Research on Feature Extraction Method of Geometric Image Recognition

NaChen¹*, Gang Zheng¹, Chunqiong Wu²

¹Lanzhou Institute of Technology, Lanzhou, China
²Yango University, FuZhou, FuJian,730050, China

Corresponding author’s e-mail:xiaohuanong312@163.com

Abstract. The existing image recognition feature extraction technology cannot effectively transmit the image information if the extracted features are not complete, sufficient and inaccurate. It will have an impact on image processing, such as reducing the accuracy and efficiency of subsequent image recognition and image tracking. At the same time, the existing technology for image processing destroys the topological structure between image pixels, and the process is complicated in the calculation of high-dimensional data space. Therefore, this paper proposes a method and steps of feature extraction for collective image recognition, and makes relevant supplementary explanation and analysis.

1. Introduction

Image recognition is a technique that uses a computer to process, analyze, and understand images to identify targets and objects in various modes. In general industrial use, an industrial camera is used to take pictures, and then the software is used to further identify and process according to the grayscale difference of the image. The image recognition software has foreign representatives such as Cognex, etc., and domestic representatives have graphic intelligence. However, the existing image recognition feature extraction technology cannot effectively transmit the image information if the extracted features are not complete, sufficient and inaccurate. It will have an impact on image processing, such as reducing the accuracy and efficiency of subsequent image recognition and image tracking. At the same time, the existing technology for image processing destroys the topological structure between image pixels, and the process is complicated in the calculation of high-dimensional data space.

2. Extraction Method

In view of the problems existing in the existing technology, the proposed image recognition feature extraction method includes the following steps:

Step1: conduct two-dimensional discriminant feature learning on the original training samples included in the original training set;

Image feature learning modelling by compact local intra-class divergence and separation of local inter-class divergence can effectively maintain the topological structure and intrinsic correlation between image pixels. Based on the 1 norm metric, the robustness of image description is improved. Then, by optimizing a feature decomposition problem, a projection matrix which can be used for 2D robust feature extraction of image samples is obtained:
M-dimensional projection matrix is \( V = [V_1, V_2, \ldots, V_N] \), among them \( V_j = v_j, j = 1, 2, \ldots, N \), find the proper projection of \( M \times L \) 's base matrix \( W = [W_1, W_2, \ldots, W_N] \) and the coefficient matrix of \( L \times N \) is \( H = [H_1, H_2, \ldots, H_N] \), make: \( V \approx WH \), it is expressed as the form of vector scalar product:

\[
V_j \approx \sum_{i=1}^{N} W_i H_{ij}
\]

For a given set of M-dimensional data vectors \( V_{M \times N} \), Where \( N \) is the number of data samples in the set, decomposed into two matrices \( W_{M \times L} \) and \( H_{L \times N} \), the product of these two matrices is similar to \( V_{M \times N} \). Choose \( L \) less than \( M \) or \( N \), make \( W \) and \( N \) less than the original matrix \( V \), and make the decomposition matrix reduce the dimension of the original data matrix \( V \).

In this study, in the process of image recognition, the original observation image to be fused is actually the objective real world that introduced noise formation in the imaging process. Therefore, effectively reducing or eliminating these noises can improve the sharpness of the image. In the projection matrix decomposition algorithm, assuming \( V = WH + \varepsilon \), \( \varepsilon \) represents noise. At this time, noise tends to converge in the iterative algorithm, which is just in line with the process of image fusion. Therefore, in connection with the image fusion process, if it is assumed that the observed image is \( V \), the real image is \( W \), and the noise is \( \varepsilon \), then \( V \) can be understood as the sum of \( W \) and \( \varepsilon \), so that NMF can be effectively applied to image fusion.

The NMF algorithm can obtain a partial-based approximate representation \( WH \) for the original data matrix \( V \) by an iterative operation method. Among them, the number of columns of \( W \) that is, the number \( r \) of feature bases is a quantity to be quantified, which will directly determine the dimension of the feature subspace. For a specific data set, the dimension of the feature space hidden inside the data set is determined, that is, when the selected \( r \) is consistent with the dimension of the feature space of the actual data set, the obtained feature space and the basis of the feature space are the most significant. When \( r = 1 \), an iterative algorithm will result in a unique one containing all the features of the source data.

It can be seen from the above that NMF and image fusion can be well combined and applied. Suppose there are \( k \) observation images \( f_1, f_2, \ldots, f_k \) of multi-sensor size \( m \times n \), and the elements of each observation image are stored row by row into one column vector, so that a matrix \( V \) of \( mn \times k \) can be obtained, and \( k \) is included in \( V \). Column vector \( v_1, v_2, \ldots, v_k \), each column vector represents information of one image in \( k \) observation images, as shown in following equation. For the decomposition matrix \( V \) projection matrix decomposition, taking \( r=1 \) in the decomposition, a unique characteristic base \( W \) can be obtained. Obviously, \( W \) at this time contains the complete features of the \( k \) images participating in the fusion, and the feature base \( W \) is restored to the pixel level of the source image to obtain an image that is better than the source image.

\[
\{v_i = [f_{1i}, f_{2i}, \ldots, f_{ni}, f_{12i}, f_{22i}, \ldots, f_{ni}, f_{13i}, f_{23i}, \ldots, f_{ni}, f_{1m}, f_{2m}, \ldots, f_{ni}, f_{1n}, f_{2n}, \ldots, f_{ni}, f_{1mp}]^T
\}
\]

\[
[= [v_1 = f_1, v_2 = f_2, \ldots, v_k = f_k]
\]

Wherein the original training sample is a sample having a category label corresponding to a category of the original training sample;

**Step2:** acquiring a new training sample set including two-dimensional robust features of each original training sample, and constructing a classifier by using the new training sample set; A two-dimensional robust feature of each of the original training samples is obtained by projecting each of the original training samples using the projection matrix;

**Step3:** using the classifier to classify the sample to be tested, and obtaining a classification result corresponding to the category of the sample to be tested, wherein the sample to be tested is a sample whose category is unknown;

**Step4:** obtaining a target image from the sample classification result to be tested;

**Step5:** dividing the target image into at least two regions according to a direction in which the character in the target image is upright;

**Step6:** extract color information of each region in a plurality of color spaces; combine color information of the plurality of color spaces extracted from the at least two regions to obtain a
combined feature vector representing the target image. Specifically:

- extracting color features and adaptive LBP operator features.
- construct a multi-feature bottom rank matrix representation model.

\[
\min_{Z_1, \ldots, Z_K, E_1, \ldots, E_K} \sum_{i=1}^{K} (\|A_i\|_1 + \lambda \|E_i\|_{2,1}) + \alpha \|A\|_{2,1}
\]

s.t. \( X_i = X_i A_i + E_i, i = 1, \ldots, K \)

Where \( \alpha \) is a coefficient greater than 0, \( \|E\|_{2,1} = \sqrt{\sum_j \left( \sum_i (E_{i,j})^2 \right)} \) Used to measure errors caused by noise and wild spots; Equivalent to the following model:

\[
\min_{J_1, \ldots, J_K, S_1, \ldots, S_K, Z_1, \ldots, Z_K, E_1, \ldots, E_K} \sum_{i=1}^{K} (\|J_i\|_1 + \lambda \|E_i\|_{2,1}) + \alpha \|A\|_{2,1}
\]

s.t. \( X_i = X_i S_i + E_i \), \( A_i = J_i \), \( A_i = S_i, i = 1, \ldots, K \)

- Decompose and solve the model to obtain a sub-model;
- output the pseudo area and get the final accurate area;
- outputting an image of a combined feature vector of the target image;

The specific steps of extracting the adaptive LBP operator feature algorithm are as follows:

- Convert the image of the input system into a grayscale image, sum the grayscale values of the image pixels, and then obtain the average value:

\[
\text{avg} = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} \text{gray}(i,j)}{m \times n}
\]

- The total texture feature is used to remove the background, and the sum of the absolute values of the difference between the pixel gray value of the image and the average pixel gray value is calculated, and the average value is obtained:

\[
\text{Diffsum} = \sum_{i=1}^{m} \sum_{j=1}^{n} \left| \text{gray}(i,j) - \text{avg} \right| \quad \text{Davg} = \frac{\text{diffsum}}{m \times n}
\]

The background texture is used to remove the background, and the sliding window of 3×3 size is used to traverse the image to obtain the difference between the gray value of the central pixel and the gray value of the surrounding pixels, and the average value is obtained in each window image:

\[
\text{Areavg} = \frac{\sum_{i=0}^{7} \sum_{j=0}^{7} |g_i - g_c|}{8}
\]

- Fitting the method of calculating the adaptive threshold: \( \theta = 4 \times \sqrt{\text{Davg} + \text{Areavg}} \). Fixed other variables to solve \( I_i \):

\[
I_i = \arg \min_{J} \frac{1}{\mu} \|J\|_1 + \frac{1}{2} \|J - \langle A_i + Wz/\mu \rangle \|_F^2
\]

Solving with a singular value threshold.

The output pseudo region and get the final accurate region:

- The circumscribed matrix leaving each subspace according to the size and proportion of the area is a suspected area;
- Set a jump function \( f(i,j) \) to accurately locate the suspected area and determine the upper and lower boundaries of the precise area:
\[ f(i, j) = \begin{cases} 1, & 60 \leq c(i, j) < 75 \text{ or } -124 < c(i, j) \leq -60 \\ 0, & \text{other} \end{cases} \]

\[ c(i, j) = LBP_{8,1}(i, j) - LBP_{8,1}(i, j-1) \]

Among them, \( i = 1, 2, 3, 4, \ldots, N \), \( j = 2, 3, 4, \ldots, M \), Therefore, the sum of the number of transitions of any row \( i \) is:
\[ S(i) = \sum_{t=2}^{N} |f(i, j)| \]

If the sum of the number of jumps in any row \( S(i \geq 12) \), then this line belongs to the precise area; Scan the entire image from top to bottom, Find all rows that satisfy \( S(i \geq 12) \), and record the number of downlinks \( i \); If there are consecutive \( h \) lines that satisfy \( S(i \geq 12) \), then, a rectangular area with a width of \( M \) and a height of \( h \) can be obtained. This area is a precise area, and an area in the image that does not have this feature is excluded.

3. Method review

In summary, according to the color histogram, the information of each color dimension of each region in multiple color spaces is extracted, represented by a feature vector, and finally the color information of each region in multiple color spaces is extracted; Combining a plurality of color space color information extracted from at least two regions to obtain a combined feature vector of the target image; The information of each region in each color dimension of each color space is combined to obtain a combined feature vector of each color space in each region. The combined feature vectors of the respective color spaces are spliced to obtain a combined feature vector of each region in a plurality of color spaces, and a combined feature vector of the plurality of color spaces of the target image in at least two regions is obtained. Using the classifier to classify the samples to be tested, and obtaining classification results corresponding to the categories of the samples to be tested, including: projecting the sample to be tested by using the projection matrix to obtain a two-dimensional robust feature to be tested of the sample to be tested; Using the two-dimensional robust feature to be tested as an input of the classifier, obtaining at least one category corresponding to the sample to be tested, and determining at least one of the obtained categories corresponding to the sample to be tested and the category with the greatest measure of similarity of the sample to be tested is the category of the sample to be tested.

The present invention provides a geometric image recognition feature extraction system comprising: a two-dimensional discriminant feature learning module for performing two-dimensional discriminant feature learning on the original training samples included in the original training set; Image feature learning modeling by compact local intra-class divergence and separation of local inter-class divergence can effectively maintain the topological structure and intrinsic correlation between image pixels. Based on the 1 norm metric, the robustness of image description is improved. And, by optimizing a feature decomposition problem, obtaining a projection matrix applicable to the two-dimensional robust feature extraction of the image sample; wherein the original training sample is a sample having a category label corresponding to the category of the original training sample.

A classifier module, acquiring a new training sample set including two-dimensional robust features of each of the original training samples, and constructing a classifier by using the new training sample set; A two-dimensional robust feature of each of the original training samples is obtained by projecting each of the original training samples using the projection matrix; A classification result module, which uses the classifier to classify the sample to be tested, and obtains a classification result corresponding to the category of the sample to be tested, wherein the sample to be tested is a sample whose category is unknown; A target image acquisition module, which acquires a target image from a sample classification result to be tested; And a target image division module that divides the target image into at least two regions according to a direction in which the person in the target image is upright; A combined feature vector obtaining module of the target image, extracting color information of each region in a plurality of color spaces; The color information of the plurality of color spaces extracted from the at least two regions is combined to obtain a combined feature vector representing the target image.
In this paper, the method of dividing the spatial region of the image in the direction of the person in the target image is used. In each region, the color information of the plurality of color spaces is extracted separately, and the three-dimensional histogram of each color space is spliced into a single dimension, and then connected. The histogram of different color spaces in the region and the histogram connecting all the regions obtain the combined feature vector of the multi-region multi-color space, thereby overcoming the limitation in the tracking recognition of the target object, increasing the information amount of the color space feature, and being efficient at the same time. Characterizing the spatial information of the target object; At the same time, the invention not only can effectively maintain the topology and intrinsic correlation between image pixels, but also can effectively reduce the complexity of the model calculation process. In addition, based on the 1 norm metric, noise robustness during feature extraction can be ensured. Therefore, the above solution provides the feature extraction and classification of the image directly, thereby effectively improving the efficiency of identifying the image and the accuracy of classifying the image, and the system performance is good and the scalability is good.

4. Conclusion
This paper obtains precise regions based on color and LBP feature operators, combined with improved LRR models and morphological operations. The combination of multiple features can effectively improve the robustness and accuracy of image detection, reduce false detections, and further texture analysis of images in complex backgrounds, providing more accurate feature information.

Acknowledgments
This article is the research results of the following fund projects: General scientific research project of Gansu Provincial Department of Education, Applied Research on Feature Extraction and Recognition Algorithm of Minority Pattern Art in Gansu Province, No.2018A-128; Youth Science and technology innovation project of Lanzhou Institute of technology 2019, Research on personalized Resource Recommendation of online learning platform based on Knowledge Representation. No.2019k-009; The National Social Science Foundation of China under Grant No.15xmx035; Lanzhou Science and Technology Development Project under Grant No.2015-3-30; Funded by the ‘Kaiwu’ Innovation Team Support Project of Lanzhou Institute of Technology, No.2018KW-07.

References
[1] Biederman I., Ju G. Surface versus edge-based determinants of visual recognition[J] Cognitive Psychology, 1988. 20(1): 38-64
[2] Belongie S., Malik J., Puizicha J. Shape Matching and Object Recognition Using Shape Contexts[C]. 2006: IEEE Computer Society. 509-522
[3] Ling H., Jacobs D.W. Shape classification using the inner-distance[J].IEEE Transactions on Pattern Analysis & Machine Intelligence, 2007. 29(2): 286-299
[4] Felzenszwalb P.F., Schwartz J.D. Hierarchical Matching of Deformable Shapes[C]. 2007. CVPR ‘07. IEEE Conference on. IEEE, 2007:1-8
[5] Kumar N., Belhumeur P.N., Biswas A., et al. Leafsnap: a computer vision system for automatic plant species identification[C]. 2012. European Conference on Computer Vision. Springer-Verlag, 2012:502-516
[6] Hu R., Jia W., Ling H., et al., Multiscale Distance Matrix for Fast Plant Leaf Recognition[J]. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 2012. 21(11): 4667-4672
[7] Alajlan N., El Rube I., Kamel M.S., et al. Shape retrieval using triangle-area representation and dynamic space warping[J]. Pattern Recognition, 2007. 40(7): 1911-1920
[8] Wang X., Feng B., Xiang B., et al. Bag of contour fragments for robust shape classification[J]. Pattern Recognition, 2014. 47(6): 2116-2125
[9] Yang C., Wei H., Yu Q. Multiscale Triangular Centroid Distance for Shape-Based Plant Leaf Recognition[C]. European Conference on Artificial Intelligence. 2016
[10] Abbasi S. Efficient and Robust Retrieval by Shape Content through Curvature Scale Space[M], Image Databases And Multi-Media Search. 1996:51-58
[11] Adamek T., O’Connor N.E. A multiscale representation method for nonrigid shapes with a single closed contour[J]. Circuits & Systems for Video Technology IEEE Transactions on, 2004. 14(5): 742-753
[12] Rojanamontien M., Watchareeruetai U. Shape recognition by using Scale Invariant Feature Transform for contour[C]. International Joint Conference on Computer Science and Software Engineering. IEEE, 2017:1-6
[13] Jia Q., Fan X., Liu Y., et al. Hierarchical projective invariant contexts for shape recognition[J]. Pattern Recognition, 2016, 52:358-374