applied sciences

MDPI

Article

An Automatic Counting Algorithm for Topographic Maps Based on Videos for Map Management

Wen Cao [1], Yuzhen Tian [1], Xiaochong Tong [2,*], Weiming Yang [3], Congzhou Guo [4], Jingwen Zhu [1], He Li [2] and Dali Wang [2]

[1] School of Geoscience and Technology, Zhengzhou University, Zhengzhou 450001, China
[2] School of Geospatial Information, Information Engineering University, Zhengzhou 450001, China
[3] Institute of Logistics Science and Technology, Academy of Military Science, Beijing 100166, China
[4] Department of Basic, Information Engineering University, Zhengzhou 450001, China
* Correspondence: txchr@zxiat.org

Abstract: Topographic maps comprise a rich data source for the processing and analysis of geographic information. However, the management of paper topographic maps is still dominated by an inefficient manual method, and there is no counting equipment for topographic maps that is suitable for narrow map warehouse passages. The existing algorithms relating to paper counting require paper hardness and cannot directly obtain the internal information of topographic maps. To address the above problems, using a map inventory machine transformed from a paddle-type paper counting machine to perform a preliminary count and obtain the video data of the counting progress, an automatic counting algorithm for topographic maps for map management is proposed in this study. First, the periodic feature changes of all the targets and each individual target in the videos are analyzed. Second, a deformable parts model accelerated by fast feature pyramids and by reducing the region of interest is used for the object detection of mechanical wheels. The grayscale variance function is selected to quantify the quality change in the images. Finally, a fusion window counting model is constructed with the adaptive window and fixed window methods. The experimental results show that the average counting accuracy of the model is about 95%, which can verify the initial counting results of the machine and provide high-quality keyframes for the subsequent sheet designation recognition.

Keywords: map management; topographic map; paper map; counting; deformable parts model; fast feature pyramids; fusion window counting model

Citation: Cao, W.; Tian, Y.; Tong, X.; Yang, W.; Guo, C.; Zhu, J.; Li, H.; Wang, D. An Automatic Counting Algorithm for Topographic Maps Based on Videos for Map Management. *Appl. Sci.* **2023**, *13*, 1461. https://doi.org/10.3390/app13031461

Academic Editor: Xianpeng Wang

Received: 22 November 2022
Revised: 6 January 2023
Accepted: 20 January 2023
Published: 22 January 2023

1. Introduction

A map is a ground model established by people after abstracting and generalizing the real world, and it is an important tool for spatial cognition [1]. As a kind of map, a topographic map comprehensively represents all kinds of geographical objects in detail, and it can meet the needs of map users in many ways. It provides basic data for national construction and original data for compiling other maps [2]. Topographic maps have always used paper (or similar materials) as an important medium because compared with electronic maps, paper maps play an irreplaceable role in the national economy and in national defense construction; their advantages include the overall macro, the ease of portability, the convenience for plotting, and other characteristics [3]. The topographic maps mentioned in this paper are all paper maps, and the topographic map style is shown in Figure 1. Due to the deepening application and the expanding requirements for topographic maps, their effective management has become an urgent demand in related fields. Currently, map suppliers mainly use map warehouses to manage maps. A map warehouse is a unit responsible for the acceptance, statistics, supply, and transportation of paper maps, and while map counting is tedious and time-consuming, it is very important

work [4]. Unlike a general warehouse, the overall space in a map warehouse is huge, and different map types are stored in map cabinets at different positions and heights. In addition, map cabinets are large and numerous; so, the storage and operation environment of maps is complicated and space is limited. In actual inventory work, the relevant staff need to use small and slow tunnel vehicles to get to and from different map cabinets to check maps manually according to the requirements of the outbound lists and then bring them back uniformly for a second verification. A lot of time is consumed between the round trips and the manual inventory in this process. As a result of the special storage operation environment in a map warehouse, it is necessary to install suitable automatic map counting equipment on the narrow tunnel vehicles and design an algorithm that is compatible with this equipment. In this way, the intelligence level of map warehouses can be improved effectively, and this will be of great significance in realizing the fine management and querying of map information.

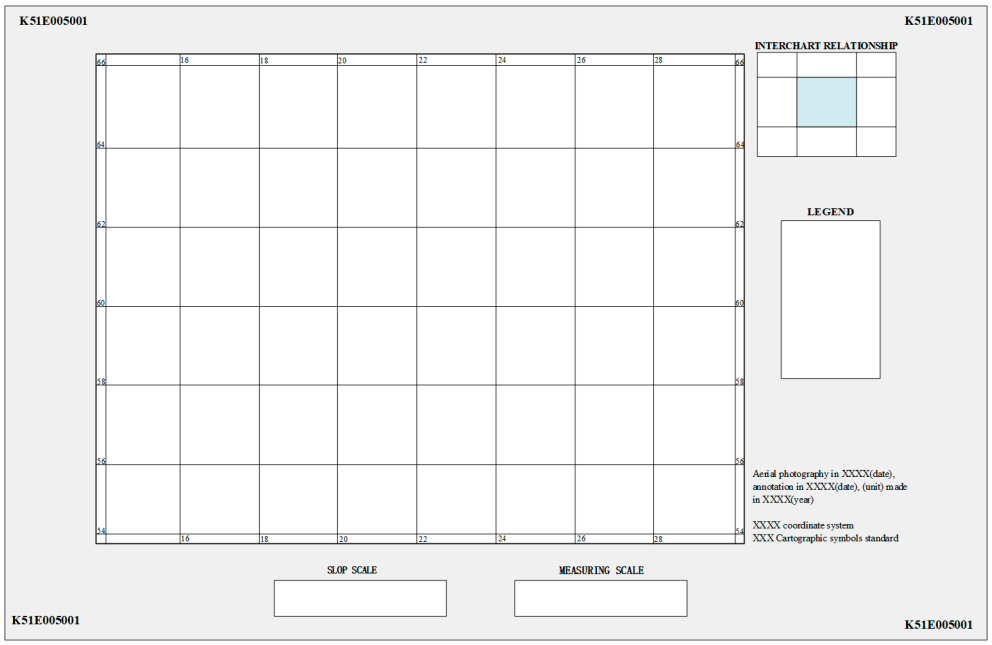

Figure 1. Style of topographic maps.

2. Related Work

In this section, some relevant works are described briefly. There is no shaped paper map inventory equipment currently available in warehouses. Considering that paper is the main medium of topographic maps, the accurate counting of them is one of the key requirements in the packaging and printing industry [5]; so, special equipment for checking topographic maps has been designed with reference to the devices related to paper counting. Today, there are two types of mature paper counting machines on the market. One is the paper counting device represented by cash counters and paper counting machines, which can quickly complete the counting of large amounts of paper by using mechanical paddles, photoelectric machines, and other instruments. The other type comprises paper transportation machines, which operate through the separation and transport of paper for counting, such as paper conveyors, paper separators, etc. However, for paper map counting in complex scenarios such as map warehouses, a non-contact photoelectric machine from the former category cannot separate the paper for the rapid recording of the internal information and lacks the ability to identify and sort. The machines in the latter category are huge in size and unstable in the separation and transportation of large-scale paper. Therefore, neither type can directly meet the daily inventory, identification, and sorting needs of map warehouses. Currently, most of the domestic and foreign research on counting machines is in the form of patents and is mainly aimed at small and thin objects with paper-

like shapes (such as banknotes, checks, etc.) [6,7]; so, it does not apply to the counting of large-scale maps and the recognition of internal information.

The traditional counting methods for paper or paper-like shaped objects (such as banknotes, hardboards, etc.) include manual, physical, and mechanical measurement [8], but these methods generally suffer from low efficiency and high cost. Therefore, with the continuous development of machine vision and image processing technology in recent years, the research on paper counting algorithms both at home and abroad has produced a new method, namely a non-contact paper counting method based on machine vision [9], which can be divided into two categories according to the placement of the papers. The first is laminated placement. The side images of the laminated papers are obtained by directly shooting the whole side or moving to shoot multiple images with a certain overlap for image stitching, and then, the shading texture features on the side are processed to achieve the task of counting. Han et al. [10] counted papers by calculating the difference in the grayscale values in the horizontal directions of the images and proposed an algorithm for counting compensation combined with the overall information to eliminate the false detection position and to recover the missed detection position. Wang et al. [11] proposed a laminated paper counting algorithm based on compressive sensing and the Hough transform, addressing the shortcomings of conventional algorithms, such as their high dependence on the quality of laminated paper and noise sensitivity. Huang et al. [12] combined a wavelet transform and the relevant morphological algorithms for processing the visual information to acquire the number of connected domains and calculated the number of corrugated papers. Luan [13] spliced the images of laminated papers with an overlap of no less than 50% and then used three methods of judging the black stripes, morphological filtering and Gabor filtering, to achieve counting after comparison and analysis. In this kind of method, if the interval between the paper and the gap is too small, this will have a great impact on the imaging effect, and the requirements for the edges of papers are relatively high, such as the requirement for regular edges with no more than 1 mm of margins misplaced, etc. The second method is arrangement placement. The whole image is obtained by photographing the paper horizontally, arranged at a certain distance; then, the edge features of the papers in the image are processed to achieve counting. Junya Sato et al. [14] extracted the edges of aligned paper images by finding the optimal threshold of the Canny edge detector with the genetic algorithm and inhibited the false detection to count six different types of facial oil blotting papers. Chen et al. [15] established template Gabor filters with different frequencies based on the edge features for pretreatment and then counted the number of papers using the grayscale value difference algorithm according to the grayscale pixel curves. Ye et al. [16] divided the edge images of arrayed answer sheets into multi-row regions of interest and then carried out forward difference on the projected curve after the accumulation of gray information; finally, they counted the number of answer sheets according to the probability of peaks, valleys, and positive and negative zero-crossing values. Such a method requires arranging the papers manually and ensuring that there is a proper distance between the sheets.

In summary, the current domestic and foreign research on counting topographic maps has mainly the following problems. This paper focuses on solving these problems:

1. Map inventory machines: at present, the counting of topographic maps is largely performed manually, which has disadvantages such as low efficiency and the wasting of resources. Because of the complex and narrow operating environments in map warehouses, none of the mature paper counting machines on the market can directly meet their daily inventory, identification, and sorting needs. Thus, there is a lack of machines suitable for the rapid inventory of topographic maps in map warehouse environments.
2. Paper counting method: the inventory and sorting of topographic maps requires not only the counting of the quantity but also the obtaining of the internal information of the maps simultaneously, e.g., classifying the map belonging to the sheet designation or scale through the identification of the sheet designation. However, the current paper counting method based on machine vision cannot obtain the internal information

of the counted objects; in addition, it demands high intervals of paper, which is unfavorable for the subsequent identification and sorting.

To address the above problems, in this study we use a map counting device transformed from a paddle-type paper counting machine for preliminary counting, and simultaneously obtain the video data of the paddle counting. Based on this, an automatic counting algorithm for topographic maps for map management is proposed. First, the map counting machine is used to paddle and turn the simulated maps, and the process is captured in multiple videos according to different paddle speeds and recording frequencies. Second, the periodic variation features of all the targets and each individual target in the videos are analyzed. Taking the mechanical wheel as the object, the accelerated deformable parts model algorithm is used for object detection, and the image gray variance function is selected to quantify the changes in image quality. Finally, the interval windows of the image clarity quantization curves are built with the object detection results, and a fusion window counting model is established by combining the two methods of the adaptive window and the fixed window. The preliminary counting results of the device are checked through the statistics of the periodic peaks in the model, and the clear frame of each key region is determined accordingly. This algorithm facilitates the subsequent identification and classification of the sheet designations and lays a solid foundation for the intelligent management of topographic maps.

3. Materials and Methods

3.1. Data Acquisition and Analysis

3.1.1. Map Inventory Machine and Data

Based on a comprehensive consideration of the above analysis, the map counting equipment used in this paper (as shown in Figure 2) was based on modifying and upgrading a paddle-type paper counting machine to complete the checking and automatic storage of map data. The following four aspects were improved. (1) At present, the principle of automatic paper counting equipment involves distinguishing the number of papers by vacuum adsorption, which requires high flatness and hardness of the paper. However, due to the inevitable folding, extrusion, and other operations during the transportation, storage, and unpacking of maps, their flatness may not meet the relevant requirements. Moreover, map paper is often soft, and there are uncertain factors in relation to counting directly with the existing counting equipment. Therefore, the cutter head and press plate of the inventory machine need to be redesigned and selected specifically for the characteristics of map paper. (2) A paddle-type paper counting machine simply separates the paper with very limited gaps and distances. The existing camera equipment can hardly obtain the map number information located at the corners of the map from its gaps, and it is difficult to complete the verification of the map counting process. Therefore, in the reconstruction of the existing paddle counting machine, it is necessary to focus on the redesign of its separation equipment so that the sheet designations can be effectively identified and read. (3) In the relevant parts of the quick inventory machine, the camera device is reasonably installed in conjunction with the transformation of the map separation equipment. This device can be used to obtain map image data during map paddling and paging and can also provide a verification method for the inventory to ensure the reliability and accuracy of the map inventory. (4) The size of the existing equipment platform is still too large, and the paper needs to be flattened before it can be counted quickly. These requirements are too complex for tunnel vehicles at high altitudes. Therefore, the equipment platform needs to be reformed to reduce the size and to improve the paper counting model so that it can effectively count and verify under relatively simple conditions.

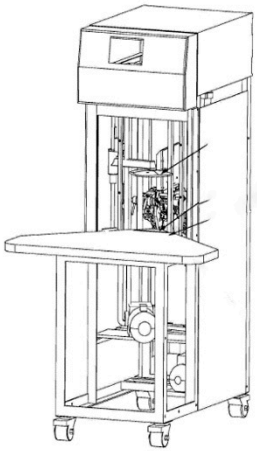

Figure 2. Map inventory machine.

The map inventory machine is used to toggle the simulated maps, take videos with different paddling speeds and recording frequencies, and intercept the videos frame by frame, as shown in Figure 3. Among them, the topographic maps in the video were printed in the same way as the map, in accordance with the standard of "Subdivision and numbering for the national primary scale topographic maps (GB/T 13989-2012)" [17] published and implemented in China in 2012, using map printing paper with low deformation to ensure map accuracy, wear resistance, and folding resistance. Moreover, considering that the contents of the topographic map are confidential, the specific map contents were discarded, and the map sheet designation was kept printed on the paper, thus making the simulated topographic maps for our research.

Figure 3. A frame in the video.

3.1.2. Feature Analysis of Map Counting Process in Video Data

Video data may be affected by various environments during the acquisition process, resulting in interference factors such as occlusion, noise, and motion blur in the targets of the videos. By fully mining all of the features of the video data, it can first be found that the mechanical wheel in the videos is clear throughout and that there is good correlation between its periodic appearance and the page turning of the topographic maps. Secondly, with each page turning and updating of the maps, the image qualities in the areas of the sheet designation have the obvious periodic variations of blur–clear–blur. Next, the mechanical wheel targets in the video are analyzed by features, and they can be specifically divided into various parts, as shown in Figure 4, where the mechanical wheel part is composed of R_1, R_2, R_3, and R_4, and A_1 is a non-mechanical wheel part. The whole target can be regarded as being composed of several objects [18], and the relative relationship among them is fixed. In each map counting process, the mechanical wheel also exists in three states, as shown in Figure 5. The first state (Figure 5a) comprises the mechanical wheel and the other parts of the machine appearing on the video screen at the same time before the equipment starts counting; the area occupied by the wheel is relatively small,

and the other parts of the machine may affect the recognition of the wheels. In the second state (Figure 5b), before and after each map page-turning action occurs during the counting process, the wheel mainly appears above the video. The third state (Figure 5c) is in the proximity of the map page-turning process; that is, when this angle appears, the map will flip at the next moment. At this time, the wheel mainly appears in the upper right corner of the video and accounts for a larger proportion of the area on the screen. According to the properties of the mechanical wheel object with multiple parts and components, taking it as the object of the target detection can achieve better detection results.

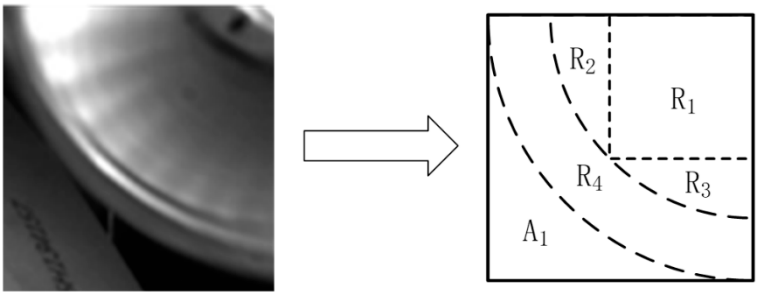

Figure 4. Composition of the mechanical wheel target component.

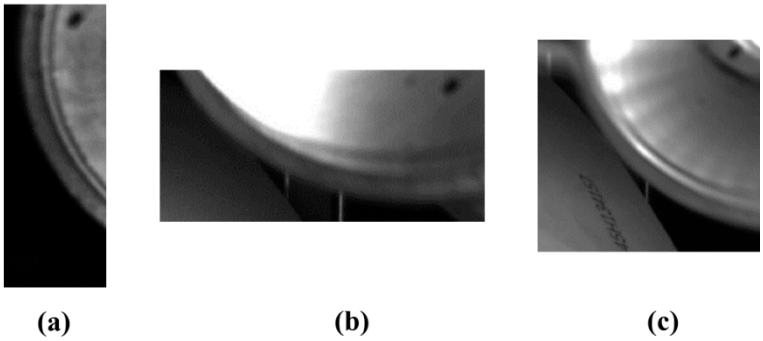

Figure 5. Three states of mechanical wheel targets. (**a**) The first state comprises the mechanical wheel and the other parts of the machine appearing on the video screen at the same time before the equipment starts counting. (**b**) The second state, before and after each map page-turning action occurs during the counting process, the wheel mainly appears above the video. (**c**) The third state is in the proximity of the map page-turning process; that is, when this angle appears, the map will flip at the next moment.

3.2. The Proposed Fusion Window Counting Algorithm

According to the periodicity of the map counting process for the video data, first the mechanical wheels in the videos are taken as the target, and they are detected by the deformable parts model (DPM) algorithm. Then, the time-consuming defect of the algorithm is improved by combining it with fast feature pyramids (FFP); this is called the FFP–DPM algorithm. For all of the video data in the research, the region of interest (ROI) is reduced through a certain number of image detection results at the beginning of the video to reduce the computation for the subsequent detection in the video, thereby shortening the operation time of the whole video; this is called the FFP–DPM–ROI algorithm. Secondly, the times of the mechanical wheel toggles in the whole video can be obtained according to the periodicity of the detection results, but considering the idling of the mechanical wheel and the need to select the clear frames of the sheet designations while counting, the change in image quality is quantified using the value of the clarity evaluation function. Finally, the object detection results are organically combined with the image quality evaluation curve to establish the fusion window counting model and realize the inventory research regarding topographic maps based on machine vision. The following describes the detailed implementation process, as shown in Figure 6.

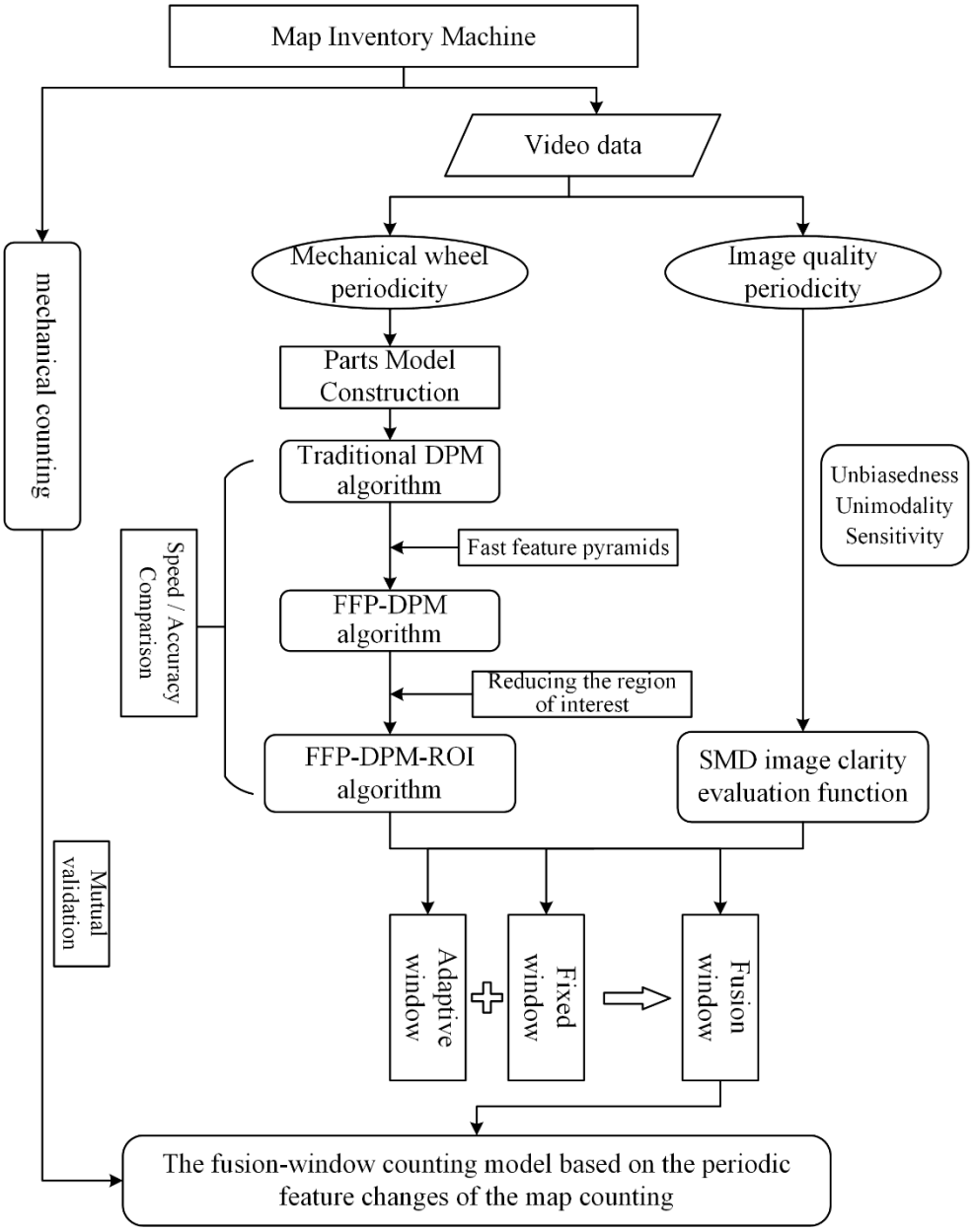

Figure 6. The flowchart of the work.

3.2.1. Deformable Parts Model Based on Fast Feature Pyramids

The DPM algorithm [19–22] can better detect different types of targets in complex natural environments by integrating the overall information of the target to be detected and the relationships between its components. The algorithm adopts a multi-component strategy to construct a mixed multi-angle deformable parts model for the existence of multiple views of the target; for the deformation problem of the target itself, it adopts a strategy based on the graph structure, which improves the adaptability of the model to the various deformation problems of the target. Moreover, ideal results have been achieved in the identification of mixed multi-angle targets, such as people and cars [23,24]; so, it is feasible to use this algorithm to detect mechanical wheel targets.

However, the traditional DPM algorithm suffers from the problems of being computationally intensive and time-consuming [25,26], and therefore, it cannot satisfy the requirements for the fast checking of topographic maps. According to the detection process of the algorithm, the detection speed mainly depends on two aspects: (1)' the feature extraction of feature pyramids and (2) the sliding window method for target localization.

The acceleration of the feature pyramid part can not only shorten the construction time of the pyramid, it can also largely reduce the computation of the sliding convolution operation. Therefore, in this paper, we draw on the idea of fast feature pyramids for optimization and adopt the FFP–DPM algorithm to detect mechanical wheels. This process mainly includes the following key stages.

1. Model training to obtain filters

The training phase is a crucial process that enables the model to play an ideal role in the detection phase. The essence of model training in the DPM algorithm is the obtaining of filters, which include three parts: a coarse-root filter describing the global contour of the target; several part filters describing the detailed features of the target; and a spatial model describing the deviation loss of each part filter relative to the root filter. The filter is a rectangular template defined by a d-dimensional weight vector, and a model for an object with n parts can be defined as an $(n + 2)$ tuple:

$$\mathbf{F} = (\mathbf{F}_0, \mathbf{P}_1, \cdots, \mathbf{P}_\mathrm{n}, b) \tag{1}$$

where $\mathbf{F}_0$ is a root filter, $\mathbf{P}_\mathrm{i}$ is a model for the ith part, and b is a real-valued bias term. Each part model is defined by

$$\mathbf{P}_\mathrm{i} = (\mathbf{F}_\mathrm{i}, v_i, d_i) \tag{2}$$

where $\mathbf{F}_\mathrm{i}$ is a filter for the ith part, v_i is the coordinate of part i relative to the root position, and d_i defines a deformation cost for each possible placement of the part relative to the anchor position.

2. Building the fast feature pyramids

As the size and position of the mechanical wheels in the image are not fixed, the traditional DPM algorithm adopts the idea of multi-scale to construct the feature pyramid, in which the features are based on the enhanced histogram of the oriented gradient proposed by Felzenszwalb (FHOG). The specific steps to build the pyramid are as follows: starting from the original image, we down-sample in each octave from the previous step $2^{(1/\mathrm{interval})}$ times to obtain the image pyramid, with 10 scales in each octave in the model (i.e., interval = 10), and then, the feature pyramid is formed by combining the characteristic diagrams obtained by computing the features on each level of the image. The idea of the fast feature pyramid [27] approximates the features of the current scale image from the features of the adjacent scale image by extrapolation, which can improve the detection speed with almost no loss of accuracy. The specific process is as follows (as shown in Figure 7).

The general feature pyramid is constructed by scaling the image I to scale s and then calculating the FHOG features of the image I_s; that is, $C_s = \Omega(\mathrm{S}(I, s))$, where $C = \Omega(I)$ denotes the feature extraction function, and S denotes the sampling function. The fast feature pyramid selects a benchmark scale ($s \in \{1, 0.5, 0.25, \cdots\}$) in each octave and calculates features by using the general feature pyramid method $C_s = \Omega(\mathrm{S}(I, s)), s \in \{1, 0.5, 0.25, \cdots\}$ for feature extraction at the key scales. Through the features of the key scales, the features at the adjacent scales are estimated using the approximate estimation method $C_s \approx \mathrm{S}(C_s, s/s')(s/s')^{-\lambda_\Omega}$, where s' is the nearest scale to s. The constant coefficient λ_Ω needs to be estimated by the training samples in advance based on specific features, and the coefficient λ_Ω of the FHOG features is -0.091.

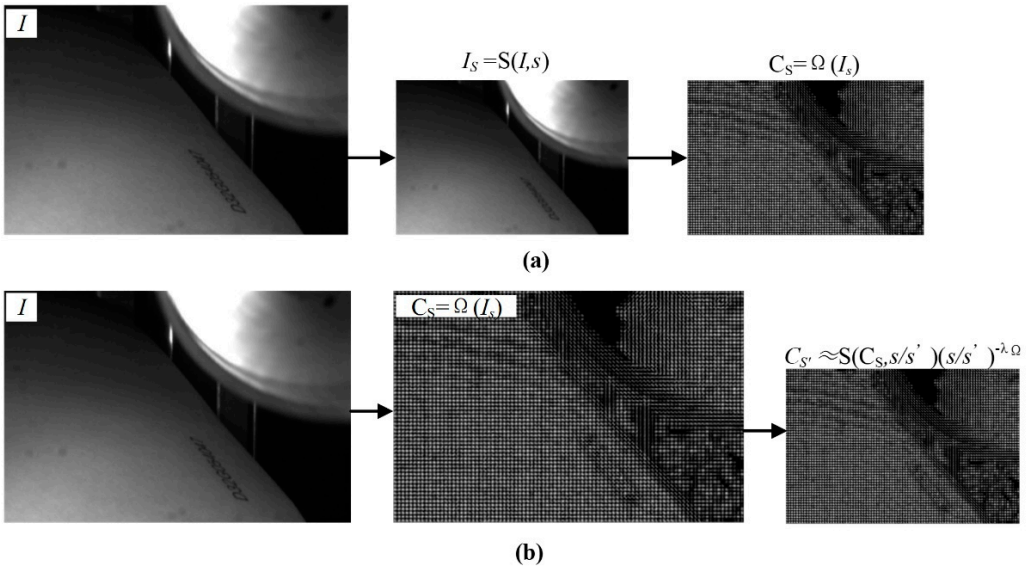

Figure 7. Schematic diagram of the fast feature pyramid: (**a**) normal computation method of FHOG features; (**b**) approximate computation method of FHOG features.

3. Model matching

The DPM algorithm adopts the sliding window method for model matching. At first, the overall convolution operation of the pyramid is performed using the root filter, and the response value of the root filter position of the level is obtained as:

$$R_{0,l}(x_0, y_0) = \mathbf{F}_0 \cdot \boldsymbol{\phi}(H, (x, y, l)) \tag{3}$$

where (x_0, y_0) is the coordinate of the root filter in its level, H is the feature pyramid, $p(x, y, l)$ is the position of level l in the specified pyramid, and $\boldsymbol{\phi}(x)$ is the feature vector at x in the image.

Similarly, the response value of the ith part filter at the lth level of the feature pyramid can be obtained, but it should be considered that it has the coordinate offset dx, dy relative to the root filter:

$$R_{i,j}(x_0 + dx, y_0 + dy) = \mathbf{F}_\mathrm{i} \cdot \boldsymbol{\phi}(H, (x_0 + dx, y_0 + dy, l)) \tag{4}$$

After calculating these filter response values, they are converted so that they have spatial uncertainty. This transformation can extend the filter's high scores to adjacent positions while also considering the deformation costs:

$$D_{i,j}(x, y) = \max_{dx,dy}(R_{i,j}(x + dx, y + dy) - d_i \cdot \boldsymbol{\phi}_d(dx, dy)) \tag{5}$$

where the $D_{i,j}(x, y)$ value means the maximum contribution to the root position score when the anchor of the ith part is placed at the location (x, y) in level l.

Then, the synthesis score for each level of the root position can be obtained as the root filter response value for that level plus the response value of the part filter after transformation and subsampling, minus the deformation cost of this position relative to the root filter, plus the deviation b:

$$Score(x_0, y_0, l_0) = R_{0,l_0}(x_0, y_0) + \sum_{i=1}^{n} D_{i,l_0-\lambda}(2(x_0, y_0) + v_i) + b \tag{6}$$

In the process of calculating $D_{i,l}$, we can also calculate the optimal position of the part—that is, the function of the anchor coordinates. After finding a root position (x_0, y_0, l_0) with a high score, the corresponding optimal position of the part can be found as follows:

$$P_{i,j}(x,y) = arg \max_{dx,dy}\left(R_{i,j}(x+dx, y+dy) - d_i \cdot \boldsymbol{\Phi}_d(dx,dy)\right) \tag{7}$$

Finally, the score of each window is calculated; then, the threshold is determined based on the training data. When the score is greater than the threshold, the window is considered to contain the mechanical wheel, and the position of the window in the image is marked with a rectangular box. After model matching, each frame of the video will produce multiple detection boxes with the mechanical wheel as the target, and one detection box corresponds to one score. The scores are ranked, and the best detection box is selected by non-maximum suppression (NMS) to obtain the final detection results.

3.2.2. Selecting the Region of Interest

According to the existing state of the mechanical wheels in the video, they only exist in the part area of each frame. However, the DPM algorithm builds a feature pyramid for the whole image and finds the target candidate regions by the sliding window method. If the picture size is large, this will lead to more candidate regions in an image and increase the algorithmic operand. Therefore, the algorithm reduces the operation time of the whole video by learning to diminish the ROI; that is, when detecting each video frame by frame, the maximum bounding rectangle of the target is determined according to the detection box of a certain number of frames after the beginning of the video, and the subsequent frames use the determined maximum bounding rectangle as the region of interest to realize the object detection.

Suppose the coordinates of the ith frame detection box are (x_1^i, y_1^i), (x_1^i, y_2^i), (x_2^i, y_1^i), and (x_2^i, y_2^i), where $x_1^i < x_2^i$, $y_1^i < y_2^i$. The coordinates of the maximum bounding rectangle determined by the image detection box of the first n frames are (X_1^n, Y_1^n), (X_1^n, Y_2^n), (X_2^n, Y_1^n), and (X_2^n, Y_2^n), where:

$$\begin{cases} X_1^n = min\{x_1^1, x_1^2, \cdots, x_1^n\} \\ X_2^n = max\{x_2^1, x_2^2, \cdots, x_2^n\} \\ Y_1^n = min\{y_2^1, y_2^2, \cdots, y_2^n\} \\ Y_2^n = max\{y_2^1, y_2^2, \cdots, y_2^n\} \end{cases} \tag{8}$$

The absolute values of the coordinate difference of the maximum bounding rectangle determined by the image detection of the first n frames and the first $n-1$ frames are $\left|X_1^n - X_1^{n-1}\right|$, $\left|X_2^n - X_2^{n-1}\right|$, $\left|Y_1^n - Y_1^{n-1}\right|$, and $\left|Y_2^n - Y_2^{n-1}\right|$, respectively.

Suppose that the threshold of the allowable coordinate difference is δ. If the coordinate difference of consecutive m frames is within the threshold range, the coordinates of the maximum bounding rectangle R can be determined as (X_1^N, Y_1^N), (X_1^N, Y_2^N), (X_2^N, Y_1^N), and (X_2^N, Y_2^N), through N frames, where $X_1^i < X_2^i$, $Y_1^i < Y_2^i$, and the relationship between n, m, and M in the first M frames can be expressed as $\{1, 2, 3, \cdots, n, n+1, \cdots, n+m, M\}$. Empirically, δ is taken as 20 pixels and m as 10. After R is determined, the subsequent images in the same video are clipped using the R for the object detection.

3.2.3. Fusion Window Counting Model

The turning frequency of the mechanical wheel can be judged from the object detection results. However, the toggle frequency does not accurately represent the number of maps due to factors such as the video shooting time, the angle, and the adhesion between papers. The object detection progress of each frame is supplemented by image quality assessment (IQA) [28,29] to improve the accuracy of the counting. In the no-reference quality assessment algorithms, image clarity [30] is one of the important indicators used to measure the quality of images and can better reflect the subjective feeling of humans.

We selected the sum of the modulus of gray difference (SMD) function, which is most applicable to the images in this study [31].

According to the relationship between the appearance of mechanical wheels and the clarity change in the images in the video, the clearer frames corresponding to sheet designations are mainly concentrated in the interval without mechanical wheels. At the same time, because the pictures in this interval are visually free of mechanical wheel occlusion, the influence of the mechanical wheels on the clarity calculation of the images can also be reduced. Therefore, we propose a fusion window counting model, combining the FFP–DPM–ROI algorithm with IQA. Firstly, the mechanical wheels are quickly detected according to the improved deformable parts model. Secondly, the SMD function is used as the clarity assessment algorithm to quantify the image quality, and the two are combined to build the counting model. The specific combination process is as follows:

Assume that the detection result and clarity value of the ith frame image are D_i and T_i, respectively. If the mechanical wheel is detected, assign D_i to 1, otherwise assign 0; then, each video will obtain the periodic pulse curve of the detection result D_i about image i. Through the above analysis, the clarity curves need to satisfy the following.

$$T_i = \begin{cases} T_i, D_i = 0 \\ 0, D_i = 1 \end{cases} \tag{9}$$

Let the length of the jth interval without the mechanical wheel be I_j, and the extremum of clarity in the interval be T_{jmax}. If the interval satisfies $\forall i \in [l, k], D_i \equiv 0$ and $D_{k+1} = 1$, then

$$\begin{cases} I_j = k - 1 \\ T_{jmax} = max\{T_i, T_{i+1}, \cdots, T_k\} \end{cases} \tag{10}$$

For each video, there is a theoretical interval without mechanical wheels in the video for each map counting process, and there is an extremum in each interval. However, the actual videos cannot accurately correspond one by one; so, the following different methods are used to define the windows in which the extremums are located in order to realize the counting. Suppose that the size of the window in which the mth extremum is located is W_m.

1. Adaptive window method: take each I_j directly as the size of the window, i.e., $W_{m1} = I_j (m = j)$; the number of extremums obtained is num_a.
2. Fixed window method: take the fixed value as the size of the window, i.e., $W_{m2} = Mo(I_j)$, where Mo is the mode function; generally, $m \neq j$ and the number of extremums obtained is num_f.
3. Fusion window method: The toggle speed of the mechanical wheel is approximately uniform; there are several intervals of different lengths, but the lengths are stable within the threshold range of a certain fixed value W_{m2}; so, the counting results of the above two methods can reflect each other. The fusion window counting model fuses the two, and if we set the final count value of the topographic maps in the video to Num, then

$$Num = \begin{cases} Num + 1, num_a = num_f \\ Num, num_a \neq num_f \end{cases} \tag{11}$$

That is, if $num_a = num_f$, the map counting result is increased by one, and this method is used to count each interval in the video to obtain the final number of maps.

4. Results

4.1. Comparison and Analysis of Mechanical Wheel Detection Results

For the experimentation, we adopted the integration of the DPM algorithm and the FFP algorithm; we used Girshick's voc-release5 and Piotr's Computer Vision MATLAB Toolbox (PMT) version 3.50 written by Piotr Dollar of UCSD in the United States. The platform adopted a combination of Visual Studio 2015 developed by Microsoft in the

United States and MATLAB R2016 developed by MathWorks in the United States, and the running environment was a PC with a Windows 10 operating system (CPU: I5-10400F 2.90GHz; RAM:16GB). In this experiment, a certain number of images, all of which were 720 pixels × 480 pixels, were selected from the frame by frame images of many videos to form the training set, test set, and validation set containing positive and negative samples of the three components of the mechanical wheels. The deformable parts model was built by training these samples and was then used to detect the videos frame by frame. The detection results of one frame are shown in Figure 8, and the precision/recall curves for the DPM algorithm and the FFP–DPM algorithm were compared and are shown in Figure 9.

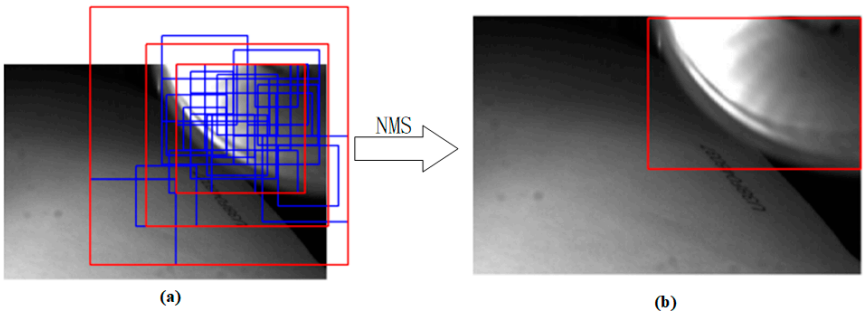

Figure 8. Target detection results: (**a**) initial detection results; (**b**) final detection results. The red rectangular boxes indicate the possible position of the detected target; there is more than one target candidate box in the figure, and the boxes will overlap with each other. The blue rectangular boxes represent the positioning results of the part model inside each red box.

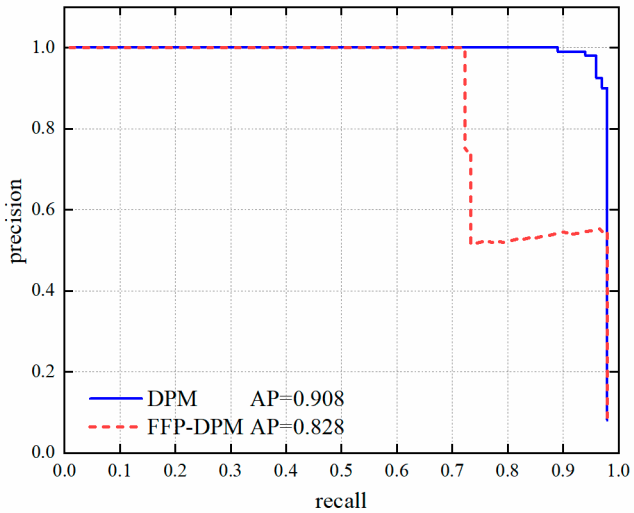

Figure 9. Precision and recall curve of DPM algorithm and FFP-DPM algorithm.

Figure 8a shows the initial detection results. The red rectangular boxes indicate the possible position of the detected target; there is more than one target candidate box in the figure, and the boxes will overlap with each other. The blue rectangular boxes represent the positioning results of the part model inside each red box. Figure 8b shows the best target bounding box obtained from Figure 7a after non-maximum suppression to eliminate the redundant candidate boxes. It can be seen that the algorithm can accurately detect the specific position of the mechanical wheels and mark it completely. The PR curve is one of the important indexes used to evaluate the performance of the trained model, and P and R in the curve indicate precision and recall. The former represents the ability of the model to detect irrelevant information, and the latter represents the ability of the model to detect relevant information; so, the higher the recall, the lower the accuracy. The AP (average precision) value is the area enclosed by the PR curve, and the larger the value, the

stronger the detection ability of the model. As can be seen in Figure 9, both the DPM and the FFP–DPM algorithm have a stronger detection capability, and the detection ability of the FFP–DPM algorithm is only slightly lower than that of the DPM algorithm.

In practical applications, the detection speed is as important as the detection accuracy. In this study, we detected many videos according to the above detection process and used the corresponding evaluation indexes to compare and analyze the detection effects of three methods: the traditional DPM algorithm, the FFP–DPM algorithm, and the FFP–DPM–ROI algorithm. The following are the detection results for the various videos, as shown in Tables 1 and 2. The detection speed adopts the average time-consumed index, and the detection accuracy adopts three indexes: accuracy (ACC), the true positive rate (TPR), and the false negative rate (FNR). The accuracy refers to the proportion of the number of images with correct detection results in the total number of images. The true positive rate refers to the proportion of the number of images with correctly detected mechanical wheels in the number of pictures with actual mechanical wheels, and the false negative rate refers to the proportion of the number of images with the incorrect detection of no mechanical wheels in the number of images with actual mechanical wheels. Finally, the time-consuming values and the accuracy of multiple video detection results were averaged, and the overall detection efficiency of the three algorithms was compared and analyzed, as shown in Figure 10.

Table 1. Comparison of detection times of different algorithms for each video (s).

Video	Number of Images	Traditional DPM		FFP–DPM		FFP–DPM–ROI	
		Total Time	Average Time	Total Time	Average Time	Total Time	Average Time
1	1812	30,360.461	16.755	8963.497	4.947	8877.310	4.899
2	6120	102,877.465	16.810	30,630.299	5.005	29,969.693	4.897
3	6389	107,624.298	16.845	31,945.032	5.000	26,866.572	4.205
4	2146	36,032.633	16.798	10,779.174	5.025	9364.335	4.366
5	2145	36,023.563	16.794	10,793.490	5.032	9327.416	4.348
average time			16.801		5.002		4.543

Table 2. Comparison of detection accuracies of different algorithms for each video.

Video	Number of Images	Traditional DPM			FFP–DPM			FFP–DPM–ROI		
		Acc	TPR	FNR	Acc	TPR	FNR	Acc	TPR	FNR
1	1812	0.995	0.990	0.010	0.986	0.963	0.037	0.984	0.965	0.035
2	6120	0.992	0.983	0.017	0.976	0.936	0.064	0.974	0.937	0.063
3	6389	0.968	0.952	0.048	0.937	0.857	0.143	0.935	0.852	0.148
4	2146	0.971	1	0	0.968	0.955	0.045	0.967	0.957	0.043
5	2145	1	1	0	0.970	0.956	0.044	0.968	0.958	0.042
average value		0.979	0.985	0.015	0.967	0.934	0.066	0.966	0.934	0.066

Table 1 shows that the traditional DPM algorithm is the slowest, with an average time of 16.801 s. After combining the fast feature pyramids to accelerate the traditional DPM algorithm, the FFP–DPM algorithm is significantly faster, with an average time of 5.002 s. The FFP–DPM–ROI algorithm, obtained by learning to reduce the region of interest of the targets to accelerate the whole video, takes an average time of 4.545 s. It has a weak acceleration effect on the FFP–DPM algorithm, but it accelerates by nearly 3.7 times compared with the traditional DPM algorithm. Table 2 shows that the accuracy and the true positive rate of the traditional DPM algorithm are the highest, and the false negative rate is the lowest. Compared with this algorithm, the FFP–DPM algorithm and the FFP–DPM–ROI algorithm decrease the accuracy and the true positive rate by about 0.01 and 0.05, respectively, and increase the false negative rate by about 0.05. The comprehensive analysis of the figures and tables shows that the accuracies of the FFP–DPM and the FFP–DPM–ROI

algorithm are only slightly reduced on the basis of the obvious acceleration effect, which meets the needs of the practical application.

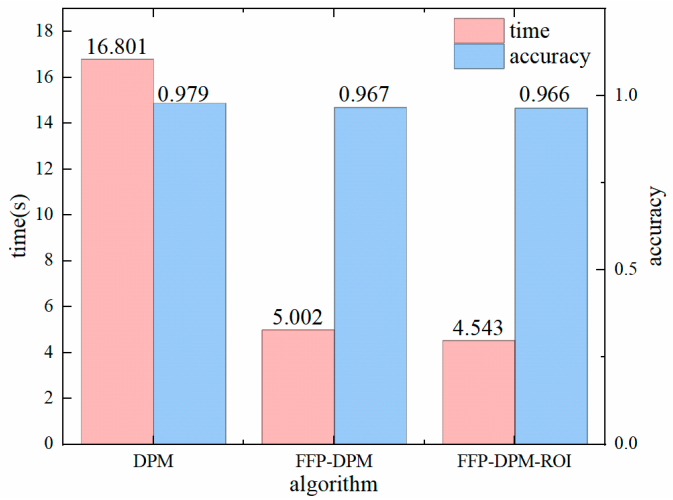

Figure 10. The average time and average accuracy of the three algorithms.

Through analyzing the accuracy of the above algorithms, we found a few false detections in the experiment results; if these values are directly used in the subsequent counting process, they will interfere with the results. According to the continuity of the process from appearance to disappearance or from disappearance to appearance of the mechanical wheels, we smoothed the anomaly detection using the detection results before and after the false detections to improve the counting accuracy, as shown in Figure 11.

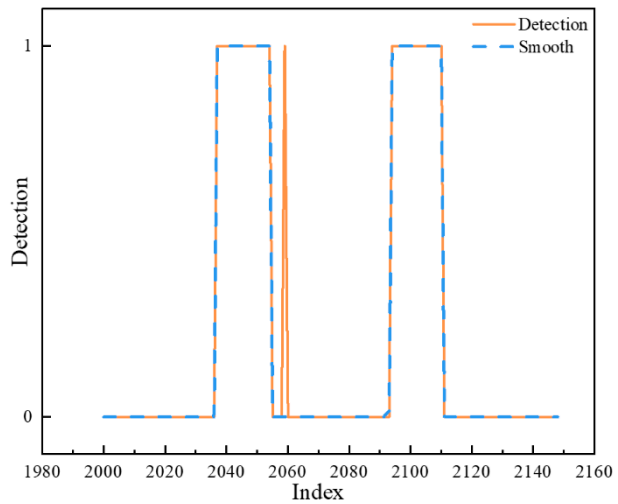

Figure 11. Abnormal detection results and smoothing.

4.2. Analysis of Paper Map Inventory Results Based on Fusion Window Counting Model

Combining the target detection results with the image quality assessment index, the counting experiments for multiple videos were carried out according to three methods: the adaptive window, the fixed window, and the fusion window. Because the FFP–DPM–ROI algorithm improved the operation speed with less loss of detection accuracy in the experimental results for object detection, we combined it with the SMD clarity assessment function (hereinafter called the FFP–DPM–ROI–SMD algorithm) for the inventory research of topographic maps. The fusion window method in this algorithm is the fusion window counting model proposed in this paper. Taking video 3 as an example, the counting process of the FFP–DPM–ROI–SMD algorithm is shown in Figure 12. The counting results of this

algorithm and the DPM algorithm combined with the SMD clarity assessment function (hereinafter referred to as the DPM–SMD algorithm) were compared and analyzed, as shown in Figure 13, where the left figure shows the counting results of the DPM–SMD algorithm and the right figure shows the counting results of the FFP–DPM–ROI–SMD algorithm. Table 3 shows the counting accuracies of the different algorithms.

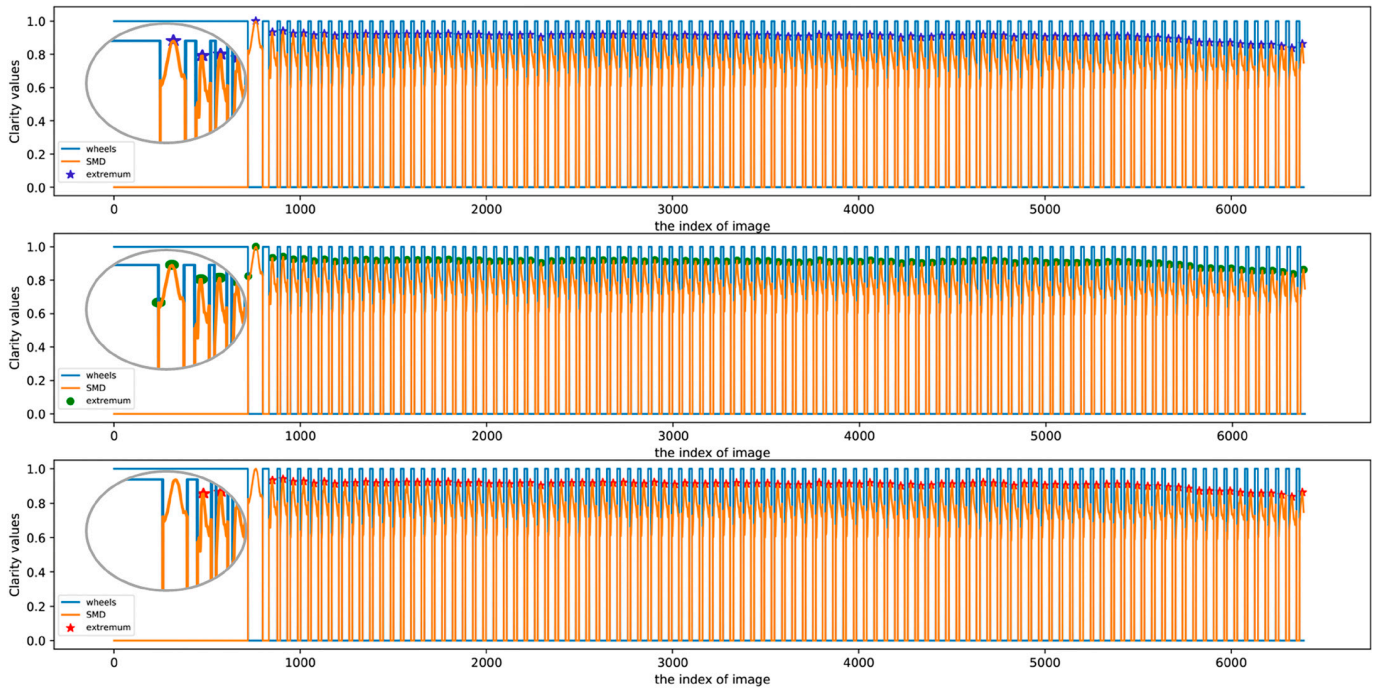

Figure 12. FFP–DPM–ROI–SMD algorithm inventory process.

Table 3. Comparison of counting accuracies of different algorithms.

Video Number	DPM–SMD			FFP–DPM–ROI–SMD		
	Adaptive Window	Fixed Window	Fusion Window	Adaptive Window	Fixed Window	Fusion Window
1	95%	83%	99%	95%	85%	98%
2	99%	95%	92%	99%	95%	93%
3	98%	96%	99%	98%	97%	99%
4	91%	88%	95%	91%	82%	92%
5	91%	88%	95%	91%	82%	92%
Average accuracy			96%			95%

As can be seen in Figure 12, for the first interval without mechanical wheels, there is one extremum in the adaptive window method and there are two extremums in the fixed window method, and according to the idea of a fusion window, the fusion window counting model does not add one to the count value for this interval. The observation of the videos shows that the topographic maps are not updated in the first cycle when the mechanical wheel appears; so, the fusion window method is more in line with the counting of the topographic maps in the actual videos. The comparison and analysis of the two algorithms in Figure 13 and Table 3 reveal that both of the adaptive window methods have the same counting results and higher accuracy, and the count results of the two fixed window methods are greatly affected by the accuracy of the object detection algorithm and have lower count accuracy. The counting results of the fusion window method for all the videos except video 2 are closest to the actual number of maps, and the average counting accuracy of the FFP–DPM–ROI–SMD algorithm is slightly lower than that of the DPM–SMD algorithm. In short, we have shown that the FFP–DPM–ROI algorithm

affects the final counting accuracy while increasing the detection speed, leading to a slight decrease in detection accuracy. It also affects the counting results of the final fusion window method, mainly by changing the counting process of the fixed window method. When we combined these results with the actual data, we found that the counting accuracies of the two algorithms for video 2 were low. This was due to the shooting angle, which caused the sheet designation area to be blocked in the later stages of the video. In addition, the clarity assessment value cannot accurately represent the image quality of the map area, and there are situations such as that of a broken edge, or there is adhesion between the maps, which also provides a reference with respect to the relevant shooting conditions for the subsequent acquisition of the video data.

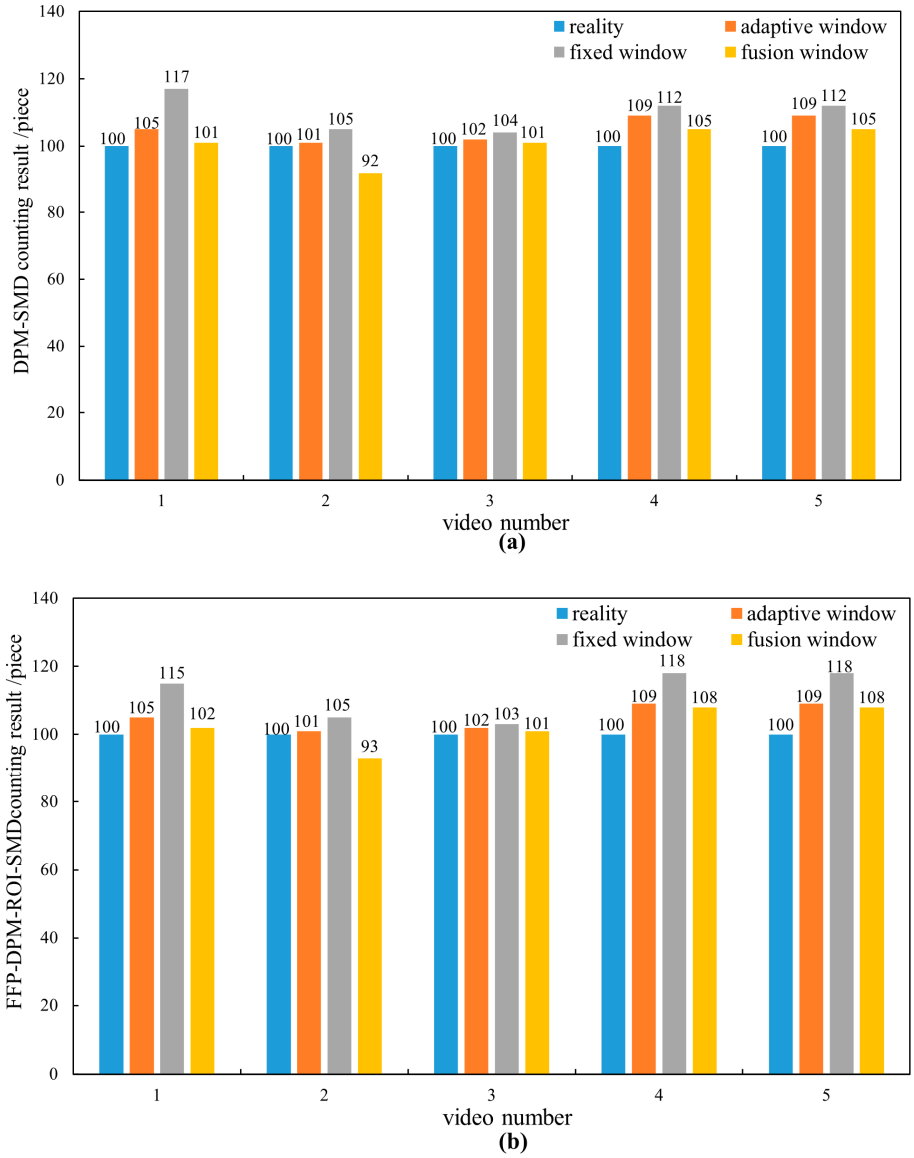

Figure 13. Counting results of different algorithms for topographic maps: (**a**) DPM–SMD method counting result; (**b**) FFP–DPM–ROI–SMD method counting result. Blue represents the actual number of maps in the videos, orange represents the counting result of the adaptive window, gray represents the counting result of the fixed window, and yellow represents the counting result of the fusion window counting algorithm.

In summary, through the experimental results, we can also find that the average time consumption of the FFP–DPM algorithm accelerated nearly 3.7 times compared with the traditional DPM algorithm, and the average time consumption of the FFP–DPM–ROI

algorithm increased by about 0.5 s compared with the FFP–DPM algorithm. The accuracy of the FFP–DPM algorithm and the FFP–DPM–ROI algorithm decreased by only 0.01 compared with the DPM algorithm. Therefore, the accuracies of the FFP–DPM and the FFP–DPM–ROI algorithm are only slightly reduced based on the obvious acceleration effect, which meets the needs of the practical application. On this basis, the average counting accuracy of the FFP–DPM–ROI–SMD counting algorithm can reach about 95%, which is mutually verified with the mechanical counting results of the map inventory machine. In addition, the algorithm can determine clearer keyframes while counting, which is beneficial to the identification and classification of the subsequent sheet designation.

5. Conclusions

Aiming at the current problems of inefficiency and incompatibility with the intelligent development needs of the manual counting of topographic maps, we used a map inventory machine transformed from a paddle-type paper counting device to perform a preliminary paper count and obtain video data for the counting, and we designed a matching automatic counting algorithm for paper topographic maps to realize the acquisition of internal information while counting so as to realize the fine and intelligent management of paper map information. A topographic map counting method based on the periodicity features of map counting in videos was proposed. First, we took the mechanical wheel in the videos as the object, constructed a fast deformable parts model by introducing the idea of fast feature pyramids to detect the target and used the learning process to reduce the region of interest to further accelerate the detection effect. Second, according to the correlation characteristics, we selected the gray variance function as the image quality assessment index. Finally, we established the fusion window counting model by combining the two counting methods of the adaptive window and the fixed window. The experimental results show that compared with the traditional DPM algorithm, the average operation time of the FFP–DPM–ROI target detection algorithm proposed in this paper was reduced by nearly 3.7 times, and the average detection accuracy was only reduced by 0.01. On this basis, the average counting accuracy of the fusion window counting model can reach about 95%, and clearer keyframes can be determined simultaneously, which fully proves the feasibility of this method.

In the actual management application of a map warehouse, it is required that the speed and accuracy of the counting should be as high as possible; therefore, we will consider optimizing the following aspects in future work. (1) We realized the acceleration by improving the feature pyramid of the DPM algorithm. Subsequently, we can improve the processing speed of the object location to achieve two-stage acceleration. (2) In the analysis of the above experiments, it was found that the video shooting environment, the angle and equipment, and the other influencing factors can directly affect the accuracy of the final counting results. Therefore, we can continuously adjust these factors through various experiments to further improve the counting accuracy.

Author Contributions: Conceptualization, W.C. and X.T.; methodology, W.C. and Y.T.; software, Y.T.; validation, Y.T., J.Z., and D.W.; formal analysis, Y.T.; investigation, X.T., C.G., and H.L.; resources, X.T. and W.Y.; data curation, Y.T.; writing—original draft preparation, Y.T.; writing—review and editing, W.C. and X.T.; visualization, Y.T.; supervision, W.C.; project administration, C.G.; funding acquisition, W.C., X.T., and W.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by The Excellent Youth Foundation of Henan Municipal Natural Science Foundation (212300410096), Program of Song Shan Laboratory (Included in the Management of Major Science and Technology Program of Henan Province) under Grant number 221100211000-03, and The National Key R&D Plan of China (2018YFB0505304).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wan, G.; Wu, Y.T. On the mathematical basis of map spatial cognition. *Acta Geod. Cartogr. Sin.* **2021**, *50*, 726–738. [CrossRef]
2. Pezeshk, A.; Tutwiler, R.L. Automatic Feature Extraction and Text Recognition from Scanned Topographic Maps. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 5047–5063. [CrossRef]
3. Liu, A.S.; Zhang, W.W.; Wang, G. Realization and research on unmarked paper map augmented reality technology. *Eng. Surv. Mapp.* **2017**, *26*, 64–70+80. [CrossRef]
4. Yang, K. The Mechanism of Tracing to Map Source and The Key Technology Research. Master's Thesis, Information Engineering University, Zhengzhou, China, 2013.
5. Babakhanova, K.A.; Varepo, L.G.; Nagornova, I.V.; Babluyk, E.B.; Kondratov, A.P. The papers printing quality complex assessment algorithm development taking into account the composition and production technological features. *J. Phys. Conf. Ser.* **2018**, *998*, 012003. [CrossRef]
6. Chen, Q.P. Method for Implementing Banknote Counting of Banknote Counting Device, and Banknote Counting Device. U.S. Patent 11,113,917 B2, 7 September 2021.
7. Oba, M. Paper Sheet Sorting and Counting Machine. U.S. Patent D850,524 S, 4 June 2019.
8. Liu, J.D. A Research on Paper Counting System based on Machine Vision. Master's Thesis, Zhengzhou University, Zhengzhou, China, 2019.
9. Mehmet, B.; Mehmet, K.; Alisan, S.; Erhan, A. An image processing based object counting approach for machine vision application. *arXiv* **2018**, arXiv:1802.05911.
10. Han, X.Y.; Wang, J.S. Design of Paper Counting Algorithm based on Texture Image. In Proceedings of the IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chengdu, China, 20–22 December 2019; pp. 145–148.
11. Wang, H.; Chen, D.F.; Wang, X.F. Laminated paper counting algorithm based on compressive sensing and hough transform. In Proceedings of the 2nd International Conference on Information Technology and Electronic Commerce, Dalian, China, 20–21 December 2014; pp. 161–166.
12. Huang, D.; Liao, S.; Yu, S.; Tian, J.; Hu, Y.; Guo, K. Research on Automatic Counting System Corrugated Paper Board Based on Machine Vision. *J. Mech. Eng.* **2017**, *53*, 79–86. [CrossRef]
13. Luan, X.M. *Research on Paper Counting System Based on Image Processing*; Shenzhen University: Shenzhen, China, 2016.
14. Sato, J.; Yamada, T.; Ito, K.; Akashi, T. Vision-Based Facial Oil Blotting Paper Counting. *IEEJ Trans. Electr. Electron. Eng.* **2019**, *14*, 899–907. [CrossRef]
15. Chen, J.Q.; Lu, R.Q.; Liang, Q.M. Design and Implementation of Paper Counting Algorithm Based on Gabor Filter. In Proceedings of the International Symposium on Big Data and Artificial Intelligence, Beijing, China, 22–24 June 2018; pp. 281–285.
16. Ye, H.W.; Guo, X.N.; Chen, J.; Yang, Y.D. Answer sheet counter system based on machine vision. *J. Nanjing Univ. Sci. Technol.* **2020**, *44*, 608–613. [CrossRef]
17. *GB/T 13989-2012*; Subdivision and Numbering for the National Primary Scale Topographic Maps. AQSIQ: Beijing, China, 2012.
18. Cao, W.; Li, R.S. Road intersections detection using deformable part models on remote sensing image. *Geomat. Inf. Sci. Wuhan Univ.* **2018**, *43*, 413–419. [CrossRef]
19. Felzenszwalb, P.F.; Girshick, B.; McAllester, D.; Ramanan, D. A Discriminatively Trained, Multiscale, Deformable Part Model. In Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Anchorage, AK, USA, 23–28 June 2008; pp. 1–8.
20. Felzenszwalb, P.F.; McAllester, D. The Generalized A*Architecture. *J. Arlificial Intell. Res.* **2007**, *29*, 153–190. [CrossRef]
21. Felzenszwalb, P.F.; Huttenlocher, D.P. Distance transforms of sampled functions. *Theory Comput.* **2012**, *8*, 415–428. [CrossRef]
22. Everingham, M.; Van Gool, L.; Williams, C.K.; Winn, J.; Zisserman, A. The pascal visual object classes (voc) challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. [CrossRef]
23. Cai, Y.F.; Liu, Z.; Sun, X.Q.; Chen, L.; Wang, H. Research on pedestrian detection technology based on improved DPM model. In Proceedings of the 2017 IEEE 7th Annual International Conference on CYBER Technology in Automation, Honolulu, HI, USA, 31 July–4 August 2017; pp. 216–219.
24. Bai, S.; Liu, Z.Y.; Chang, Y. Classify vehicles in traffic scene images with deformable part-based models. *Mach. Vis. Appl.* **2018**, *29*, 393–403. [CrossRef]
25. Chu, W.T.; Hsu, M.H.; McAllester, D. Fast object detection using multistage particle window deformable part model. In Proceedings of the 2014 IEEE International Symposium on Multimedia, Taichung, Taiwan, 10–12 December 2014; pp. 98–101.
26. Yang, J.; Li, G.; Wang, W.M.; Wang, R.G. An empirical study of deformable part model with fast feature pyramid. In Proceedings of the 2016 23rd International Conference on Pattern Recognition (ICPR), Cancun, Mexico, 4–8 December 2016; pp. 567–572.
27. Dollár, P.; Appel, R.; Belongie, S.; Perona, P. Fast feature pyramids for object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *36*, 1532–1545. [CrossRef] [PubMed]

28. Chen, M.J.; Bovik, A.C. No-reference color image quality assessment: From entropy to perceptual quality. *EURASIP J. Image Video Process.* **2019**, *2019*, 77. [CrossRef]
29. Yao, H.; Man, B.; Zou, M.; Xu, D.; Yao, J.C. No-reference noisy image quality assessment incorporating features of entropy, gradient, and kurtosis. *Front. Inf. Technol. Electron. Eng.* **2021**, *22*, 1565–1582. [CrossRef]
30. Chandler, D.M. Seven challenges in image quality assessment: Past, present, and future research. *ISRN Signal Process.* **2013**, *2013*, 905685. [CrossRef]
31. Liang, Y.T.; Zhang, W.; Li, Y. A Self-Adaption Single Image Dehaze Method Based on Clarity-evaluation-function of Image. In Proceedings of the 2018 International Conference on Advanced Mechatronic Systems (ICAMechS), Zhengzhou, China, 30 August–2 September 2018; pp. 320–325.