [-1, 1]. The larger the cosine value, the more similar the two h5 feature vectors are represented.

5. Simulation and Result Analysis of Target Search

In order to evaluate the JLHS target search method proposed in this paper, 178 lowlight images are taken at any position of the selected experimental site as experimental images for the target search. The specific content of the target search experiment is as follows: firstly, the 178 experimental images captured are preprocessed using the JBCRM method, thereby obtaining the corresponding JBCRM-enhanced images. Then, Höschl IV [37], Yin [38], Chhabra [39], Kopparthi [40], and the JLHS method proposed in this paper are used to match the JBCRM-enhanced images with the images in the offline database, respectively, thus obtaining the database image that is most similar to the target image. The position coordinates corresponding to this database image are the position coordinates of the target image. The performance comparison results of each target search method are shown in Table 7.

Table 7. Performance comparison results of each target search method (* represents the process including image preprocessing and target search).

Search Method	Höschl IV	Yin	Chhabra	Kopparthi	Ours
Average search error/m	1.21	0.26	0.31	0.12	0.098
Average deviation angle/°	3.29	0.90	0.89	0.83	0.52
Average search time/s	1.31	0.98	0.57	0.58	0.36
Average whole process search time */s	1.34	1.01	0.60	0.61	0.39

As can be seen from Table 7, compared with other target search methods, the average search error of the JLHS method is reduced by 91.90% at most and 18.33% at least. The average search time is reduced by 72.52% at most and 36.84% at least. As a result, the JLHS method proposed in this paper significantly increases the target search accuracy while decreasing the search time.

In 178 sets of target search experiments, the accuracy rate of each target search method is shown in Figure 11. Among them, the formula for the search accuracy is as follows:

$$P = \frac{m}{N},\tag{24}$$

where *P* represents the accuracy rate, *m* represents the number of correctly searched samples, and *N* represents the number of searched samples.



Figure 11. Comparison of the accuracy of each target search method.

As can be seen from Figure 11, compared to the other four methods, the JLHS method proposed in this paper has the least number of erroneous search samples and the highest search accuracy. Meanwhile, compared to the other four approaches, the JLHS method improves the accuracy by 12.83% at most and 1.16% at least. Therefore, the JLHS method proposed in this paper effectively overcomes the problem of a difficult target search in indoor low-light environments.

6. Conclusions

Aiming at the difficulty of target searching in indoor low-light environments, this paper proposes a JLHS target search method based on JBCRM image preprocessing enhancement. The JBCRM approach solves the problem of difficult feature extraction and gives superior visual data for the succeeding target search task by enhancing the dark area features and eliminating noise interference during the image preprocessing stage. Compared to other image-enhancement techniques, the PSNR of the JBCRM-enhanced images is boosted by 34.24% at most and 2.61% at least. The Laplace operator is increased by 54.47% at most and 3.49% at least. From the evaluation metrics, the JBCRM-enhanced images have less noise, higher clarity, and more details. In terms of feature extraction, the maximum increase in the number of feature points in JBCRM-enhanced images is 303.44%, and the minimum increase is 20.51% as compared to the original low-light images. In the target search phase, the JLHS method designed in this paper improves the matching accuracy between the target image and the offline database image by combining the local feature SIFT and high-level semantic features to describe the image, thus boosting the target search accuracy. Compared with other target search methods, the average search error of the JLHS method is only 9.8 cm, and the average search time is only 360 ms. Experimental results demonstrate the effectiveness of the proposed method in the task of target searching in indoor low-light environments, which is able to obtain the position information of the target more accurately. In our future work, we will reduce the dimension of the local feature descriptor SIFT by using the effective dimension reduction algorithm, which improves the efficiency of SIFT feature extraction and further shortens the search time.

Author Contributions: Methodology, software, writing—original draft preparation, and writing—review and editing, Huapeng Tang; formal analysis, Huapeng Tang and Danyang Qin; resources and project

administration, Danyang Qin; supervision, Danyang Qin, Jiaqiang Yang, Haoze Bie, Yue Li, Yong Zhu, and Lin Ma. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Fundamental Scientific Research Funds of Heilongjiang Province (2022-KYYWF-1050), National Natural Science Foundation of China (61971162), and the Open Research Fund of National Mobile Communications Research Laboratory Southeast University (No. 2023D07).

Data Availability Statement: The data presented in this study are available upon request from the corresponding authors. The data cannot be made public for privacy reasons.

Acknowledgments: We thank Heilongjiang University for supporting the experimental scenes. We also gratefully thank the reviewers for their thorough review and are extraordinarily appreciative of their comments and suggestions, which have significantly improved the quality of the publication.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Ling, Z.; Liang, Y.; Wang, Y.; Shen, H.; Lu, X. Adaptive extended piecewise histogram equalisation for dark image enhancement. *IET Image Process.* 2015, 9, 1012–1019. [CrossRef]
- Wang, P.; Wang, Z.; Lv, D.; Zhang, C.; Wang, Y. Low illumination color image enhancement based on Gabor filtering and Retinex theory. *Multimed. Tools Appl.* 2021, 80, 17705–17719. [CrossRef]
- 3. Garg, A.; Pan, X.W.; Dung, L.R. LiCENt: Low-light image enhancement using the light channel of HSL. *IEEE Access* 2022, *10*, 33547–33560. [CrossRef]
- 4. Huang, Z.; Wang, Z.; Zhang, J.; Li, Q.; Shi, Y. Image enhancement with the preservation of brightness and structures by employing contrast limited dynamic quadri-histogram equalization. *Optik* **2021**, 226, 165877. [CrossRef]
- Santhi, K.; Banu, R.S. DW Adaptive contrast enhancement using modified histogram equalization. *Opt.-Int. J. Light Electron Opt.* 2015, 126, 1809–1814. [CrossRef]
- Rahman, Z.U.; Jobson, D.J.; Woodell, G.A. Multi-scale retinex for color image enhancement. In Proceedings of the 3rd IEEE International Conference on Image Processing, Lausanne, Switzerland, 19 September 1996; pp. 1003–1006.
- Jobson, D.J.; Rahman, Z.U.; Woodell, G.A. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Trans. Image Process.* 1997, *6*, 451–462. [CrossRef] [PubMed]
- 8. Krishnan, N.; Shone, S.J.; Sashank, C.S.; Ajay, T.S.; Sudeep, P.V. A hybrid low-light image enhancement method using Retinex decomposition and deep light curve estimation. *Optik* **2022**, *260*, 169023. [CrossRef]
- 9. Wang, Y.; Zhang, Z. Global attention retinex network for low light image enhancement. J. Vis. Commun. Image Represent. 2023, 92, 103795. [CrossRef]
- Li, C.; Guo, J.; Porikli, F.; Pang, Y. LightenNet: A convolutional neural network for weakly illuminated image enhancement. *Pattern Recognit. Lett.* 2018, 104, 15–22. [CrossRef]
- 11. Wei, C.; Wang, W.; Yang, W.; Liu, J. Deep retinex decomposition for low-light enhancement. arXiv 2018, arXiv:1808.04560.
- 12. Lore, K.G.; Akintayo, A.; Sarkar, S. LLNet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognit.* **2017**, *61*, 650–662. [CrossRef]
- 13. Unar, S.; Wang, X.; Zhang, C.; Wang, C. Detected text-based image retrieval approach for textual images. *IET Image Process.* 2019, 13, 515–521. [CrossRef]
- 14. Cui, C.; Lin, P.; Nie, X.; Yin, Y.; Zhu, Q. Hybrid textual-visual relevance learning for content-based image retrieval. *J. Vis. Commun. Image Represent.* **2017**, *48*, 367–374. [CrossRef]
- 15. Singha, M.; Hemachandran, K.; Paul, A. Content-based image retrieval using the combination of the fast wavelet transformation and the colour histogram. *IET Image Process.* **2012**, *6*, 1221–1226. [CrossRef]
- Varish, N. A modified similarity measurement for image retrieval scheme using fusion of color, texture and shape moments. *Multimed. Tools Appl.* 2022, *81*, 20373–20405. [CrossRef]
- 17. Pedrosa, G.V.; Batista, M.A.; Barcelos, C.A.Z. Image feature descriptor based on shape salience points. *Neurocomputing* **2013**, 120, 156–163. [CrossRef]
- Batur, A.; Tursun, G.; Mamut, M.; Yadikar, N.; Ubul, K. Uyghur printed document image retrieval based on SIFT features. *Procedia* Comput. Sci. 2017, 107, 737–742. [CrossRef]
- Kan, S.C.; Cen, Y.G.; Cen, Y.; Wang, Y.H.; Voronin, V.; Mladenovic, V.; Zeng, M. SURF binarization and fast codebook construction for image retrieval. J. Vis. Commun. Image Represent. 2017, 49, 104–114. [CrossRef]
- Zhu, H. Massive-scale image retrieval based on deep visual feature representation. J. Vis. Commun. Image Represent. 2020, 70, 102738. [CrossRef]
- 21. Wu, Q. Image retrieval method based on deep learning semantic feature extraction and regularization softmax. *Multimed. Tools Appl.* **2020**, *79*, 9419–9433. [CrossRef]
- 22. Bai, C.; Chen, J.N.; Huang, L.; Kpalma, K.; Chen, S. Saliency-based multi-feature modeling for semantic image retrieval. J. Vis. Commun. Image Represent. 2018, 50, 199–204. [CrossRef]

- Allani, O.; Zghal, H.B.; Mellouli, N.; Akdag, H. A knowledge-based image retrieval system integrating semantic and visual features. *Procedia Comput. Sci.* 2016, 96, 1428–1436. [CrossRef]
- Chen, C.; Zou, H.; Shao, N.; Sun, J.; Qin, X. Deep semantic hashing retrieval of remotec sensing images. In Proceedings of the IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 1124–1127.
- Ren, Y.; Ying, Z.; Li, T.H.; Li, G. LECARM: Low-light image enhancement using the camera response model. *IEEE Trans. Circuits* Syst. Video Technol. 2018, 29, 968–981. [CrossRef]
- Grossberg, M.D.; Nayar, S.K. Modeling the space of camera response functions. *IEEE Trans. Pattern Anal. Mach. Intell.* 2004, 26, 1272–1282. [CrossRef]
- 27. Liu, J.; Xu, D.; Yang, W.; Fan, M.; Huang, H. Benchmarking low-light image enhancement and beyond. *Int. J. Comput. Vis.* **2021**, 129, 1153–1184. [CrossRef]
- Xiong, W.; Liu, D.; Shen, X.; Fang, C.; Luo, J. Unsupervised Low-light Image Enhancement with Decoupled Networks. In Proceedings of the 2022 26th International Conference on Pattern Recognition (ICPR), Montreal, QC, Canada, 21–25 August 2022; pp. 457–463.
- 29. Shen, L.; Yue, Z.; Feng, F.; Chen, Q.; Liu, S.; Ma, J. Msr-net: Low-light image enhancement using deep convolutional network. *arXiv* 2017, arXiv:1711.02488.
- Cai, J.; Gu, S.; Zhang, L. Learning a deep single image contrast enhancer from multi-exposure images. *IEEE Trans. Image Process.* 2018, 27, 2049–2062. [CrossRef]
- Wang, W.; Wei, C.; Yang, W.; Liu, J. Gladnet: Low-light enhancement network with global awareness. In Proceedings of the 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), Xi'an, China, 15–19 May 2018; pp. 751–755.
- Fu, X.; Zeng, D.; Huang, Y.; Liao, Y.; Ding, X.; Paisley, J. A fusion-based enhancing method for weakly illuminated images. *Signal Process.* 2016, 129, 82–96. [CrossRef]
- Wang, S.; Zheng, J.; Hu, H.M.; Li, B. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Trans. Image Process.* 2013, 22, 3538–3548. [CrossRef]
- Guo, X.; Li, Y.; Ling, H. LIME: Low-light image enhancement via illumination map estimation. *IEEE Trans. Image Process.* 2016, 26, 982–993. [CrossRef]
- 35. Al-Ameen, Z. Nighttime image enhancement using a new illumination boost algorithm. *IET Image Process.* **2019**, *13*, 1314–1320. [CrossRef]
- 36. Dong, X.; Pang, Y.; Wen, J. Fast efficient algorithm for enhancement of low lighting video. In Proceedings of the ACM SIGGRAPH 2010 Posters, Los Angeles, CA, USA, 26–30 July 2010; Association for Computing Machinery: New York, NY, USA; p. 1.
- 37. Höschl, I.V.C.; Flusser, J. Robust histogram-based image retrieval. *Pattern Recognit. Lett.* 2016, 69, 72–81. [CrossRef]
- Yin, Y. Research on Image Similarity Retrieval Algorithm Based on Perceptual Hashing. Master's Thesis, Kunming University of Science and Technology, Kunming, China, 2020.
- Chhabra, P.; Garg, N.K.; Kumar, M. Content-based image retrieval system using ORB and SIFT features. *Neural Comput. Appl.* 2020, 32, 2725–2733. [CrossRef]
- 40. Kopparthi, S.; Nynalasetti, K.K.R. Content based image retrieval using deep learning technique with distance measures. *Sci. Technol. Hum. Values.* **2020**, *9*, 251–261.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.