Research on image matching methods in computer vision

Zizhou Shi

School of Yali high school, No. 428 Laodong Western Road, Changsha, China

Abstract. Today, computer vision has shown a variety of roles in our lives, making people's life more convenient. Also, a variety of artificial intelligence models and algorithms have emerged for computer vision. Image matching is an important technique in the field of computer vision to find the similarities between two images or multiple images with the help of matching algorithms to achieve scientific and accurate processing of images. Therefore, summarizing the effective approximate solution to this problem as well as the future research is the main part for current research. The paper firstly describes the basic elements of the image matching technique. Then, some representative traditional image matching algorithms proposed in the field of computer vision research in recent years are summarized and reviewed. Finally, the future research directions and research ideas of image matching are discussed, providing reliable guidance and reference for subsequent research.

Keywords: Computer vision, image matching, attention methods, development trend.

1. Introduction

With the development and application of artificial intelligence, people's lives have improved a lot. For example, the application of automatic driving technology has greatly reduced the frequency of traffic accidents, and driver fatigue from driving for a long time has also been lessened. Also, Google Translate allows people to translate languages in different countries in real-time using the camera, facilitating communication between people in different countries. Among these achievements, computer vision has played an important role. Computer vision is a field of artificial intelligence that enables computers and systems to derive meaningful information from digital images, videos, and other visual inputs. When the images are sent to a computer, it interprets them as a collection of pixels. After these pixel combinations are processed and input into the model for training, the computer can distinguish the similarities and differences between different images, and then achieve image recognition. Image matching is the key step of image recognition technology. Image matching is a process of looking for the target image in the original image, by analyzing the content, features, structure, relationship, texture, and grayscale value, and then judging the similarity and consistency between images. It has been applied to many aspects of life, such as license plate recognition, and face recognition.

The literature summarizes image matching algorithms and its characteristics, and the literature discusses the future research outlook. The purpose of this paper is to provide a more comprehensive literature review for scholars in this direction and also to provide a reference for those who are going to conduct research in this field.

2. Components of graph matching methods in computer vision Essentials

The principle of different image matching algorithms is different, but the process of image matching is basically the same. Image matching algorithm is composed of the following four parts, i.e., similarity measurement, feature space, search space, and search strategy.

Similarity measurement is a comprehensive measure to assess the degree of similarity between two things. The closer two things are, the greater their similarity measure will be, while the more different two things are, the smaller their similarity measure will be. In image matching, people can adjust the accuracy of image matching by adjusting the similarity measurement.

Feature space refers to the specific parameters and conditions of the picture in the link of picture matching. In computer image processing, it is necessary to grasp the relationship between these specific parameters and keep them relatively average to make the training achieve a better effect [1].

The Search space is the space with estimated parameters. The parameter content is preliminarily summarized to form a certain range of space, and finally, the parameter interval of different graphs is matched. Such matching is more in line with the running characteristics of the computer so that the computer can better achieve recognition.

The search strategy is the path and solution chosen for the search. By comparing the image parameters, a suitable search scheme is selected, and the choice of search scheme determines the accuracy and comprehensiveness of the search. The optimization of the search strategy is the best way to control the matching error and improve the matching quality.

3. Image matching algorithm

According to the principle, these algorithms can be divided into image matching algorithm based on gray value, image matching algorithm based on feature, image matching algorithm based on frequency domain analysis, image matching algorithm based on neural network, and image matching algorithm based on semantic recognition.

3.1. Image matching algorithm based on gray value

These types of algorithms will compare the grayscale matrix of the target image with all possible window grayscale matrices in the test image, and then use a method of similarity measurement to obtain the optimal solution as the matching result. There are many image-matching algorithms based on the gray value, like NCC (normalized cross-correlation), MAD (Mean Absolute Differences), SAD (Sum of absolute differences), SSD (Single Shot MultiBox Detector), and MSD (Most significant digital). One representative algorithm NCC is introduced in detail as follows.

NCC is a matching method based on image gray-level information. In the matching process, the computer will specify a matrix of a certain size, search and match in the gray array of the image according to the similarity measure method, and return the matching result between [-1, 1]. The closer the result is to 1, the more similar the images are, and vice versa.

S is the target image, and its size is M*N; T is the test image, and its size is m*n. R is the matching value returned by NCC calculation. The numerator is the sum of the gray value of a point minus the average gray value, and the denominator is the multiplicative accumulation of the gray standard deviation of two images

$$R(i,j)=\frac{\sum_{m=1}^{M}\sum_{n=1}^{N}[S^{i,j}(m,n)-\overline{S^{i,j}}]\times[T^{m,n}-\overline{T}]}{\sqrt{\sum_{m=1}^{M}\sum_{n=1}^{N}[S^{i,j}(m,n)-\overline{S^{i,j}}]^{2}\cdot\sum_{m=1}^{M}\sum_{n=1}^{N}[T^{m,n}-\overline{T}]^{2}}}$$

The NCC algorithm has strong anti-white noise ability and high accuracy under the condition of little gray level variation and geometric distortion. But at the same time, this method is susceptible to local illumination changes, and it needs to traverse all the combinations of matrices in the image, which takes a long time to compute.

3.2. Image matching algorithm based on feature

Another important image-matching algorithm is based on the feature. Firstly, it will extract the features of the image and regenerate them into feature descriptors; then, it will match the features of two images according to the similarity degree of descriptors. Image features can be divided into points, lines (edges), regions (surfaces), and so on, as well as local and global features. Because the feature extraction of regions (surfaces) is comparably complex and time-consuming, people mainly choose to use point features and line (edges) features.

Line features are important features of the image and are the discontinuity of the distribution of the image's characteristics (such as pixel gray and textures). The edge part of an image concentrates most of the information of the image, and the edge structure and characteristics of an image are often

an important part of determining the characteristics of the image. Image edge detection is mainly used to enhance the contour edges, details, and gray jumping parts in the image to form a complete object boundary, so as to achieve the purpose of separating the object from the image or detecting the area representing the same object surface. By far the most common method is to detect discontinuities in luminance values by the first and second derivatives. The first method is differentiation. The purpose of the differential method is to use the differential operation to find the rate of change of the signal, and strengthen the role of the high-frequency component so that the outline is clear.

Feature points are some special places in the image. We can put the images in the corner, edge, and block as the image represents. It is made up of Key-point and Descriptors. The key point is the position of the feature point in the graph, some also include information about orientation and size, and the descriptors are usually vectors that describe the key information of surrounding pixels according to some designed way. The descriptor is in accordance with the principle of "appearance similar features should have a similar descriptor". Therefore, as long as the descriptor of the two feature points is at a similar distance on the vector space, people can consider them to be of the same feature points. So, an intuitive method to extract features is to identify the angular point between different images and determine their corresponding relationship. In this approach, the corner is known as the characteristic. However, simple feature points aren't suitable for some cases. For example, someplace that seems to be a feature point might change after getting a closer view, and the appearance of feature points might change when the camera turns. Therefore, people designed some more stable features like SIFT (scale invariant feature transform).

In 2004, D.Lowe, University of British Columbia, came up with a new algorithm, Scale Invariant Feature Transform (SIFT) in his paper, Distinctive Image Features from Scale-Invariant Keypoints, which extract keypoints and compute its descriptors [5]. There are mainly four steps involved in SIFT algorithm. The first step is scale-space extrema detection. In it, Laplacian of Gaussian is found for the image with various σ values. LoG acts as a blob detector which detects blobs in various sizes due to change in σ. In short, σ acts as a scaling parameter. But this LoG is a little costly, so the SIFT algorithm uses Difference of Gaussians which is an approximation of LoG. The second step is keypoint localization. Once potential keypoints locations are found, they have to be refined to get more accurate results. They used Taylor series expansion of scale space to get a more accurate location of extrema, and if the intensity at these extrema is less than a threshold value, it is rejected. The third step is orientation assignment. Orientation is assigned to each keypoint to achieve invariance to image rotation. A neighbourhood is taken around the keypoint location depending on the scale, and the gradient magnitude and direction is calculated in that region. The fourth step is keypoint descriptor. Descriptor is created and it is represented as a vector to form keypoint descriptor. And finally matching the keypoints, keypoints between two images are matched by identifying their nearest neighbors [6].

Because SIFT feature is a local feature of image, it keeps invariant to rotation, scale scaling and brightness changes, and also maintains a certain degree of stability to Angle of view change, affine transformation and noise. At the same time, it also has the property of multi-quantity, which can extract a large number of feature vectors from a small amount of data.

3.3. Image matching algorithm based on neural network

It is an algorithmic mathematical model that simulates the behavior characteristics of an animal neural network and performs distributed parallel information processing. This kind of network depends on the complexity of the system, by adjusting the relationship between internal nodes to achieve the purpose of processing information. If the image data is processed and input into the artificial neural network for training, people can train the model that can recognize the image [3].

At first, process images asi required, graying the color of them, correcting them by rotation, and extracting their features. Then, design the network structure, select the appropriate input layer, hidden layer, and output layer. After building the network structure, select appropriate samples to train it. In the end, get a model that can be used to recognize images.

Its matching accuracy is high, and has characters like Generalization, nonlinear mapping, adaptive and self-organization, and high degree of parallelism abilities. But it also needs a lot of time to train the model, so its learning efficiency is comparably low.

3.4. Image matching algorithm based on frequency domain analysis

It converts the original image and target image in the spatial domain into the frequency domain by mathematical transform and uses similarity measurement to determine the matching degree between these two images. Such algorithms mainly include the image matching algorithm based on the Fourier transform, image matching algorithm based on the Gabor transform and image matching algorithm based on wavelet transform [2].

This type of algorithm has a little dependence on the grayscale value, the sharp detection peak, and high matching precision. It also needs the conversion between the spatial and frequency domains, so the matching efficiency is undesirable.

3.5. Image matching algorithm based on semantic recognition

The semantic recognition will use the image semantic information like color, lines, and contours. And by analyzing the features, the high-level semantic features are constructed. The semantic features are used to judge the similarity of the images to locate the target image and realize the image matching [4].

The semantic features represent the understanding of the image content, close to the human's intuitive perception of the image meaning, which is the key step to realize intelligent image matching.

Nevertheless, core problems still exist to be solved: how to get high-level semantic features from visual features processing? How to extract the semantic features accurately and make them consistent with the image meaning? How to solve the problem that image meaning is subjective……researchers still need to do subsequent analysis to figure out these problems' answers.

4. Conclusion

This paper set out to compare the advantages and disadvantages of different image recognition algorithms. It explains in detail the principle and advantages and disadvantages of image recognition technology such as gray algorithm and feature algorithm and provides a deeper understanding of image recognition technology based on artificial intelligence. However, this paper also has several limitations: it only explains a few representative algorithms and does not comprehensively introduce the field of image analysis. We hope that this article can provide guidance and help to people concerned with computer vision research so that they can determine the algorithms and models suitable for their own project easily.

References

[1] J Wang, M Zhang, "Research progress of image matching algorithm," Journal of atmospheric and environmental optics, 2 (1), 11-15 (2007).

[2] J Liu, X He, W Chen, "A9 SIFT feature image matching algorithm based on Wavelet Transform." Computer simulation, 28 (1), 257-260 (2011).

[3] X Huang, W Bao, "Image feature point matching algorithm based on improved Hopfield neural network", 31 (9), 1961 - 1964 (2010).

[4] Y Zhao, "Research on image semantic representation and metric learning," Henan: The PLA Information Engineering University, 1-123 (2015).

[5] David G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," Computer Science Department University of British Columbia, (2004).

[6] J Zhang, J Li, S Zhang, "A hybrid image matching algorithm based on line feature and SIFT point feature".