



# Research on Image Recognition Algorithm Based on Depth Level Feature

Chan Zhang<sup>(✉)</sup> and Jia Luo

School of Information Technology, Guangdong Industry Polytechnic,  
Guangzhou, Guangdong, China  
zhangchan332@163.com

**Abstract.** In order to solve the problem that the traditional image recognition algorithm can not guarantee a relatively stable recognition accuracy and poor robustness under a variety of interferences, the image recognition algorithm based on depth level features is studied. After preprocessing, such as filtering and enhancing, the image to be recognized is segmented. The segmented image is input into convolution neural network, and the feature of depth level is extracted from the neural network. The feature points of the extracted deep level features are matched to realize image recognition, and the algorithm of image recognition based on the deep level features is designed. Compared with the traditional image recognition algorithm, the designed image recognition algorithm can ensure a more stable recognition accuracy and has better robustness.

**Keywords:** Depth level features · Image recognition · Algorithm research

## 1 Introduction

Image recognition technology plays an extremely important role in various fields. Good recognition technology is the key. How to improve the recognition rate and speed is of great significance, which is directly related to the practicability and security of image recognition. Image data, as a common information source, is characterized by huge information content, complex structure and redundancy. Therefore, compared with other types of data, digital image processing is more difficult. Image features are the centralized and simplified expression of image information. Due to different shooting angles, shooting environments or photographers, the same object will present different shapes on different images, however, the features used to express the object should be as stable as possible [1]. Therefore, image features are designed to express the invariable and essential information of the object in the image. No matter what shape the object presents in the image, the features of the object in the image can be used as invariable marks, which can be effectively identified by the computer.

Image recognition is one of the important research topics in the field of computer vision and pattern recognition. It has a wide range of application prospects in practice, and it is also an important tool to realize the intelligent society in the future [2]. Deep level feature is an important theory in deep learning. Based on convolution neural network, the objects of input network are expressed in deep level feature, which is

helpful for subsequent processing. Based on the above analysis, this paper designs an image recognition algorithm based on depth level.

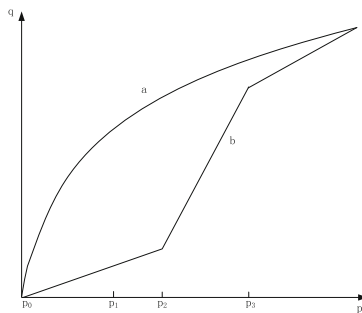
## 2 Research on Image Recognition Algorithm Based on Depth Level Feature

### 2.1 Image Preprocessing

In practical application, the image to be recognized will contain interference factors that affect the recognition accuracy due to the influence of acquisition equipment, personnel, environment and other factors. Therefore, image preprocessing is needed before image recognition.

Firstly, the image is filtered to remove the noise in the image. When the number of pixels in a region is less than a fixed threshold, it is judged as noise, and the region is filled as background color to be removed. The size of each connected region is calculated by identifying the connected region of the whole picture. That is, all pixels belonging to the same connected domain are marked with the same region. Taking the 4-connected identification method as an example, first apply for a piece of identification space, scan the image from top to bottom, from left to right, for any pixel with a value of 0 in the image, the pixel above is  $t$ , and the pixel on the left is  $r$ . If  $r = t = 0$ , a new identifier will be given to  $p$ ; if  $r$  and  $t$  have one of 1, a marker of 1 will be given to  $p$ ; if  $r = t = 1$ , and there is the same marker, the marker will be given to  $p$ ; if  $r = t = 1$ , but there are different markers, a smaller value will be given to  $P$ , indicating that the two markers are equivalent. Merge all the equivalent tags, scan again until the tags no longer change, and complete the image filtering. After image filtering, the image details are enhanced to facilitate the subsequent extraction of image features [3].

In this paper, image enhancement is realized by gray level transformation or histogram equalization. The gray level transformation does not depend on the position of the pixel in the image. Using the transformation  $T$ , the original brightness  $P$  in the range  $[p_0, p_k]$  is transformed into a new range  $[q_0, q_k]$ . The gray level transformation for enhancing image contrast is shown in the following figure (Fig. 1).



**Fig. 1.** Gray level transformation

The curves a and b in the image are two different mapping of gray level changes. The  $[p_0, p_1]$  and  $[p_2, p_3]$  areas in a and b are the areas of image contrast enhancement. The pixel value difference between these two areas is enlarged. For 256 color gray-scale image, only 256 bytes of storage space is needed to save a look-up table. The content of the table is the transformed gray-scale value. The original gray-scale value is used as the look-up index of the table, and the pixel value of each point in the image is searched and replaced, that is, the gray-scale transformation of the image is completed. In order to make the object to be recognized have the same gray value in the whole brightness range, histogram equalization is used.

For an image with  $G$  gray levels and size of  $M \times N$ , create an array  $H$  with length of  $G$ , and initialize its elements to 0. Establish image histogram: scan each pixel, and increase the corresponding  $H$  value. When the pixel  $p$  has brightness  $g_p$ , there are:

$$H[g_p] = H[g_p] + 1 \tag{1}$$

Thus, a cumulative histogram  $H_c$  is formed:

$$\begin{cases} H_c[0] = H[0] \\ H_c[p] = H_c[p - 1] + H[p] \quad p = 1, 2, \dots, G - 1 \end{cases} \tag{2}$$

Set up  $T[p]$ , Make  $T[p]$  meet the following formula:

$$T[p] = \text{round} \frac{G - 1}{NM} H_c[p] \tag{3}$$

Rescan the image, for the pixel point with gray value of  $g_p$ , make the new gray value of  $T[g_p]$ , and get the image after histogram equalization [4]. After image pre-processing, the image is segmented.

## 2.2 Image Segmentation

Because a digital image contains too much information, it will bring great interference to the intelligent recognition of the object in the image. In order to separate some areas of useless information from the areas of interest, we need to use the means of image segmentation. In this paper, the edge based segmentation method is used to segment the preprocessed image.

When detecting the edge of the target, firstly roughly detect its contour points, then connect the detected contour points according to certain principles, test and link the missing contour points, and remove the wrong boundary points at the same time [5]. Use the  $3 \times 3$  operator template as shown in the figure below to segment the image after detecting the edge (Fig. 2).

$Z_1$	$Z_2$	$Z_3$
$Z_4$	$Z_5$	$Z_6$
$Z_7$	$Z_8$	$Z_9$

**Fig. 2.**  $3 \times 3$  operator template

Define the gradients in horizontal  $g_x$  and vertical  $g_y$  directions:

$$\begin{cases} g_x = \frac{\partial f}{\partial x} = (Z_7 + Z_8 + Z_9) - (Z_1 + Z_2 + Z_3) \\ g_y = \frac{\partial f}{\partial y} = (Z_3 + Z_6 + Z_9) - (Z_1 + Z_4 + Z_7) \end{cases} \quad (4)$$

Define gradients in the direction of  $g'_x$  and  $g'_y$  diagonals:

$$\begin{cases} g'_x = (Z_2 + Z_3 + Z_6) - (Z_4 + Z_7 + Z_8) \\ g'_y = (Z_6 + Z_8 + Z_9) - (Z_1 + Z_2 + Z_4) \end{cases} \quad (5)$$

According to this definition, the generated operator template is as shown in the following figure (Fig. 3):

-1	-1	-1	-1	0	1
0	0	0	-1	0	1
1	1	1	-1	0	1

-1	1	1	-1	-1	0
-1	0	1	-1	0	1
-1	-1	0	0	1	1

**Fig. 3.** Prewitt operator template

After calculating the gradient, the image is segmented according to the gradient. After image segmentation, deep features of image are extracted.

### 2.3 Deep Feature Extraction of Image

The image deep feature extraction in this paper is carried out in the convolution neural network. The segmented image is input into the input layer of the convolution neural network. After processing in the neural network, the feature map is generated, and then the deep feature of the image is extracted from the feature map [6].

Each layer of CNN structure contains several feature graphs, each of which is expressed from different scales and layers. Low level feature images can be understood directly through visualization. They capture the edge of the object and the sharp changes of the pixel information in the image. The detail texture structure of the original image is still preserved in the feature image. The visualization results of the feature map on the high level are not easy to understand, and the high level features are abstracted and iterated to express the semantic information of the image.

In CNN hierarchical structure, each layer of feature map expresses image information from different aspects. Making full use of the information of feature map to describe image can make the constructed image features have stronger expression ability. Because the feature images at different levels have different sizes, the higher the level is, the smaller the size of the feature image is. Therefore, in order to facilitate the extraction of hierarchical features, it is necessary to adjust the feature images at different levels to the original image size. In this paper, the bilinear interpolation method is used to sample the feature map at different levels, making it have the same size as the original image, stacking all the feature maps to build a three-dimensional matrix  $F \in R^{N \times H \times W}$ , where H is the height of the image, W is the width of the image, and N is the number of feature maps. F can be expressed as:

$$F = [up(F_1), up(F_2), \dots, up(F_L)] \quad (6)$$

Among them,  $up$  is the upper sampling operation, and the number of characteristic graphs at the level of  $up(F_i) \in R^{N_i \times H \times W}$  and  $N_i$  [7]. For any region Q on the image, its descriptor can be represented by N-dimensional vector as  $D_Q \in R_N$ , and the i-dimensional  $d_i$  of the descriptor represents the result of describing the corresponding region pixels on the i-level feature map.

DHF feature extraction process is as follows: the trained CNN model is used in the original image, and the convolution output layer is selected to construct the image hierarchy. Re adjust the size of the feature map to the original image. For the feature extraction unit on the original image, determine its corresponding area on the feature map, and count the pixel information contained in the area on the feature map. The low-level feature image still retains rich texture information, and the method of information entropy is used to count the pixels in the feature area. The maximum and minimum values of pixels in the statistical region are divided into several intervals. Count the number of pixels in each interval  $n_i, i = 1, 2, \dots, bins$ ,  $bins$  is the number of divided intervals, and calculate the probability of pixels falling into each interval  $p_i$  is the total number of regional pixels.

$$p_i = \frac{n_i}{total} \quad (7)$$

where *total* is the total number of area pixels. Calculate the information entropy of area pixels:

$$H = - \sum_{i=1}^{bins} p_i \cdot \log_2 p_i \quad (8)$$

Information entropy is a kind of quantity used to measure the degree of information confusion. Specifically, in this paper, the location distribution of significant changes in the region with lower entropy rate is relatively centralized, and the location of significant changes with higher entropy rate is widely distributed in the whole characteristic region [8]. The feature graph at high level has strong abstract semantic information. In this paper, region average method is used to reduce the influence of noise in the region. Finally, the DHF feature descriptor of feature extraction unit is constructed by integrating the high and low level feature map information. Complete the deep feature extraction of the image to be recognized, match the feature points, and realize the image recognition.

## 2.4 Image Recognition

After the feature points in the image are extracted, they need to be matched and correlated. Image feature matching and association is to find the feature points extracted from one image and find the matching feature points in another image. This step is completed by calculating the feature descriptors of the extracted image feature points, and then calculating the similarity between the feature descriptors. Image feature descriptor is a kind of vector which can describe the local structure features of image quantitatively. It can fully represent the local shape and texture structure features near the feature points. After extracting feature points, the matching and association of feature points are based on the similarity between feature descriptors. The cosine distance of the included angle shown in the following formula is used to calculate the similarity.

$$d = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}} \quad (9)$$

In the above formula,  $x_i$  and  $y_i$  are the feature points to be matched. Generally speaking, the smaller the distance between two vectors is, the higher the similarity is. However, in the process of feature matching and association, it is simply considered that the feature points with the smallest distance between feature descriptors are matched points, which may cause some mismatches. If there is a feature point A in the reference image, its feature descriptor is  $Desc_a$ , there may be two mismatches: one is

that the real feature points corresponding to A in the image to be matched are not extracted, at this time, only other wrong feature points can be matched; the other is that in the image to be matched, there may be multiple feature points with the minimum distance between the feature descriptor and  $Desc_a$ , at this time, there is a one to many matching problem [9].

In order to reduce the problem of mismatching, we use the ratio of the nearest distance to the next nearest distance to associate feature point pairs [10]. Suppose that there are two feature points  $B_1$  and  $B_2$  in the image to be matched, their feature descriptors are  $Desc_{b1}$  and  $Desc_{b2}$  respectively. Compared with the feature descriptors of all other feature points, the distance between  $Desc_a$  and  $Desc_{b1}$  is the smallest, while the distance between  $Desc_a$  and  $Desc_{b2}$  is the second smallest. At this time, it is not directly considered that point A and point  $B_1$  are matched, but the distance between  $Desc_a$  and  $Desc_{b1}$  divided by the distance between  $Desc_a$  and 12 is calculated, when the ratio is less than a certain threshold, point a and point  $B_1$  are considered to match each other.

Through feature matching and association, the next step of image matching is to solve a certain spatial transformation model according to the spatial coordinates of these matching feature points. For two images that need to be matched, one of them can be regarded as the result of some spatial transformation for the other, and the two can be modeled according to some spatial transformation model. In image matching, affine transformation is generally used to model the spatial transformation relationship between the reference image and the image to be matched. The mathematical model is shown in the following formula.

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & t_x \\ a_3 & a_4 & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{10}$$

Among them,  $(x, y)$  and  $(x', y')$  represent the coordinates of the midpoint of the two images respectively,  $(t_x, t_y)$  represents the amount of translation, and  $a_j, j = 1, 2, 3, 4$  represents the parameters of image rotation, scaling and other transformations. After the parameters  $t_x, t_y$  and  $a_j$  are calculated by certain methods, the transformation relationship of spatial coordinates between the two images can be obtained.

Assume that there are N pairs of matching feature points found in feature matching and association steps. For affine transformation model, it needs at least 4 pairs of matching feature points to solve its parameters. For the remaining  $n - 4$  pairs of matching points, assume that the coordinate of reference image is  $A_i(i = 1, 2, \dots, N - 4)$ , and the coordinate of matching point corresponding to  $B_i(i = 1, 2, \dots, N - 4)$  is, take  $A_i$  into the result obtained in the previous step, and calculate the transformed coordinate, assuming that it is 4  $A'_i(i = 1, 2, \dots, N - 4)$ . Calculate the distance  $D_i$  between  $A'_i$  and  $B_i$ . If the distance is less than a preset threshold T, the correct matching point of the point is considered as the internal point. If it is not satisfied, the point is considered as the wrong matching point, which is called the external point. Make a note of all the inner points, find 4 new pairs in the matching point pairs, perform the above steps again, and record the inner points every time. After

repeating the above process many times, compare the number of interior points obtained each time. The parameter obtained in the most number is the required solution parameter. Using the obtained parameters, the spatial matching of feature points is completed and the matching results are output. At this point, the research of image recognition algorithm based on depth level feature is completed.

### 3 Algorithm Test Experiment

In order to verify the performance of the algorithm, we choose the image data as the experimental data in YouTube labeled video dataset (<https://research.google.com/youtube8m/>), exclude the unavailable data, take the remaining data as the experimental sample set, and carry out the comparative experiment with three kinds of commonly used images A cognitive algorithm is designed. The experimental hypothesis was verified by comparison.

#### 3.1 Experimental Content

In this section, we compared the robustness of the two image recognition algorithms to evaluate the advantages and disadvantages of the two algorithms. The experimental method was comparative experiment. The experiment verification process is completed under the control of other variables except the experimental variation.

The object of the experiment is 300 images randomly selected from the library, and each image is then selected 5 images with higher similarity as interference items. The experiment was divided into three groups, the first group added noise to the experimental image; the second group did blur processing, namely the reverse processing of sharpening processing; the third group reduced the brightness of the image. The image recognition algorithm designed in this paper will be compared with the recognition algorithm based on neural network, the recognition algorithm based on information fusion matching and the recognition algorithm based on template matching. Taking the above three algorithms as the reference group of the comparative experiment, the algorithm proposed in this paper is the experimental group, recording the relevant experimental data, and completing the experiment.

Add quantization interference to the experimental image according to the quantization table of external interference degree of the experimental image shown below. 300 images were randomly divided into three groups. After adding different interferences, the experiment was carried out (Table 1).

Using the image recognition algorithm based on the depth level feature designed in this paper and its comparative recognition algorithm to recognize the image with interference, comparing the image recognition accuracy of different algorithms, and judging the robustness of four image recognition algorithms under the external interference.

**Table 1.** Quantization table of external interference degree of experimental image

Degree of interference	Add noise /pixels	Fuzzy processing /%	Reduce the luminance $\text{cd/m}^3$
1	10	20	0.05
2	20	40	0.10
3	30	60	0.15
4	40	80	0.20

### 3.2 Experimental Result

Under the noise interference, the recognition accuracy of the proposed recognition algorithm and the recognition algorithm based on neural network for the image of the experimental object is shown in the table below (Table 2).

**Table 2.** Recognition accuracy under noise interference/%.

Degree of interference	Experimental group algorithm	Reference group algorithm
1	96.42	88.97
2	95.84	84.43
3	95.76	81.69
4	95.43	79.75

It can be seen from the above table that the recognition accuracy of the experimental group algorithm is higher than that of the reference group algorithm after adding different levels of noise to the image. With the increase of the interference degree, the recognition accuracy of the experimental group algorithm has no significant change, and the recognition accuracy of the reference group algorithm has a downward trend, and the decline range is large. It shows that the recognition algorithm proposed in this paper can guarantee higher recognition accuracy when adding noise interference to the image.

After the experimental image is fuzzy processed, the recognition accuracy of the proposed recognition algorithm and the recognition algorithm based on information fusion matching for the experimental object image is shown in the table below (Table 3).

**Table 3.** Recognition accuracy after fuzzy processing/%.

Degree of interference	Experimental group algorithm	Reference group algorithm
1	91.16	74.28
2	93.27	72.75
3	91.56	71.97
4	92.48	67.31

It can be seen from the above table that the recognition accuracy of the experimental group algorithm is higher than that of the reference group after image fuzzy processing, and does not fluctuate with the increase of the fuzzy degree, but the recognition accuracy of the reference group algorithm will decrease with the increase of the interference degree.

After reducing the gray level of the image, the recognition accuracy of the proposed recognition algorithm and the recognition algorithm based on information fusion matching for the experimental object image is shown in the table below (Table 4).

**Table 4.** Recognition accuracy after gray level reduction/%.

Degree of interference	Experimental group algorithm	Reference group algorithm
1	85.47	65.64
2	86.05	64.59
3	87.16	61.53
4	86.28	60.44

After analyzing the above table, the accuracy of the two groups of image recognition algorithms is affected after the gray level of the image is reduced. However, with the gray level getting lower and lower, the accuracy of the reference group algorithm is getting lower and lower, but the recognition rate of the experimental group algorithm does not fluctuate greatly. In conclusion, the image recognition algorithm based on the depth level features designed in this paper can guarantee high recognition accuracy and good robustness under various levels of image interference.

## 4 Conclusions

In this paper, an image recognition algorithm based on depth level features is designed. Firstly, the image is preprocessed. Secondly, the segmented image is input into the input layer of convolution neural network, and the feature mapping is generated in the neural network. Finally, the deep features of the image are extracted from the feature mapping. Finally, the effectiveness of the algorithm is proved by the comparative experiment of three traditional image recognition algorithms. The algorithm designed in this paper can ensure high recognition accuracy and good stability under different levels of interference.

## References

1. Yang, J., Zhou, T., Guo, L., et al.: Lung tumor image recognition algorithm based on cuckoo search and deep belief network. *J. Comput. Appl.* **38**(11), 3225–3230 (2018)
2. Gao, Y., Wang, B., Guo, Z., et al.: Clothing image recognition and classification based on HSR-FCN. *Comput. Eng. Appl.* **55**(16), 144–149 (2019)

3. Liu, Y., Yang, J., Wang, C.: Optimization of behavior recognition algorithm based on low-level feature modeling. *Sci. Technol. Eng.* **18**(21), 69–75 (2018)
4. Zhang, Q., Guan, S., Xie, H., et al.: Image adaptive target recognition algorithm based on deep feature learning. *J. Taiyuan Univ. Technol.* **49**(04), 592–598 (2018)
5. Huang, X., Ling, Z., Li, X.: Discriminative deep feature learning method by fusing linear discriminant analysis for image recognition. *J. Image Graph.* **23**(04), 510–518 (2018)
6. Yi, L., Ya, E.: Based on Gobor feature and deep belief network face recognition methods. *Comput. Simul.* **34**(11), 417–421 (2017)
7. Sang, S., Feng, H., Wang, L.: Image true-false decision algorithm based on improved SURF coupled hierarchical clustering. *Comput. Eng. Des.* **38**(06), 1602–1607 (2017)
8. Li, Y., Yang, Q., Lei, H., et al.: Improved Criminisi algorithm based on DIBR. *Comput. Eng. Des.* **38**(05), 1287–1291 (2017)
9. Li, Q., You, X., Li, K., et al.: Deep hierarchical feature extraction algorithm. *Pattern Recognit. Artif. Intell.* **30**(02), 127–136 (2017)
10. Ma, X., Sang, Q.: Handwritten signature verification algorithm based on LBP and deep learning. *Chin. J. Quantum Electron.* **34**(01), 23–31 (2017)