

Research Article

Zhaosheng Xu*, Suzana Ahmad, Zhongming Liao, Xiuhong Xu, and Zhongqi Xiang

Image feature extraction algorithm based on visual information

<https://doi.org/10.1515/jisys-2023-0111>

received August 03, 2023; accepted October 10, 2023

Abstract: Vision is the main sensory organ for human beings to contact and understand the objective world. The results of various statistical data show that more than 60% of all ways for human beings to obtain external information are through the visual system. Vision is of great significance for human beings to obtain all kinds of information needed for survival, which is the most important sense of human beings. The rapid growth of computer technology, image processing, pattern recognition, and other disciplines have been widely applied. Traditional image processing algorithms have some limitations when dealing with complex images. To solve these problems, some scholars have proposed various new methods. Most of these methods are based on statistical models or artificial neural networks. Although they meet the requirements of modern computer vision systems for feature extraction algorithms with high accuracy, high speed, and low complexity, these algorithms still have many shortcomings. For example, many researchers have used different methods for feature extraction and segmentation to get better segmentation results. Scale-invariant feature transform (SIFT) is a description used in the field of image processing. This description has scale invariance and can detect key points in the image. It is a local feature descriptor. A sparse coding algorithm is an unsupervised learning method, which is used to find a set of “super complete” basis vectors to represent sample data more efficiently. Therefore, combining SIFT and sparse coding, this article proposed an image feature extraction algorithm based on visual information to extract image features. The results showed that the feature extraction time of X algorithm for different targets was within 0.5 s when the other conditions were the same. The feature matching time was within 1 s, and the correct matching rate was more than 90%. The feature extraction time of Y algorithm for different targets was within 2 s. The feature matching time was within 3 s, and the correct matching rate was between 80 and 90%, indicating that the recognition effect of X algorithm was better than that of Y algorithm. It indicates the positive relationship between visual information and image feature extraction algorithm.

Keywords: image feature extraction algorithm, visual information, feature extraction time, feature matching time, correct matching rate, scale invariant feature transform, sparse coding

* **Corresponding author: Zhaosheng Xu**, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Shah Alam Branch, Shah Alam, Selangor Darul Ehsan, 40450, Malaysia; Xinyu College, Xinyu, 338004, Jiangxi, China, e-mail: xuzhaosheng@xyc.edu.cn

Suzana Ahmad: Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Shah Alam Branch, Shah Alam, Selangor Darul Ehsan, 40450, Malaysia, e-mail: suzana@tmsk.uitm.edu.my

Zhongming Liao: Xinyu College, Xinyu, 338004, Jiangxi, China, e-mail: liaozhongming@xyc.edu.cn

Xiuhong Xu: Jiangxi New Energy Technology Vocational College, Xinyu, 338004, Jiangxi, China, e-mail: xxh258639049@163.com

Zhongqi Xiang: Xinyu College, Xinyu, 338004, Jiangxi, China, e-mail: lxsy20110927@163.com

1 Introduction

In modern computer vision systems, feature extraction and segmentation are two key issues, which together constitute the core of the whole vision system. In machine learning, pattern recognition, and image processing, feature extraction begins with an initial set of measurement data and establishes derived values (features) aimed at providing information and nonredundancy, thus facilitating the subsequent learning and generalization steps and bringing better interpretability in some cases. Image feature extraction is a process that uses certain methods to extract the local structure, content features, and global features of the image. Through feature extraction, the essential features with high information content, distinguishability, and high repeatability can be extracted from complex images.

There are three main methods of traditional image feature extraction: the method based on the statistical model, the method based on the artificial neural network, and the method based on knowledge. The artificial neural network has been a research hotspot in the field of artificial intelligence since the 1980s. It abstracts the human brain neural network from the perspective of information processing to establish a simple model and forms different networks according to different connection modes. The first two methods are relatively mature. However, most of these methods are based on theory. When faced with the problem of image segmentation, these methods seem inadequate. Therefore, many scholars have proposed different improved algorithms based on the analysis of the previous work.

Aiming at the problem of image segmentation, this article proposes an image feature extraction algorithm based on visual information. It uses an algorithm combining scale invariant feature transform (SIFT) and sparse coding to segment the image and extract its features. Then, by comparing the feature extraction time, feature matching time, and correct matching rate between the algorithm in this article and the conventional algorithm, it is concluded that visual information can improve the effect of image feature extraction algorithm. The innovation of this article is that visual information is a classic tool for two-dimensional image recognition and analysis, and the image feature recognition algorithm is to recognize and analyze images. The combination of the two not only makes the research significant but also has a new perspective and perfect idea, which points out a new direction for the development of image feature recognition algorithms in the future.

2 Related work

In computer vision, image recognition is an important research direction. It involves many fields such as mathematics, machine learning, image processing, and pattern recognition, and its research also has a lot of achievements. Liu *et al.* proposed an intelligent transportation system image extraction scheme based on threshold. The research results showed that this method provided a safe and intelligent image extraction method, which could be used for further analysis of the intelligent transportation system [1]. Ganji *et al.* developed a set of algorithms to extract architectural environment features from Google aerial and street view images, reflecting the microscopic features of urban location and different functions of buildings. The research results showed that the model based on feature extraction showed higher prediction ability, which highlighted the higher accuracy of the proposed method compared with the geographic information system layer based only on aerial images. The comparison with other neural network methods and traditional land use regression models showed the advantages of Bayesian regularized artificial neural network model in spatial interpolation of air quality [2]. After the disaster, the structural engineer team collected a large number of images from the damaged buildings to obtain new knowledge and learn from the event. However, in many cases, the collected images were captured without sufficient spatial background. When the damage is serious, it might even be difficult to identify the building. It is necessary to access the images of the predisaster conditions of these buildings to accurately identify the cause of failure or the actual loss of the buildings. To solve this problem, Lenjani *et al.* developed a method to automatically extract the building image in advance from the 360° panoramic image. To demonstrate the capability of this technology, taking the residential building in the holiday beach in Rockport, Texas, USA, as an example, he used the geographic marker images collected in the

actual postdisaster building survey mission. This method could successfully extract residential building images from Google Street View images captured before the event [3]. Hyperspectral images (HSI) are a powerful source of reliable data in various remote sensing applications, but most traditional frequency band selection methods cannot fully explain the interactions between spectral bands nor can they evaluate the representation and redundancy of selected wave band subsets. Therefore, Esmaeili first tested a supervised frequency band selection method that allows for the selection of the required number of frequency bands and embedded a three-dimensional convolutional layer deep network in genetic algorithms. The proposed method was evaluated, and satisfactory results were obtained. The accuracy has increased from 6 to 21%; the accuracy of each evaluation mode is between 90 and 99% [4]. The development of space-based hyperspectral sensors, advanced remote sensing, and machine learning can assist in crop yield measurement, modeling, prediction, and monitoring to prevent losses and global food security. However, accurate and continuous spectral features can only be provided through hyperspectral imaging, which is crucial for using cutting-edge algorithms for large-scale crop growth monitoring and early yield prediction. Therefore, Farmonov et al. used the new-generation Deutsche Zentrum für Luft und Raumfahrt Earth Induction Imaging Spectrometer (DESI) images to classify the main crop types of Mez Å heyas. Research has shown the potential of DESI data in observing the growth and predicting the harvest of different crop types, which is crucial for farmers, small farmers, and decision-makers [5]. HSIs provide rich information in various applications due to their unique features, but some issues reduce the accuracy of HSI classification. In order to improve accuracy in the case of limited training samples, Ghaderizadeh proposed a multiscale dual branch residual spectral space network that focuses on human-machine interaction classification models and named it MDRSSN. Applying the advanced feature of multiscale abstract information extracted by convolutional neural networks to image processing can improve the classification accuracy of complex hyperspectral data [6]. However, the aforementioned studies are all surface analysis for image recognition, and the lack of research on image feature extraction algorithm. Therefore, a scientific method is urgently needed for verification.

Aiming at the aforementioned problems, the use of visual information to analyze image feature extraction algorithms has become a hot topic for increasingly scholars and has been studied in a large number of related fields. Among them, Gao et al. proposed a hierarchical short-term and short-term memory method with adaptive attention for image and video captions. At the same time, he designed a hierarchical long-term and short-term memory network, taking into account low-level visual information and high-level language context information to support title generation. The experimental results showed that the method achieved the most advanced performance in most evaluation indicators of the two tasks [7]. The extraction of key visual information from images containing natural scenes is a challenging task and an important step for visually impaired people to recognize information based on tactile graphics. Yoon et al. proposed a saliency region extraction method based on global contrast enhancement and saliency cutting to improve the recognition process of visually impaired people. The research results showed that this method was conducive to extracting significant objects from images containing natural scenes, generating simple but important edges, and providing information for visually impaired people [8]. These studies have illustrated the applicability of visual information in the field of image and laid a solid foundation for combining it with image feature extraction algorithms.

3 Image feature extraction algorithm of visual information

3.1 Introduction of relevant technologies

3.1.1 Visual information processing process

Both biological vision and machine vision face the same processing target, namely, visual information. Its application field is shown in Figure 1. Machine vision is a branch of artificial intelligence that is developing

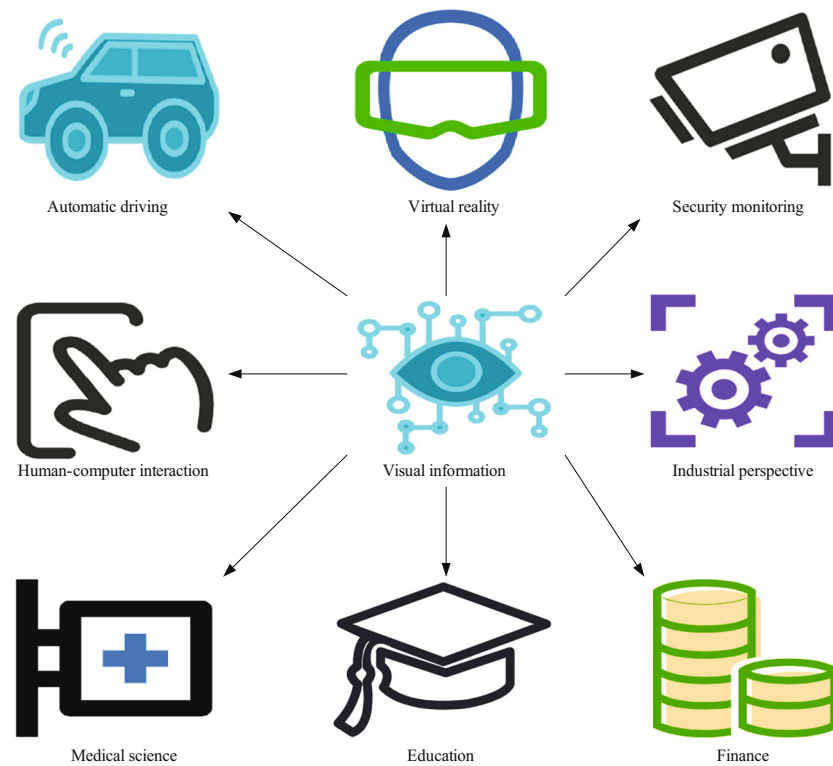


Figure 1: Application field of visual information.

rapidly. In short, machine vision is to use machines instead of human eyes to measure and judge. There are many descriptions of visual information, including brightness, color, and shape [9,10]. However, brightness is essential information. Human eyes can only recognize in the visible area. At the same time, the difference in light intensity also greatly affects the information obtained by the visual system. With the basic information of brightness, the color, shape, and motion state of objects are also essential. For example, it is difficult to find a frog still in the grass. However, if it keeps beating, even if it is wearing protective colors, people can easily find it. For a man walking on a snowy mountain in a white ski suit, his whereabouts are not easy to be found. However, if he changed into a red ski suit, it might immediately attract people's attention. The research shows that the object appears obvious through the stimulation of various information. Generally, moving objects are more likely to cause the response of the visual system. Color stimulation is a subjective feeling. Existing research shows that cone cells can be divided into three categories in vision. Each type has a strong response to visible light of a specific spectrum, and these three kinds of light are red, green, and blue. Therefore, the basic principle of three primary colors is not difficult to understand [11].

3.1.2 Image segmentation

To better extract the target features in the image, the image should be segmented first, as shown in Figure 2 [12,13]. For the target in the image, its pixel value is generally determined by its gray distribution function. Generally, the area with gray value less than the maximum gray value of the image is taken as the edge, and the area with gray value greater than the maximum gray value is taken as the target. Generally, the target pixel gray distribution function can be described by two parts. One is the mean function, which can be used to reflect the difference between the pixel value and the target gray value. The second is the standard deviation function, which can be used to reflect the difference between pixels. The pixel value is the value given by the computer when the original image is digitized. It represents the average brightness information of a small

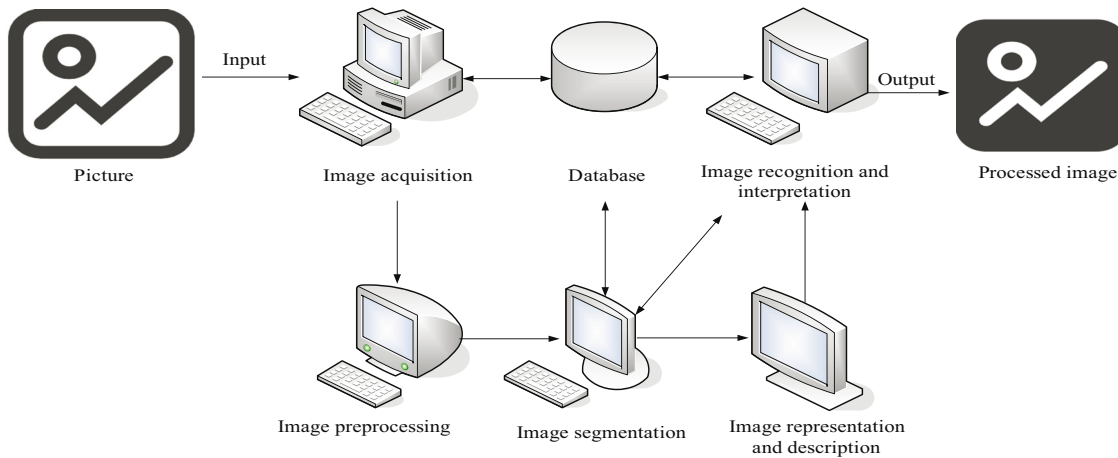


Figure 2: Process flow of image segmentation.

square in the original, or the average reflection (transmission) density information of the small square. To get better segmentation results, the median filtering algorithm can be used to denoise the image.

3.1.3 Local feature extraction

Local feature is one of the key factors of image segmentation, which is the description of a certain region in the image [14,15]. For the algorithm proposed in this article, more useful information can be extracted after image segmentation because it can avoid introducing additional calculation in region segmentation. In the region of interest, if the pixels in the neighborhood are Gaussian filtered and then these pixels are normalized, the corresponding normalized window becomes smaller. If the pixels in these neighborhoods have similarities with a certain pixel after Gaussian filtering, it can be used to describe the pixel, so that the features to be extracted after image segmentation can be obtained.

3.2 Advantages of visual information

3.2.1 Making full use of computer hardware resources

The feature extraction process of visual information is the analysis and processing of visual information, while the computer is responsible for processing data, storing data, and analyzing data. Therefore, in essence, it is a judgment of the utilization of computer resources [16,17]. Computer is a modern electronic computing machine for high-speed computing. It can perform numerical calculation, logical calculation, and memory storage. It is a modern intelligent electronic device that can automatically process massive data at high speed according to the program. There is a large amount of computation in the image feature extraction algorithm. For example, Canny operator in the edge detection operator needs a lot of data operations. For the image feature extraction process of visual information, it is not necessary to consider the amount of data operation, but only the need to consider the analysis and processing of image features to complete the feature extraction. The process of image feature extraction based on visual information does not need to consume a lot of computing resources and storage resources. Therefore, this method has the advantage of making full use of computer resources. In addition, the method is highly fault tolerant and reliable because the computer is used for calculation. If there is an error in this method, it causes problems such as too much computation for recalculation or replacement of this method, which affects the performance of the algorithm. However, a large number of calculations involved in the process of image feature extraction based on visual cues are all completed by computer, so it is

highly fault tolerant and reliable. Even if there are errors in the calculation results of this method due to some reasons, it does not cause the performance of the algorithm to decline or fail to work.

3.2.2 Independent of human visual physiological structure and characteristics

Because human visual physiological structure and features are different from machines, in the image feature extraction algorithm based on visual information, it does not depend on human visual physiological structure and features [18,19]. Human perception of the environment is not only a complex process but also an extremely complex and profound process. First, after receiving the information, human senses analyze and process it to form their own understanding of the world. Then, human beings form their own judgment through comprehensive analysis and comparison. In the image feature extraction algorithm based on visual information, it does not depend on human visual physiological structure and features, which means that this method is a kind of imitation of human visual physiological structure and features. When people do not know or observe the environment (such as in the dark environment), the results of the computer image feature extraction algorithm may be wrong or inaccurate. For example, when performing image segmentation, the result of image segmentation is wrong due to ignorance or errors in the observing environment. Therefore, this article uses a dynamic threshold method to determine the accurate threshold of each pixel in the image. The dynamic threshold is calculated according to the visual perception parameters (such as color, brightness, and contrast) of each pixel before the subjects segment the selected image. In the algorithm, these visual perception parameters of each pixel are dynamically updated, and the final test results are obtained.

3.3 Image feature extraction algorithm combining SIFT and sparse coding

Conventional image feature extraction algorithms mainly apply color feature information to dictionary learning when learning features, but do not make full use of the gradient information of objects [20,21]. At the end of the 20th century, people proposed a feature description method of SIFT, that is, to extract the gradient direction histogram of the object and sample it with high density to retain the information of the object in the gradient direction. SIFT feature algorithm not only has scale rotation invariance but also has a certain affine invariance, which can be used to obtain a stable image block feature representation. Therefore, this article combines SIFT and sparse coding to extract image features and proposes an image feature extraction algorithm based on visual information. First, the algorithm uses SIFT features to conduct centralized sampling of samples and then uses the obtained feature matrix for analysis. Second, K-singular value decomposition (K-SVD) algorithm is used for dictionary learning. In the learning process, the nearest neighbor method is used to obtain the sparse representation of the block, to preserve the details of the image as much as possible. Finally, the pyramid pool method is used to connect various features to obtain the feature representation of the image. Its advantage is that the dictionary obtained by directly learning the features of the pixel block using sparse coding can save the overall features of the pixel block (including the spatial position relationship between pixels, geometric features, and the relationship between pixel values, etc.). At the same time, with the help of the learned pixel block dictionary, the sparse coding of pixels can be obtained.

3.3.1 K-SVD dictionary learning

In this algorithm, in addition to using dictionary learning to obtain sparse coding for the color, depth, and normal vectors of the original block, SIFT features are constructed and sparsely coded according to the gray information of the image. Normal vector is a concept of spatial analytic geometry. The vector represented by a line perpendicular to a plane is the normal vector of the plane. The sparse coding is solved by the K-SVD update of the original pixel block. However, in the final sparsity coding, K-nearest neighbor is needed to solve the sparsity problem of SIFT features. The advantage of this method is that it can set larger sparsity to preserve

more detailed features. For the pixel block features that directly use the original pixel, the Orthogonal Matching Pursuit method is used to obtain the sparse representation of each pixel and set it to a low sparsity to maintain the most important features of each pixel block. As shown in formula (1), the dictionary F is learned from the data matrix B . A represents a sparse coding matrix. Dictionary learning can be achieved by minimizing the objective function.

$$\min_{F, A} \|B - FA\|_F^2 \text{ d.u. } \forall o, \|f_{co}\|_2 = 1, \|a_{cu}\|_0 \leq I_c. \quad (1)$$

In formula (1), the square term of the G norm is $\|\cdot\|_F^2$, and the zero norm is $\|\cdot\|_0$. They can be used to obtain the sparsity, that is, the number of nonzero elements. Each column in data matrix B has a pixel block feature or SIFT feature of a different information channel. In dictionary $F = [f_1, \dots, f_o, \dots, f_n] \in T^{q \times n}$, column vector f_o represents the o th word of the dictionary. Sparse matrix $A = [a_1, \dots, a_u, \dots, a_m] \in T^{n \times m}$ can be calculated from dictionary F and sample B . Among them, a_u is the sparse encoding of the corresponding b_u . Let $C = [D, V, F, M]$ be different features extracted from the image. Among them, D, V, F , and M are the feature matrix, dictionary and sparse coding corresponding to the SIFT feature, color, depth, and object surface normal vector information image block.

3.3.2 Extraction and representation of block features

3.3.2.1 Sparse coding based on block SIFT feature

Each image is collected with twenty 128-dimensional SIFT features to obtain feature matrix B_D . Then, dictionary F_D is iteratively calculated based on the SIFT feature matrix. To retain details as much as possible, the sparsity should be within $I_D = 200$.

$$\mu_D = \{\beta_{11}, \dots, \beta_{ok}, \dots, \beta_{nm}\}. \quad (2)$$

In formula (2), n and m are the number of SIFT sampling rows and columns. β_{ok} is the sparse coding of row o and column k of the sampling fast.

3.3.2.2 Image block feature calculation based on image pixel information

First, the pixels are sparsely encoded. The maximization method is used to obtain the feature expression of the image unit, and the feature expression of multiple image units is combined to form the block feature expression, to obtain the image block feature of the first layer [22,23]. On this basis, block features are used to learn the dictionary of the second layer and find the sparse coding of the corresponding block.

The sparse encoding of each pixel is set to $q_h = \{v_{1c}, \dots, v_{nc}\}$. Among them, $h \in \{1, 2, 3, \dots, 16\}$ represents the number of pixels contained in each cell. c is the characteristic information of color, depth, image, normal vector, etc. Let the expression of cell characteristics be:

$$vg = \max_{o=1, \dots, n_c} \{q_{1o}, \dots, q_{16o}\}. \quad (3)$$

Then the characteristics of the image block can be expressed as follows:

$$\alpha = \{\alpha_1, \dots, \alpha_{16}\}. \quad (4)$$

After obtaining the image block features, the K-SVD algorithm is used again to calculate the second layer of sparse coding based on image blocks.

3.3.3 Extraction and representation of image level features

Finally, SIFT is combined with other colors, depth, and other information to achieve sparse coding of feature expression. The article uses the pyramid pooling method to extract SIFT-based sparse coding, block feature-based sparse coding and image block features. These three image features can be combined to obtain the characteristics of the object.

For sparse coding ϕ s based on the SIFT feature, its corresponding subregion feature solution is as follows:

$$dg_{\delta}^{zt} = \max_{u=1, \dots, i} \{\delta_{u1}^{zt}, \dots, \delta_{up}^{zt}\}. \quad (5)$$

In formula (5), z is the number of layers, whose value is $\{0, 1, 2\}$, and i corresponds to the subregions of different layers. u represents the dimension of sparse coding, and p represents the characteristic number of t subregion in the z -layer. δ represents the sparse encoding μ_D corresponding to SIFT, block feature corresponding to block α , and sparse encoding α_s . All final image features are expressed as follows:

$$\Gamma = \{dg_{\mu_D}(h), dg_{\alpha}(v), dg_{\alpha}(f), dg_{\alpha}(mz), dg_{\alpha_s}(v), dg_{\alpha_s}(f), dg_{\alpha_s}(mz)\}. \quad (6)$$

Among them, $dg_{\mu_D}(h)$, $dg_{\alpha}(v)$, $dg_{\alpha}(f)$, and $dg_{\alpha}(mz)$ are features based on gray scale and SIFT sparse coding, color and image blocks, depth and image blocks, and object surface normal vector and image block, respectively. The same feature representation is used for dg_{α_s} , but it is obtained by combining with sparse coding of block features.

4 Experiment on image feature extraction algorithm of visual information

The main tasks of this article are given in three aspects. First, the feature matching performance of image feature extraction algorithm based on visual information is evaluated. Second, the correct matching rate of image feature extraction algorithm based on visual information (X algorithm) and conventional image feature extraction algorithm (Y algorithm) is compared from the aspects of scene rotation, illumination, blur, scale change, etc. Finally, the image feature extraction algorithm based on visual information is used to complete the recognition of the specified object in the scene and mark it.

The experiment uses the Logitech C910 camera to collect images. The software platform uses Microsoft Visual Studio2022 and the OpenCV (Open Source Computer Vision Library) computer vision library to complete the programming of relevant algorithms.

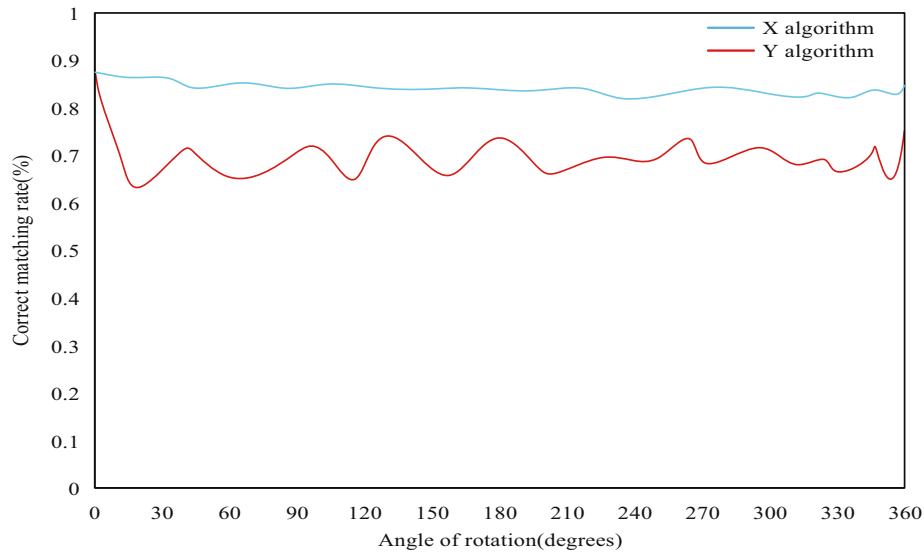
4.1 Characteristic matching performance evaluation

The experiment mainly verifies the feature matching performance of the two algorithms in the case of rotation and Gaussian noise in similar scenes. The plane two-dimensional image used in the experiment is divided into two data sets according to the different changes. For the image set used for the rotation change experiment, different original scene images and images within the 360° rotation range around the image center are used. For images with Gaussian noise, different original scene images and images with different degrees of Gaussian noise are used.

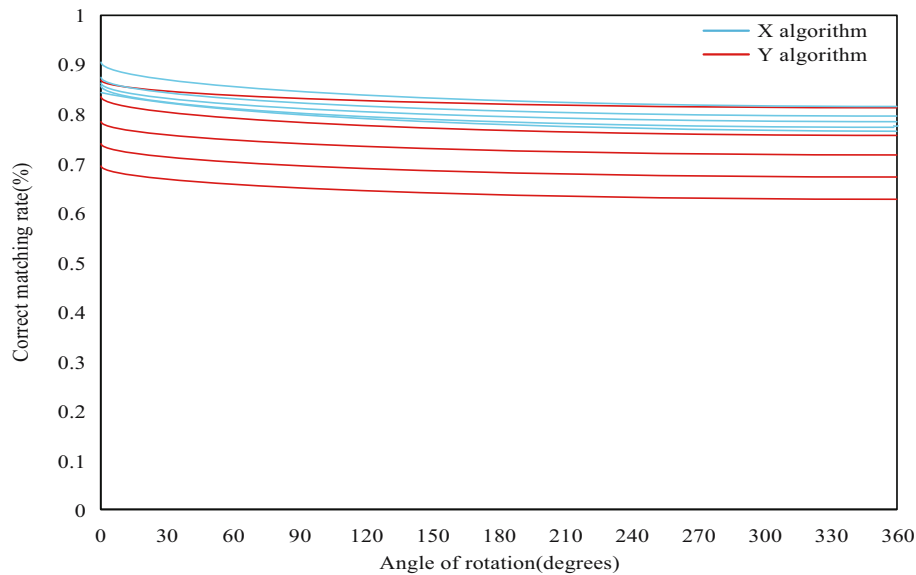
The measurement of all experiments is expressed by the correct feature matching ratio between the original image and each changed image. The VS2008 platform is used for programming, and the OpenCV platform is used for different algorithms. The most suitable matching pair is found through a simple pattern matching algorithm, and the corresponding plane homography matrix is estimated to estimate the number of correct matching points.

4.2 Comparative experiments of the effect of rotation invariance and Gaussian noise

On the one hand, this article used these two algorithms to test the correct matching rate between the rotated image and the original image. On the other hand, Gaussian kernel noise was added to the rotated image at the same time. The noise levels were 0, 10, 20, 30, and 50, respectively, and the change of matching performance was recorded, as shown in Figure 3.



(a)



(b)

Figure 3: Comparison of rotation invariance and Gaussian noise of the two algorithms. (a) Rotation invariance contrast, (b) Gaussian noise effect.

Figure 3(a) shows that with the increase of rotation angle, the correct matching rate of the X algorithm fluctuated, but the decline was not obvious, and its correct matching rate remained between 80 and 90%. The correct matching rate of Y algorithm fluctuated greatly and decreased significantly, and its correct matching rate has been around 70%. Figure 3(b) shows that with the increase in noise level, the matching performance of X algorithm and Y algorithm has decreased, but the decline arc of X algorithm is very small. The decline of Y algorithm was very obvious, and its matching performance decreased by about 5% for each level of noise increase. Figure 3 shows that the performance of X algorithm is greatly better than that of Y algorithm in terms of rotation invariance contrast and Gaussian noise. It shows that visual information plays an essential role in improving the correct matching rate and matching performance of the image feature extraction algorithm.

4.3 Effect of feature matching experiment

To compare the performance of X algorithm and Y algorithm in the real-time scene, such as feature matching rate and matching speed, the matching effect experiment of the image in the real-time scene with rotation, scale, blur and illumination changes has been carried out.

4.3.1 X algorithm feature matching findings

In this article, the X algorithm was utilized to test the matching effect of similar scenes on images with changes in illumination (A), plane rotation (B), scale (C), and Gaussian blur (D). The results are shown in Figure 4.

It can be learned from Figure 4(a) that the total number of feature matching, the number of wrong feature matching, and the matching rate of X algorithm were 280 pairs, 10 pairs, and 96.4%, respectively, under the change of illumination. When the plane rotation changed, they were 80 pairs, 7 pairs, and 91.3%, respectively. When the scale changed, they were 369 pairs, 30 pairs, and 91.9%, respectively. When Gaussian blur changed, they were 184 pairs, 6 pairs, and 93.4%, respectively. Figure 4(b) shows that the matching speed of X algorithm in the four scenarios was 0.45, 0.17, 0.58, and 0.24 s. Figure 4 shows that the matching rate of the X algorithm in the four scenarios has been maintained at more than 90%, and the time was also within 1 s. The matching effect is flawless.

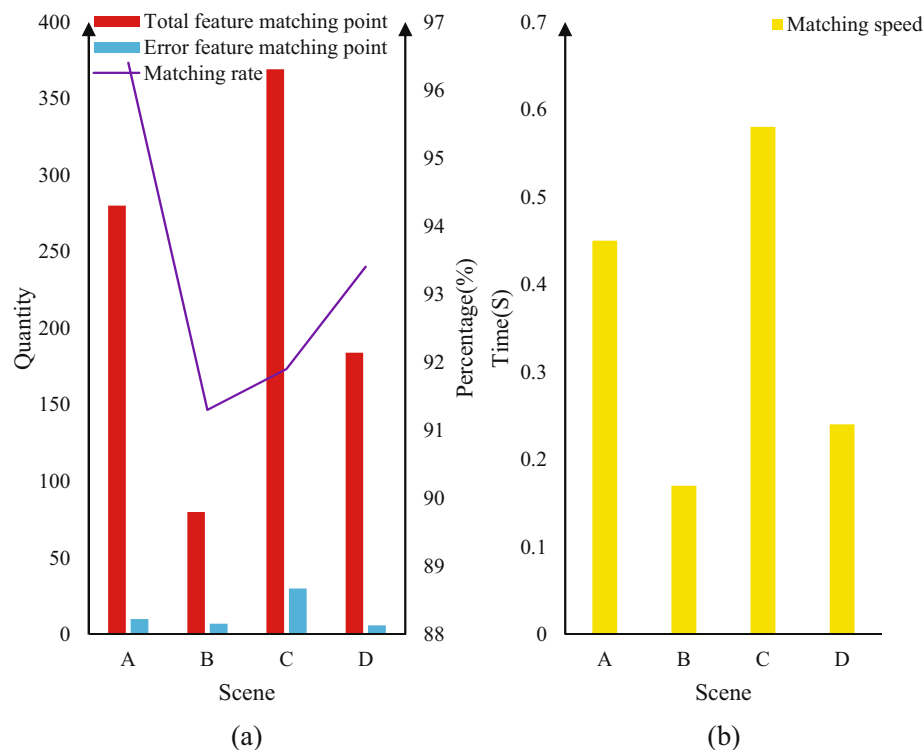


Figure 4: Feature matching rate and matching speed of X algorithm in different scenarios. (a) Feature matching rate and (b) matching speed.

4.3.2 Y algorithm feature matching result

The experiment remains the same, and only the algorithm is changed. The results are shown in Figure 5.

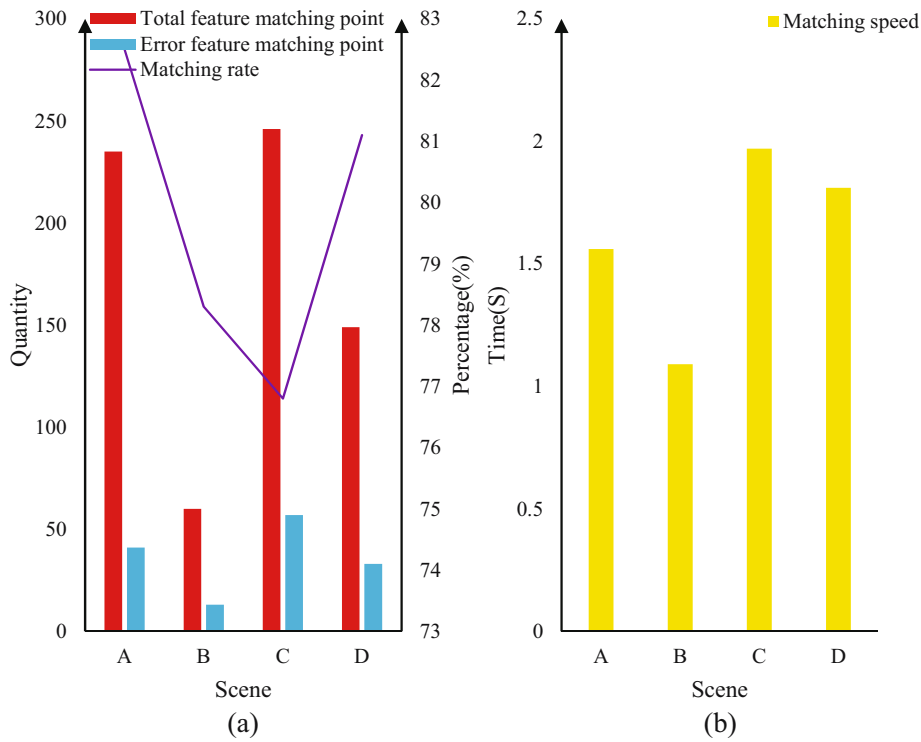


Figure 5: Feature matching rate and matching speed of Y algorithm in different scenarios. (a) Feature matching rate and (b) matching speed.

Figure 5(a) shows that the total number of feature matching, the number of wrong feature matching, and the matching rate of Y algorithm were 235 pairs, 41 pairs, and 82.5%, respectively, under the change of illumination. When the plane rotation changed, they were 60 pairs, 13 pairs, and 78.3%, respectively. When the scale changed, they were 246 pairs, 57 pairs, and 76.8%, respectively. When Gaussian blur changed, they were 149 pairs, 33 pairs, and 81.1%, respectively. Figure 5(b) shows that the matching speed of Y algorithm in the four scenarios was 1.56, 1.0, 1.97, and 1.81 s. Figure 5 shows that the matching rate of Y algorithm in the four scenarios has been maintained at about 80%. The time was also within 1–2 s, and the matching effect was far less than that of X algorithm.

To sum up, the X algorithm has fast extraction speed and strong stability against rotation changes and image noise. In the pursuit of matching accuracy, matching speed, and real-time, the algorithm has a strong application value.

4.4 Effect of target recognition experiment

To understand the experimental effect of target recognition of X algorithm and Y algorithm, four pictures were prepared and identified in this article. The feature extraction time, feature matching time, and correct matching rate were counted. Among them, *E* is an object with relatively poor surface texture. *F* shows an object with relatively rich surface texture. *G* shows an object with rich surface texture. *H* shows objects in similar scenes. The recognition results of different targets are shown in Figure 6.

Figure 6(a) shows that the feature extraction time, feature matching time, and correct matching rate of X algorithm were 0.41, 0.33 s, and 93.7%, respectively, for objects with relatively rich surface texture. For objects with relatively rich surface texture, they were 0.27, 0.54 s, and 99.1%, respectively. For objects with rich surface texture, they were 0.34, 0.59 s, and 97.4%, respectively. For objects in similar scenes, they were 0.48, 0.96 s, and 90.3%, respectively. Figure 6(b) shows that the feature extraction time, feature matching

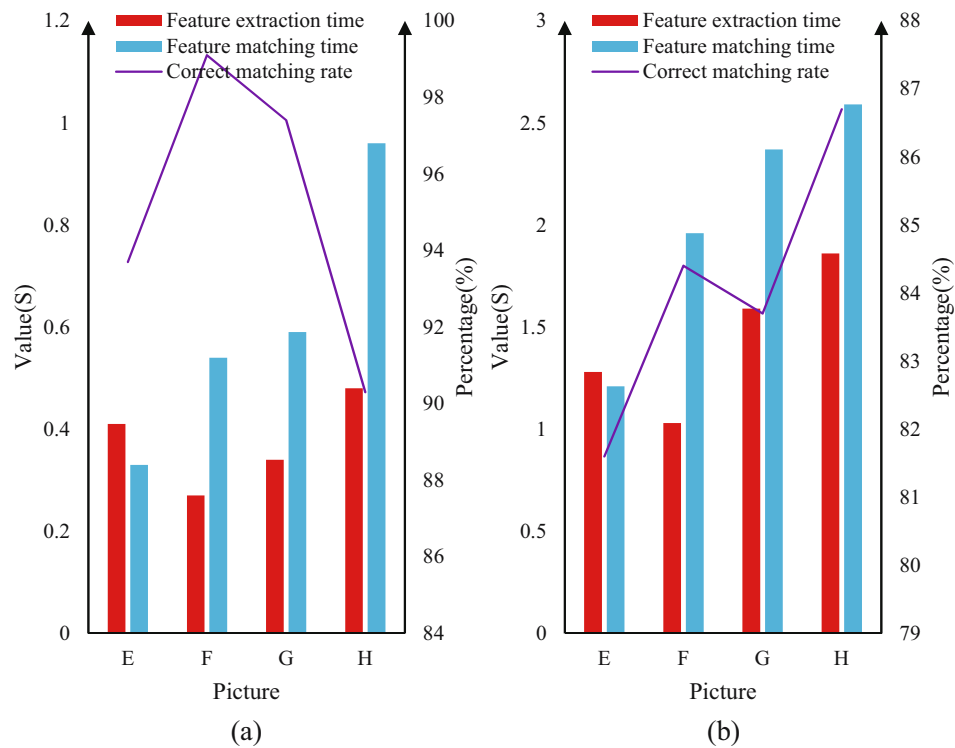


Figure 6: Recognition effect of two algorithms on different targets. (a) X algorithm and (b) Y algorithm.

time, and correct matching rate of Y algorithm were 1.28, 1.21 s, and 81.6%, respectively, for objects with relatively rich surface texture. For objects with relatively rich surface texture, they were 1.03, 1.96 s, and 84.4%, respectively. For objects with rich surface texture, they were 1.59, 2.37 s, and 83.7%, respectively. For objects in similar scenes, they were 1.86, 2.59 s, and 86.7%, respectively. Figure 6 shows that the feature extraction time of X algorithm for different targets was within 0.5 s, while the feature matching time was within 1 s. The correct matching rate was above 90%. The feature extraction time of Y algorithm for different targets was within 2 s, while the feature matching time was within 3 s. The correct matching rate was between 80 and 90%. It shows that the recognition effect of X algorithm is better than that of Y algorithm, which further indicates that the recognition effect of image feature extraction algorithm has been greatly improved by visual information.

5 Conclusions

This article proposed an image feature extraction algorithm based on visual information. The algorithm combines the ability to detect and recognize the region of the image, to obtain better image features. First, the two algorithms were compared with rotation invariance, and the influence of noise change was analyzed. Then, illumination, plane rotation, scale, and Gaussian blur changes were tested. Finally, the relationship between visual information and image feature extraction algorithm was verified by feature extraction time, feature matching time, and correct matching rate. Through the analysis of experimental results, it can be seen that this method can effectively recognize and match the features of the image, and the matching rate has been kept above 90%. Therefore, the algorithm proposed in this article is applicable to a variety of situations, which has high efficiency, accuracy, and good versatility. Because of the lack of comprehensive knowledge and ability, the article has studied the effect of target recognition experiments, but the experimental pictures are relatively small. Although it has certain representativeness, accidental factors may still occur, which

makes the conclusion of the article have certain limitations. Therefore, it is hoped that increasingly scholars can actively participate in this research and expand the research objects and experimental pictures to improve the experimental conclusions and accelerate the application of visual information in the field of image feature extraction.

Funding informations: This article is the research result of the science and technology project of Jiangxi Provincial Department of Education (Project Name: Research on Deep Learning-Based Video Image Target Detection Tracking Modeling Analysis, Project No.: GJJ212325).

Author contributions: Zhaosheng Xu: Work concept or design. Xiuhong Xu: The data collection. Zhongming Liao: Draft paper. Suzana Ahmad: Make important revisions to the paper. Zhongqi Xiang: Approve final paper for publication.

Conflict of interest: The authors declare that there is no conflict of interest with any financial organizations regarding the material reported in this manuscript.

Data availability statement: Data are available upon reasonable request.

References

- [1] Liu Y, Yang C, Sun Q. Thresholds based image extraction schemes in big data environment in intelligent traffic management. *IEEE Trans Intell Transp Syst.* 2020;22(7):3952–60.
- [2] Ganji A, Minet L, Weichenthal S, Hatzopoulou M. Predicting traffic-related air pollution using feature extraction from built environment images. *Environ Sci Technol.* 2020;54(17):10688–99.
- [3] Lenjani A, Yeum CM, Dyke S, Billionis I. Automated building image extraction from 360 panoramas for postdisaster evaluation. *Comput-Aided Civ Infrastruct Eng.* 2020;35(3):241–57.
- [4] Kosari A, Sharifi A, Ahmadi A, Khoshsima M. Remote sensing satellite's attitude control system: Rapid performance sizing for passive scan imaging mode. *Aircr Eng Aerosp Technol.* 2020;92(7):1073–83.
- [5] Farmonov N, Amankulova K, Szatmari J, Sharifi A, Abbasi-Moghadam D, Mirhoseini Nejad SM, et al. Crop type classification by DESIS hyperspectral imagery and machine learning algorithms. *IEEE J Sel Top Appl Earth Obs Remote Sens.* 2023;16:1576–88.
- [6] Ghaderizadeh S, Abbasi-Moghadam D, Sharifi A, Tariq A, Qin S. Multiscale dual-branch residual spectral-spatial network with attention for hyperspectral image classification. *IEEE J Sel Top Appl Earth Obs Remote Sens.* 2022;15:5455–67.
- [7] Gao L, Li X, Song J, Shen HT. Hierarchical LSTMs with adaptive attention for visual captioning. *IEEE Trans Pattern Anal Mach Intell.* 2019;42(5):1112–31.
- [8] Yoon H, Kim B-H, Mukhriddin M, Cho J. Salient region extraction based on global contrast enhancement and saliency cut for image information recognition of the visually impaired. *KSII Trans Internet Inf Syst (TIIS).* 2018;12(5):2287–312.
- [9] Pushpita K. Improvement and implementation of machining and positioning method of intelligent construction machinery components relying on machine vision. *Kinetic Mech Eng.* 2021;2(2):45–53.
- [10] Zeng X, Wang Z, Hu Y. Enabling efficient deep convolutional neural network-based sensor fusion for autonomous driving. 2022. arXiv preprint arXiv:2022.11231.
- [11] Khaleefah SH, Mostafa SA, Mustapha A, Nasrudin MF. Review of local binary pattern operators in image feature extraction. *Indones J Electr Eng Comput Sci.* 2020;19(1):23–31.
- [12] Wang S, Ding C, Zhang N, Liu X, Zhou A, Cao J, et al. A cloud-guided feature extraction approach for image retrieval in mobile edge computing. *IEEE Trans Mob Comput.* 2019;20(2):292–305.
- [13] Yang B, Liu M, Wang Y, Zhang K, Meijering E. Structure-guided segmentation for 3D neuron reconstruction. *IEEE Trans Med Imaging.* 2022;41(4):903–14.
- [14] Du H, Wang J, Liu M, Wang Y, Meijering E. SwinPA-Net: Swin transformer based multiscale feature pyramid aggregation network for medical image segmentation. *IEEE Trans Neural Netw Learn Syst.* 2022;1–12. doi: 10.1109/TNNLS.2022.3204090.
- [15] Balasamy K, Shamia D. Feature extraction-based medical image watermarking using fuzzy-based median filter. *IETE J Res.* 2023;69(1):83–91.
- [16] Zhang Y, Poon T-C, Tsang PWM, Wang R, Wang L. Review on feature extraction for 3-D incoherent image processing using optical scanning holography. *IEEE Trans Ind Inform.* 2019;15(11):6146–54.
- [17] Jiang J, Ma J, Chen C, Wang Z, Cai Z, Wang L. SuperPCA: A superpixelwise PCA approach for unsupervised feature extraction of hyperspectral imagery. *IEEE Trans Geosci Remote Sens.* 2018;56(8):4581–93.

- [18] Bashar A. Survey on evolving deep learning neural network architectures. *J Artif Intell.* 2019;1(2):73–82.
- [19] Grange C, Barki H. The nature and role of user beliefs regarding a website's design quality. *J Organ End User Comput.* 2020;32(1):75–96.
- [20] Venkatesh B, Anuradha J. A review of feature selection and its methods. *Cybern Inf Technol.* 2019;19(1):3–26.
- [21] Santoso MH, Larasati DA, Muhathir M. Wayang image classification using MLP method and GLCM feature extraction. *J Computer Sci Inf Technol Telecommun Eng.* 2020;1(2):111–9.
- [22] Varuna Shree N, Kumar TNR. Identification and classification of brain tumor MRI images with feature extraction using DWT and probabilistic neural network. *Brain Inform.* 2018;5(1):23–30.
- [23] Rasti B, Hong D, Hang R, Ghamisi P, Kang X, Chanussot J, et al. Feature extraction for hyperspectral imagery: The evolution from shallow to deep: Overview and toolbox. *IEEE Geosci Remote Sens Mag.* 2020;8(4):60–88.