The application of computer vision in logistics warehouse security management

Mengxiao Wang, University of Minnesota

Abstract: The safety management of logistics and warehousing plays an important role in the survival of modern enterprises. With the development of China's logistics industry and the information age, the application of computer vision technology, intelligent video monitoring system, security simulation technology to achieve the security of logistics warehouse, logistics warehouse management intelligent and modern. In the actual logistics warehousing scene, the external light changes, noise interference and the complex situation of the scene will affect the detection of logistics targets. How to obtain a good representation of the target is the key to detect and extract the target that users are interested in effectively and accurately. To solve this problem, a target tracking algorithm based on correlation filtering and convolutional neural network is proposed in this paper, and the feasibility of the algorithm is verified by experimental results.

Key words: computer vision; Image processing; Correlation filtering; Convolutional neural network; Target tracking; Fourier

0. Introduction

According to the national standard of Logistics Terminology, the concept of modern logistics refers to the flow process of corresponding goods from the place where the goods are supplied to the place where the goods are received. With the continuous development of our society, the rapid and vigorous development of e-commerce such as JINGdong, Suning and Taobao has brought development to China's logistics industry, but also brought serious challenges. E-commerce and logistics industry promote each other, so that e-commerce has become a powerful force in China's logistics market. In recent years, China's information technology has made rapid development, the Internet of Things, big data, artificial intelligence and other technologies continue to emerge, and has been integrated into the society in all industries, this to China's traditional industry into a new boost, China's logistics industry is no exception. The new technology represented by computer vision and intelligent optimization algorithm has been integrated into China's logistics industry, promoting the rapid development of China's logistics industry. As the safety management of logistics and warehousing is of great significance to the logistics industry, with the continuous development of e-commerce and new technologies, there are new and higher requirements for the safety management of logistics and warehousing, and the application of intelligent and modern logistics and warehousing safety management is born. This paper will discuss the application of computer vision in logistics and storage safety management.

1. Computer vision

The concept of computer vision is to put forward important symbols or values in videos and images, analyze these information fully, and finally detects, recognize and track targets. Simply put, computer vision is about allowing the computer to see and understand real life images [1]. In the 1950s, computer vision emerged and was subsequently applied to the recognition and analysis of two-dimensional images, such as microscopic images or optical characters [2]. In the 1960s, researchers used computer programming languages to convert 2d images into 3D structures and analyze these 3d structures. In the 1970s, the Artificial Intelligence Laboratory of Massachusetts Institute of Technology opened a computer vision course taught by Professor Horn, and Professor Mart (the same lab as Professor Horn) proposed for the first time that the most important problem in vision research was representation [3]. By the 1980s and

1990s, computer vision had made rapid development and formed a new theoretical framework based on perceptual features, which was gradually applied to the industrial environment [4]. In the 21st century, computer vision has many new development trends. For example, computer graphics and computer vision have been deeply integrated, and many applications based on computer vision have emerged in an endless stream. Computer vision technology has various forms of applications in security, logistics, transportation, medical treatment and robotics [5].

Generally speaking, target tracking is one of the basic problems in the field of computer vision. Its task is to continuously estimate the trajectory of the image sequence in the following successive frames while determining the initial state of the target. In many fields of real-time vision, target tracking plays a very important role, especially in logistics and warehouse security management. At present, target tracking model can be generally divided into generative and discriminative models. Model generation is mainly to form a fixed model according to the performance characteristics of the target, and then conduct minimized pattern matching under the condition of the model to find the most appropriate matching window [10]. As a typical generated model tracking algorithm [1], L1 APG (Accelerated Proximal Gradient) uses the dictionary sparse to represent the candidate target and takes the candidate target with the minimum reconstruction error and the sparsest coefficient as the tracking result. Discriminant tracking model is to classify tracking problems into binary categories, train classifiers with training data, and distinguish targets in the background. A relatively classical tracking algorithm is KCF (Kernelized Correlation Filters) [2]. It mainly uses ridge regression model and performs Fourier transform on the template with the introduced loop structure. Since it does not invert the matrix in ridge regression, the tracking efficiency and speed are greatly improved. In recent years, target tracking algorithm based on CF (Correlation Filter) has attracted extensive attention, and its computational efficiency and competitive effect highlight great advantages. CF adds Fourier transform to realize the effect of reducing computation. This idea has led to many characteristic tracking algorithms, such as KCF tracking algorithm with multi-channel characteristics. In 2016, Martin Danelljian is proposed on the basis of relevant filtering algorithm, with CNN (Convolution Neural Network, convolution neural network) + HOG + CN as feature combination, this method greatly reduced the feature dimension and application characteristics of the original subsets, realized the feature extraction of simplified, avoids the filter redundant [3]. However, CF - based target tracking still has some disadvantages, such as manual feature extraction, which cannot capture semantic information of the target, and lack of training data. In order to overcome the above problems, some researchers have introduced the feature of deep convolution. Although this method [4] can achieve certain effects and improve robustness, it cannot capture or track real-time targets. Aiming at the problem of large computational load of extracting CNN features, the improved CNN algorithm is fused with relevant filters, and FFT (Fast Fourier Transform) is introduced to reduce the computational load.

2. Correlation filtering

Generally speaking, correlation filter is a kind of learning discriminant classifier, which mainly determines the target object by searching the maximum response value of scene graph. Simply put, in the selected scenario, the response to the corresponding background is lower and the response to the target of interest is higher. A method or method of forming a multichannel image or data through a single channel signal, which can simplify symbols. However, in the actual operation process, not only one-dimensional single-channel images are processed, but more gradient direction histogram (HOG) and multi-color images (three channels of R, G and B) are processed.

F is taken as the training signal of MXN, and the training sample is all cyclically shifted F. $f_{m,n}\in\{0,1,...,M-1\}X\{0,1,...,N-1\}$ as the shifted sample, g(m, n)= $e^{-\frac{(m-M/2)^2+(n-N/2)^2}{2\sigma^2}}$ is the Gaussian function, δ is the kernel size. The correlation filter H with the same size f is:

$$\min\|h\otimes f-g\|_2+\lambda\|h\|_2 \tag{1}$$

In the formula above, λ is a regularized parameter, $\otimes$ is a circular convolution symbol. The target position is searched by Applying Fourier change through Formula 2, which is as follows:

$$H^*=\frac{G\bullet F^*}{F\bullet F^*+\lambda} \tag{2}$$

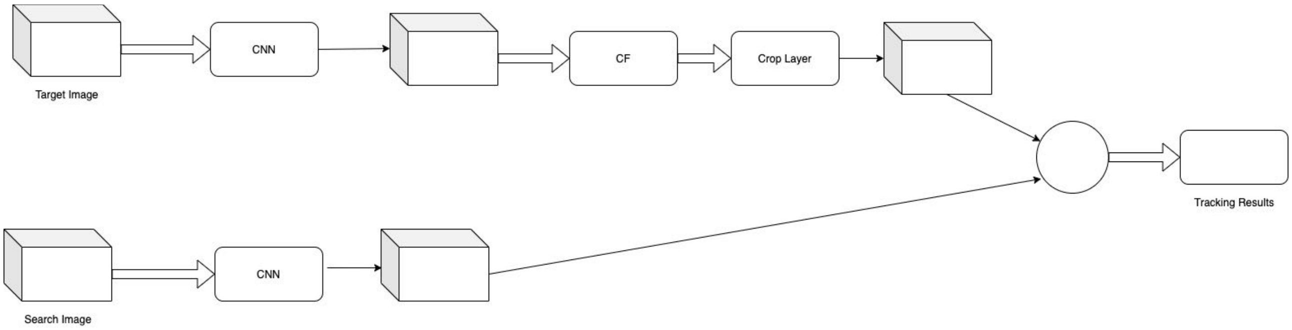
In the formula above, $\bullet$ is the operation of multiplying two vectors, $*$ is the conjugate symbol. $F=$ £ (f), $H=$ £ (h), $G=$ £ (g). Once the search image Y (next frame) arrives, the reaction graph Z is described as follows:

$$z=£^{-1}(H\bullet Z) \tag{3}$$

Formula (2) and formula (3) are combined as follows:

$$z=£^{-1}(\frac{G\bullet F}{F\bullet F+\lambda}\bullet Z)=£^{-1}(\frac{G\bullet(F\bullet Z)}{F\bullet F+\lambda}) \tag{4}$$

Relevant filter tracking algorithm, in general, flexible use of cyclic matrix migration produces some of the classifier training samples, the samples matrix has the characteristics of cyclic matrix, easy will change to solve problem of matrix for the Fourier Ye Yu in solving vector dot product with low amount of calculation, this greatly reduces the amount of calculation of the algorithm. After a lot of calculation, the optimal classifier is obtained, and the new target position is the position at the maximum response value. Finally, the fast detection target is realized, and the new target position is applied to update the classifier. The target tracking algorithm of correlation filtering generally adopts a fixed target scale, but when the target is blocked, the target scale changes or the target is lost, there is no corresponding method or measure to deal with the above problems. In order to optimize the correlation filter module, this paper proposes a symmetric network structure based on correlation filter, which is a multi-domain learning framework based on CNN, where useful expressions are obtained from a specific domain and domain-independent expressions are separated [5]. The system frame diagram is as follows:



2. Correlation filtering and correlation filtering network

As shown in Figure 1, this method applies the framework network based on multi-domain learning and integrates the correlation filter module of X and correlation operation. The formula for the above changes is as follows:

$$\mathrm{h}_{p,x,p}(x',y') = sw(f_p(x'))f_p(y')+b \tag{5}$$

In the above formula, y' represents the search area, x' represents the target area, and f_p represents the CNN with learning rate of P.

ω=ω(x)

The calculation of ω=ω(x) of CF module realizes the standard template obtained from the training feature graph X, thus solving the ridge regression problem in The Fourier domain [2]. Where the two scalar parameters s(weight) and b(deviation) are 2, they make the fraction range more suitable for logistic regression.

A large-scale context-based image correlation filter is very important for the training process. The idea of minimum squares was added. Although good results were achieved, this would introduce the boundary problem of CF into the network, so the Crop layer [6] was increased and the middle part was retained. The forward propagation of the network added a CNN-based CF tracker, which was unable to realize the end-to-end training of previous algorithms. This paper proposed a method to realize the end-to-end training of CF. Simply put, it is to input the derivative in the template of CF to realize the end-to-end training of CF.

3. Convolutional neural network

Convolutional neural network is a kind of multi-layer neural network, which is good at processing the related machine learning problems of images, especially large images. Through a series of methods, convolutional network successfully reduces the dimensionality of the image recognition problem with a large amount of data, and finally enables it to be trained. Classical convolutional neural network mainly includes pooling layer, convolutional layer, full connection layer, pooling layer and Softmax regression layer. Convolutional layer is a very important layer structure in CNN network, through which the feature graph is obtained, and the quality of the graph directly affects the processing of subsequent layers. To put it simply, the convolutional layer applies the feature map of the previous layer and the convolution kernel for local connection to obtain the local features of the image, and finally calculates the shared weight value to the new feature map. The pooling layer is also called the lower sampling layer. Its task is to undertake the convolutional layer, which is mainly to carry out corresponding feature dimensionality reduction for some feature diagrams behind the convolutional layer, thus reducing the computation and the complexity of the network. All connection also known as special convolution layer, and the convolution layer different points, one by one in the whole connection layer neurons and before all the neurons in a layer of connection, the effect is a dimension transformation, simply put, is a high dimension matrix data into low dimension matrix, to consolidate and then have the characteristics of the differential capacity [11]

In computer vision, CNN has achieved good results and been widely used. Literature 7 applied large-scale data sets to train CNN and efficient GPU, and finally realized image classification and improved performance [7]. In 2018, Harbin Institute of Technology proposed STRCF (spatial-temporal regularization Correlation Filters) algorithm by adding Temporal and Spatial regularization into the framework of DCF (Discriminative Correlation Filters). This method has achieved good results in the field of Correlation filtering and tracking [8]. The online PA based approach is similar to SRDCF (Spatial Regularized Discriminative Correlation Filters) for multiple training images and is more robust when the appearance is Discriminative.

Although CNN has achieved great success, tracking algorithm cannot improve performance due to lack of large-scale training data. Literature [9] proposed a learning method based on CNN pool [9], but compared with the accuracy of manual feature extraction method, the performance of this method is not greatly improved, and the training data depth network is lacking. Literature [10] proposes a new method, which constructs a large data set for image classification and realizes the transfer of pre-training [10]. However, the difference between tracking task and classification task is not obvious.

Different from other methods, the algorithm proposed in this paper applies large-scale visual tracking data, and then pre-trains CNN, and finally achieves good results.

4. Multi-domain learning network

In the text, in order to train pre-training depth CNN, multi-Domain Network (MDNet) for multi-domain training is applied [5]. MDNet refers to a learning method that applies training data from multiple fields and domains to the learning process. Multidomain learning has been widely used in natural language processing. However, in computer vision, there are few discussions on multi-domain learning. For example, Duan et al. used the domain-weighted combination of Hoffman and SVM for video concept detection, and then proposed the classification of object mixed transformation model.

MDNet is divided into a domain-specific layer and a shared layer. A specific domain layer has a dichotomy layer for a class of objects, which is used to distinguish background from foreground. The sharing layer is mainly used to learn general object representations. The network mainly includes the RGB-type image that receives input, and it has 5 hidden layers (including 2 full connection layers and 3 convolutional layers). Finally, corresponding to the K domain, a full connection layer mainly passes through K branches ($fc6^1$—$fc6^K$) (training sequence). The convolutional layer is the part of the network corresponding to VGG-M. The size of the feature map is adjusted mainly through the input size. The two full connection layers behind it correspond to 512 output units. In order to distinguish the background and target of each field, each K branch contains a binary classification layer with a cross entropy loss classifier. The $fc6^1$—$fc6^K$ domain specific layer and the preceding layer are used together as the shared layer. MDNet networks have many advantages, such as a smaller network size compared to the usual recognition network, the application of some special tracking data for training, and in order to effectively distinguish between background and target, specific domain classification for the same class of objects.

5 Tracking algorithms

The network itself is mainly used to measure only the similarity between two image blocks and to track and evaluate the network forward propagation online. In order to better use this network in the target tracking of images, it should be closely combined with the tracker logic program. This algorithm mainly uses simple tracking algorithm to evaluate (the practicability of similar functions). Some evaluation of online tracking algorithms mainly apply simple forward mode to network evaluation. Generally speaking, the target position estimated in the previous frame of the latest frame is taken as the center, and the search area is extracted, and then the search area is compared with the target's features. The position with the highest score is the target's new position.

6. Experimental results and analysis

In order to verify the tracking performance of target tracking algorithm (KCF) based on correlation filtering and convolutional neural network in intelligent monitoring environment, three video sequences with different backgrounds and resolutions were selected for comparison experiment. The data is extracted from the video information extracted from the warehouse of an e-commerce company, as shown in the following table.

Table 1. Test video details

Video	Name	Resolution	Frame Number
Video 1	Test1	352*288	241
Video 2	Test2	640*360	292
Video 3	Test3	640*480	901

In order to verify the feasibility of the algorithm, the text algorithm and KCF algorithm were compared and analyzed in the three videos in table 1 above. The average results are as follows.

Table2. Comparison of the median accuracy and running speed of the two algorithms

Algorithms	Median Accuracy	Running speed/fps
Text algorithm	85.1	75.3
KCF	84.6	72.5

For video sequences of different scenes, the tracking effect of KCF algorithm is very good. The experimental results show that the algorithm is robust to rapid movement, short-term occlusion, scale change and chaotic scene, and it also improves the adaptability of different tracking scenes and verifies the feasibility of the algorithm.

7. Conclusion

In the logistics warehousing scene, the external light changes, noise interference and the complex situation of the scene will affect the detection of the logistics target. With convolution neural network of target tracking based on correlation filtering algorithm can effectively and accurately detect and extract the user interested target, improve the computer vision technology in logistics warehouse space management, the problems of and identify items of state information and location information, realize the efficient management of warehouse space, enhance the informationization level of warehouse management system, can better adapt to modern logistics management mode for warehouse space control requirements.

References

[1] Ji H . Real time robust L1 tracker using accelerated proximal gradient approach[C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2012.

[2] 罗海波, 许凌云, 惠斌,等. 基于深度学习的目标跟踪方法研究现状与展望[J]. 红外与激光工程, 2017, 046(005):6-12.

[3] Danelljan M , Robinson A , Khan F S , et al. Beyond Correlation Filters: Learning Continuous Convolution Operators for Visual Tracking[J]. 2016.

[4] He Z , Zhang Z , Jung C . Fast Fourier Transform Networks for Object Tracking Based on Correlation Filter[J]. IEEE Access, 2018:1-1.

[5] Zhang X , Zhang X , Du X , et al. Learning Multi-Domain Convolutional Network for RGB-T Visual Tracking[C]// 2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). 2018.

[6] Jack Valmadre, Sridha Sridharan, Simon Lucey. Learning Detectors Quickly with Stationary Statistics[J]. lecture notes in computer science, 2014, 9003:99-114.

[7] Krizhevsky A , Sutskever I , Hinton G . ImageNet Classification with Deep Convolutional Neural Networks[C]// NIPS. Curran Associates Inc. 2012.

[8] Li F , Tian C , Zuo W , et al. Learning Spatial-Temporal Regularized Correlation Filters for Visual Tracking[J]. 2018.

[9] LI H, LI Y, PORIKLI F. DeepTrack: Learning Discriminative Feature Representations by Convolutional Neural Networks for Visual Tracking [C] / / British Machine Vision Conference. Nottingham,2014.

[10] 赵明瀚, 王晨升. 基于视频的人数识别方法综述[J]. 软件, 2013(03):18-20+62.