Journal of Image Processing Theory and Applications (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/jipta.2024.070109
ISSN 2560-6239 Vol. 7 Num. 1

Denoising of Traffic Statistics in Multi-frame Video Sequences

Dawei Zhang[1,a], Dan Huang[2,b]

[1]School of Electronic Information Engineering, Beihai Vocational College, Xizang Street, Beihai, China

[2]School of General Education, Beihai University of Art and Design, New Century Street, Beihai, China

[a]davichang@foxmail.com, [b]danh367@126.com

Keywords: Denoising, Traffic Statistics, Video Sequences

Abstract: The advancements in technology have facilitated the interconnection of all things, with computer vision technology emerging as a prominent field in society. As cities continue to evolve towards greater levels of intelligence, the utilization of cameras as the primary means of collecting visual data for road traffic flow monitoring has become a ubiquitous sight. The current research endeavors to devise a sophisticated vehicle flow denoising algorithm that leverages the power of multi-frame video sequences. By processing images captured from road monitoring systems, this algorithm effectively filters out noise within vehicle flow data, enabling the acquisition of more precise vehicular statistics. Consequently, it facilitates a deeper analysis of vehicle counts and densities, ultimately providing vital data support for managing urban pressures and enhancing road decongestion efforts.

1. Introduction

Traffic flow data serves as one of the crucial foundations for traffic management and planning, as it provides valuable insights into road usage, traffic congestion, transportation efficiency, and other pertinent aspects. However, in practical applications, due to various factors, the multi-frame video traffic flow data collected often suffers from noise interference, posing difficulties for subsequent data analysis and prediction. As a result, researching how to eliminate noise from multi-frame video sequence traffic flow data has become an urgent and pressing issue that needs to be addressed.

2. Related Works

Traffic flow denoising refers to improving the accuracy of traffic flow estimation by removing interferences and misdetections from videos. It primarily encompasses detection methods based on image processing and those based on deep learning. Therefore, denoising techniques for traffic flow are closely related to vehicle detection methods. In multi-frame video sequences, all elements apart from vehicles can be categorized as noise. Consequently, traffic flow denoising techniques are intimately tied to advancements in vehicle detection methodologies.

Currently, scholars are mainly tackling the issue of noise in traffic flow data from various perspectives, including image processing, machine learning, and statistical modeling. Li[1] proposes a residual encoder-decoder based on multi-attention fusion attention module (RED-MAM) for ultrasound image denoising, and this method achieves state-of-the-art performance on all three ultrasound datasets. To enhance the efficiency of predicting future traffic flow trends within a transportation network, Xiao[2] propose a traffic flow prediction approach based on Constrained Dynamic Graph Convolutional Recurrent Network (C-DGCRN). Compared to state-of-the-art methods like Adaptive Graph Convolutional Recurrent Network, C-DGCRN showed a reduction of 7%, 9%, and 11% in these metrics. These methods have achieved some results, but they still have certain limitations in complex scenarios. For example, in high-density traffic scenarios, traditional methods frequently encounter difficulties in precisely identifying and counting each vehicle. Furthermore, the substantial requirement for labeled data and computational power posed by certain deep learning algorithms hinders their practical applicability in real-world situations, therefore, the research on vehicle flow denoising algorithms remains challenging and urgent. This paper aims to eliminate background noise from multi-frame video sequences using traditional algorithms, with the goal of obtaining more accurate vehicle statistics.

2.1. Denoising Algorithm Based On Image Processing

The denoising method based on image processing is traditional method, which mainly utilizes the traditional computer vision technology to achieve the accurate detection of the vehicle, and then the picture other than the vehicle is used as the noise, and the noise is processed using the computer vision technology to achieve the purpose of statistics and detection of the vehicle.

Deep learning based vehicle denoising methods utilize deep neural network models trained with large amounts of labeled data to achieve accurate vehicle detection and denoising of video sequences. These methods generally include steps such as data preprocessing, network construction, network training, and object detection. Among them, data preprocessing is used to normalize and enhance the input data, while network construction is used to design the deep neural network model, then network training optimizes and learns the parameters of the model, and target detection is used to achieve the localization and identification of vehicles through the network model.

An autoencoder is an unsupervised learning model that consists of an encoder and a decoder. The encoder compresses the input image into a low-dimensional representation, while the decoder attempts to reconstruct the original image from this representation. By training an autoencoder, a compressed representation can be learned that ignores noise and retains useful information. Generative Adversarial Networks (GANs) consist of a generator and a discriminator, and they generate realistic samples through adversarial training. The generator in a GAN tries to convert a noisy image into a clear image, while the discriminator attempts to distinguish between real clear images and the generated results. Convolutional Neural Networks (CNNs) exhibit strong expressive power in image processing, and by designing appropriate network structures and loss functions, CNNs can be utilized for image denoising tasks. A common approach to better restore details and textures is to use deep Convolutional Neural Networks (CNNs) with residual connections. The Variational Autoencoder (VAE) is a generative model that possesses the capability to learn the probability distribution of input data. By training a VAE, we can enable it to acquire low-dimensional representations of images, and subsequently reconstruct those images using sampling techniques. This process concurrently achieves both image denoising and the generation of clear image samples.

The mean filter is a fundamental denoising technique that reduces noise in images by calculating the average of pixel values within a pixel and its neighborhood, and using this average as the output pixel value. Its advantages include simplicity of computation and high effectiveness in eliminating salt-and-pepper noise. However, it may also lead to blurring of fine details in the image. In contrast, the median filter is a nonlinear processing technique that selects the median value within a pixel's neighborhood as the output. This method is extremely effective in eliminating salt-and-pepper noise while maintaining a good level of edge detail in the image. However, when processing color images, it may introduce color distortion. The Gaussian filter, on the other hand, utilizes the normal distribution (Gaussian function) to smooth images, effectively reducing random noise but potentially sacrificing some high-frequency details. The bilateral filter combines spatial distance between pixels with similarity in pixel values for weighted averaging, achieving both noise reduction and edge clarity preservation, but with increased computational complexity and limited effectiveness against large noise. The non-local means filter, based on global image information for denoising, is highly effective in removing uniform noise while preserving structural information in the image, but it requires significant computational resources and longer processing times.

2.2. Denoising Algorithm Based On Deep Learning

In order to conduct a thorough research on traffic flow denoising algorithms, it is imperative to obtain suitable video datasets. In this study, we have selected video footage from traffic surveillance cameras as our primary data source. To ensure the high quality and accuracy of the data, we have carefully chosen representative traffic segments as our observation areas, and deployed high-resolution cameras to precisely capture and record vehicle movements. (shown in Figure 1).

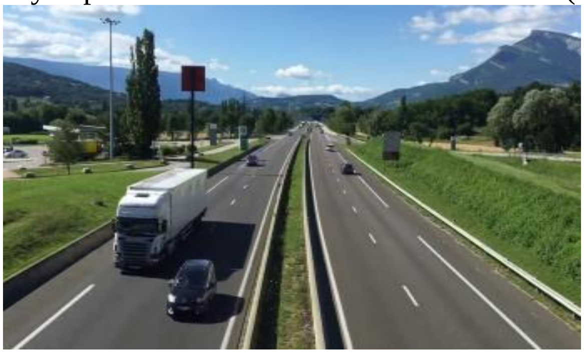

Figure 1: A Multi-frame video

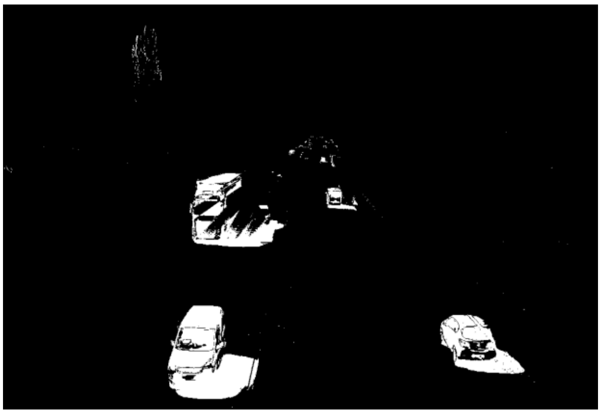

Figure 2: Multi-frame video after binarization processing

In Figure 1, one can clearly observe the original image of a specific frame from a video surveillance sequence, which retains all the details and color information of the scene. In contrast, Figure 2 showcases the result of this frame after undergoing a meticulously designed binarization process. This process takes into full account the characteristics of the scene and employs a flexibly adjusted empirical threshold, rather than a rigid fixed value, achieving a simplification of the image from colorful to black and white while preserving crucial structural and contour information. Such a processing approach facilitates subsequent image processing and analysis tasks.

3. Video Sequence Image Denoising Method Design

3.1. Experimental Environment

In order to thoroughly investigate the vehicle flow denoising algorithm based on video analysis technology, we meticulously designed and executed a series of experiments aimed at comprehensively evaluating the performance and practical application effects of this algorithm. Table 1 provides a detailed list of the hardware and software environment configurations utilized in this experiment. This configuration scheme ensures sufficient computational resources, enabling efficient support for complex and diverse video analysis tasks, thereby laying a solid foundation for the research and validation of the algorithm.

Table 1: Experimental Environment

Processor	Intel Core i7 9750
GPU	RTX 3080Ti
RAM	64G
OS	Ubuntu 22.04 LTS
Language	Python
Vision library	OpenCV 3.4.1.15
IDE	Pycharm
Editor	Jupyter notebook

3.2. Algorithm Design

Binarization is an image processing technique that simplifies images by reducing them to only two grayscale levels: black and white. In the binarization process, each pixel in the image is converted to either black or white based on its grayscale value, with black typically representing a lower grayscale level (such as 0) and white representing a higher grayscale level (such as 255). This conversion significantly facilitates the separation of target objects from the background and simplifies subsequent image processing and analysis tasks.

The primary goal of binarization is to classify the information in an image by setting an appropriate threshold, enabling clear distinction between target objects and the background. Depending on how the threshold is set, binarization can be divided into two main methods: global binarization and adaptive binarization.

● Global Binarization: In this method, a single, uniform threshold is applied to the entire image for binarization. This threshold is determined based on the overall characteristics of the image (such as the average grayscale value or histogram analysis) and is suitable for images with distinct grayscale differences between the background and target objects.

● Adaptive Binarization: Unlike global binarization, adaptive binarization automatically adjusts the threshold for different regions of the image. This method considers the grayscale distribution characteristics of local image regions, enabling more effective separation of target objects from the background under varying lighting conditions or complex backgrounds, thereby improving the accuracy and robustness of binarization.

The global binarization method employs a fixed, global threshold to uniformly process the entire image. This approach is highly effective when there are significant grayscale differences between the target objects and the background in the image, enabling effective separation of the target objects from the background. However, in complex scenarios such as uneven illumination or the presence of noise in the image, global binarization may yield unsatisfactory results, including unclear identification of target objects or excessive background noise.

In contrast, the adaptive binarization method is more flexible, dynamically setting thresholds based on the local neighborhood information surrounding each pixel in the image. This approach can effectively adapt to environments with uneven illumination or noise, ensuring more ideal binarization results across different regions of the image. Common adaptive binarization techniques include methods based on local mean, local median, and the Otsu algorithm. These methods dynamically calculate and apply thresholds by analyzing the grayscale characteristics of the pixel's surrounding neighborhood, achieving precise binarization processing for different regions within the image.

Whether it is global binarization or adaptive binarization, selecting the appropriate threshold or threshold calculation method is crucial. An excessively high threshold may result in the loss of target object information, while an excessively low threshold may introduce excessive background noise. Therefore, in practical applications, it is necessary to adjust parameters and conduct experimental verifications based on the specific characteristics of the image and the requirements of the task to find the optimal threshold setting scheme.

Multi-frame sequence video is composed of a series of closely connected image frames, with a particular focus on preserving the detailed changes and dynamic characteristics at each time point. Unlike conventional video, which often displays continuous scenes at a fixed frame rate, creating a smooth visual effect but retaining only one image frame per time point, multi-frame sequence video achieves more detailed recording by increasing the sampling rate of the image sequence, capturing more information at each time point.

These additional image frames demonstrate immense value in post-production, as they can be used for in-depth extraction of visual features, more precise analysis, and effective enhancement of video quality. Multi-frame sequence video has a wide range of applications, particularly in cutting-edge fields such as computer vision, image processing, and computational photography. For instance, in the field of motion tracking, utilizing continuous image sequences enables more accurate tracking of object positions and motion trajectories. In video enhancement, multi-frame sequence technology can effectively reduce noise, improve image quality, and enhance detail representation. Additionally, this technology is widely applied in advanced visual processing tasks such as generating super-resolution images, implementing motion compensation and blending.

At its core, video is composed of frames, each of which is essentially a static image, viewed as a form of bitmap. Bitmaps, in turn, are composed of countless pixel points. Image binarization, a common visual processing technique, simplifies the grayscale value of each pixel in an image to either 0 or 255 (as shown in Formula 1). This process significantly simplifies image information, markedly reduces the consumption of computational resources, and facilitates subsequent processing.

$$g(x,y)=\begin{cases}255 & f(x,y)\geq Threshold \\ 0 & otherwise\end{cases} \tag{1}$$

In visual processing target object can be considered as foreground and non-target object can be considered as background. So, all targets other than vehicles in this project can be considered as background. After binarization of the video, the expected background becomes foreground, and this part is the noise that needs to be processed (as shown in Fig. 2).

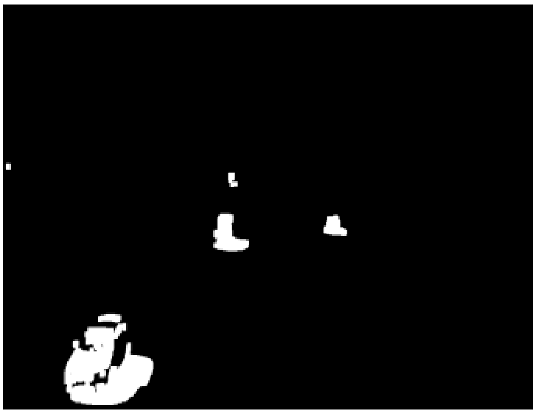

Figure 3: binarization video after denoising

Handling video noise involves various methods, but the core objective remains the same: to minimize its interference with the foreground content of the video. In OpenCV, for images that have already been binarized (where the foreground is typically represented by white), we can employ morphological operations such as erosion and dilation to further process the noise.

The basic concept of erosion operation is to define a structuring element B and translate it over the image X. At each position a, if B is fully contained within the corresponding region of X, that position a is recorded. The set of all a points that satisfy this condition is the result of X being eroded by B. This process effectively "shrinks" the foreground objects by one layer, thereby removing fine noise points along their edges (as shown in Figure 4). The calculation formula for erosion can be referred to in relevant literature or in Formula 2 of OpenCV's official documentation.

Through erosion, we can effectively remove noise points adhering to the edges of foreground objects, but it may also result in a reduction in the size of the foreground objects themselves. To restore the size of the foreground objects or further process remaining noise, dilation is often performed after erosion. In contrast to erosion, dilation "expands" or "enlarges" the foreground objects, filling in the areas reduced by erosion and potentially eliminating some isolated noise points.

$$A - B = \{x \mid B_x \subseteq A\} \tag{2}$$

The fundamental process of dilation operation is to translate the predefined structuring element B over the image X. Whenever B is translated to a position a, if B overlaps with X at that position, that position a is recorded. The set of all a points satisfying the above condition is referred to as the result of X being dilated by B (as shown in Figure 5). The calculation formula for dilation can be referred to in relevant literature or in Formula 3 of OpenCV's official documentation (note that "Formula 3" here is a placeholder; actual usage requires consulting the specific documentation).

Through dilation, we can "expand" or "enlarge" the foreground objects to fill in the areas reduced by erosion and potentially eliminate some isolated noise points, thereby restoring or enhancing the morphology of the foreground objects.

$$A \oplus B = \{x \mid (B)_x \cap A \neq \Theta\} \tag{3}$$

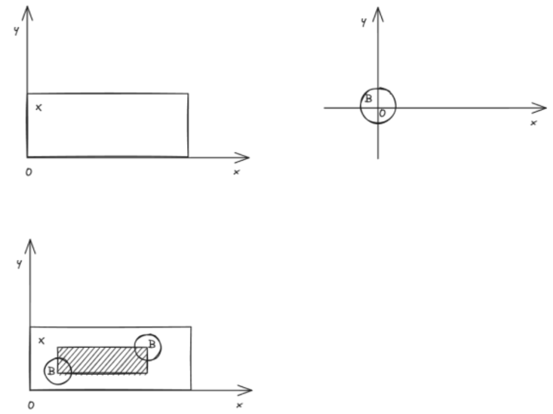

Figure 4: Erosion

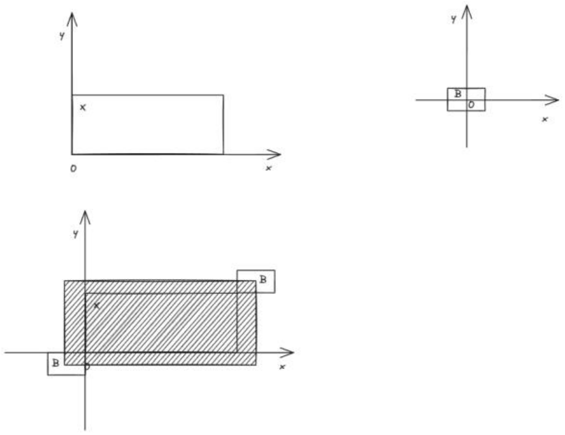

Figure 5: Dilation

After several adjustments of the parameter, the noise in the video sequence is denoised, and the effect graph shown in Fig. 6 is obtained, and it can be seen that most of the noise has been removed.

After the denoising work is completed, in order to better mark the vehicle, need to draw the outline of the detected target, the use of common minimum outer rectangle can be (as shown in Fig. 7). Based on the minimum area rectangle algorithm, the basic idea of the algorithm is as follows:

Step 1: find the smallest rectangle containing the set of points whose sides are parallel to the X and Y axes;

Step 2. for each rectangle, calculate its area;

Step 3. for the rectangle with the smallest area among all the rectangles, rotate it so that its edges can be rotated arbitrarily.

The vehicles in the multi-frame sequence video automatically draw the outline, the problem in front of the distant target is small, almost difficult to distinguish, is it necessary to recognize? In order to better solve this problem, a detection region is drawn in the video (as shown by the blue line in Fig. 8), and those passing through this region are counted, while those beyond this region are not counted, so that both the statistical accuracy can be improved and the computational consumption can be reduced. The position of the blue line is not fixed, it depends on the actual needs, it uses a coordinate value, when the detected object exceeds this coordinate value will be recorded.

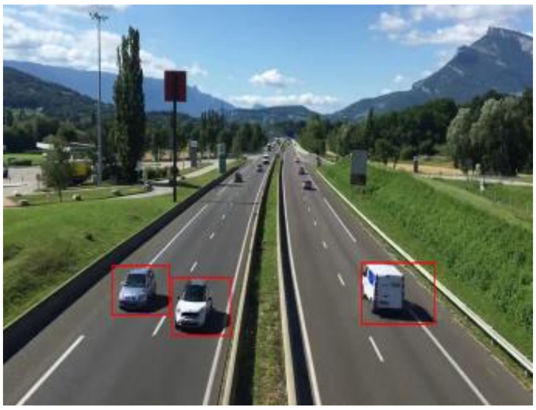

Figure 6: Drawing vehicle contours

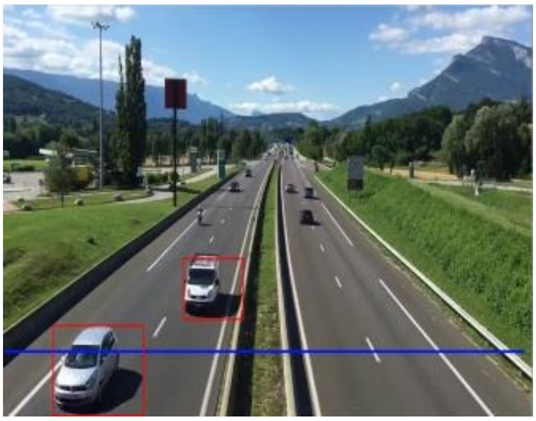

Figure 7: Identify boundaries

When the above work is completed, the flow statistics can be carried out in two ways: one is to count the number of vehicles whose coordinate value crosses the blue line, so that the approximate traffic flow can be calculated; the other is to access the counting system and realize networking, so that multi-road flow monitoring can be realized. Through the detailed analysis and discussion of the experimental results, the performance and effect of the denoising algorithm can be obtained, and the algorithm can be optimized and improved to provide a valuable reference for further optimization and improvement. Provide valuable reference for further optimization and improvement of the algorithm.

Note that this algorithm is not perfect and counts only the number of objects that cross the blue line, not the number of cars. An object is counted when it crosses the blue line, regardless of whether it is a car or not. The traffic monitoring system does not recognize cars, so the statistics are approximate.

4. Conclusion

In this study, video analysis technology was innovatively employed to tackle the issue of noise in traffic flow data. In contrast to traditional methods that heavily rely on sensors and other equipment, which are prone to interference from external environmental factors during data collection, this study ingeniously leveraged widely installed surveillance cameras and other infrastructure, combined with advanced image processing and machine learning techniques, effectively achieving denoising of video data and significantly enhancing the robustness and stability of the data.

However, due to practical constraints in data collection and processing, the dataset utilized in this study is relatively small in scale. Therefore, before applying the research findings to real-world scenarios, it is necessary to further validate the algorithm's widespread applicability and generalization capabilities under different conditions.

Furthermore, while the algorithm proposed in this study performs well in various common situations, its performance may be somewhat impacted in extreme weather conditions or in particularly complex road traffic scenarios, revealing certain limitations. To address this issue, future research can focus on the in-depth optimization and improvement of the algorithm to enhance its adaptability and stability in a wide range of complex environments.

Acknowledgements

2023 Guangxi University Young and Middle-aged Teachers' Basic Scientific Research Ability Improvement Project "Research on Semantic Segmentation of Sugar Planting Plots Based on Remote Sensing Images" (2023KY1493)

References

[1] LI Yancheng et al. RED-MAM: A residual encoder-decoder network based on multi-attention fusion for ultrasound image denoising [J]. Biomedical Signal Processing and Control, 2023, 79(P1)

[2] Xiao H ,Zhao Z ,Yang T .A traffic flow prediction method based on constrained dynamic graph convolutional recurrent networks[J].Engineering Applications of Artificial Intelligence, 2024, 133(PE):108486.