Hindawi
International Journal of Digital Multimedia Broadcasting
Volume 2021, Article ID 6641633, 10 pages
https://doi.org/10.1155/2021/6641633

Research Article

Vehicle Reidentification via Multifeature Hypergraph Fusion

Wang Li,[1] **Zhang Yong**,[2] **Yuan Wei**,[1] **and Shi Hongxing**[1]

[1]*College of Computer Science, Open University of China, Beijing, China*
[2]*Beijing University of Technology, Beijing, China*

Correspondence should be addressed to Wang Li; wlpolo@ouchn.edu.cn

Received 29 December 2020; Revised 21 February 2021; Accepted 7 March 2021; Published 18 March 2021

Academic Editor: Marco Roccetti

Vehicle reidentification refers to the mission of matching vehicles across nonoverlapping cameras, which is one of the critical problems of the intelligent transportation system. Due to the resemblance of the appearance of the vehicles on road, traditional methods could not perform well on vehicles with high similarity. In this paper, we utilize hypergraph representation to integrate image features and tackle the issue of vehicles re-ID via hypergraph learning algorithms. A feature descriptor can only extract features from a single aspect. To merge multiple feature descriptors, an efficient and appropriate representation is particularly necessary, and a hypergraph is naturally suitable for modeling high-order relationships. In addition, the spatiotemporal correlation of traffic status between cameras is the constraint beyond the image, which can greatly improve the re-ID accuracy of different vehicles with similar appearances. The method proposed in this paper uses hypergraph optimization to learn about the similarity between the query image and images in the library. By using the pair and higher-order relationship between query objects and image library, the similarity measurement method is improved compared to direct matching. The experiments conducted on the image library constructed in this paper demonstrates the effectiveness of using multifeature hypergraph fusion and the spatiotemporal correlation model to address issues in vehicle reidentification.

1. Introduction

Currently, traffic video surveillance plays a vital role in ensuring public safety. A critical part of traffic video surveillance in the urban and major cities is the monitoring of vehicles, which include detection, tracking, and classification. Targeted reidentification technology has emerged as a prudent application in the field of vehicle recognition, particularly important for public safety departments in its efforts to track target vehicles in intricate urban transportation networks. The main task of vehicle reidentification is to search for images of the same vehicle captured by multicameras in various areas on the precondition that the target vehicle is known in the surveillance videos. Vehicle reidentification differs from vehicle detection, classification, and tracking and can be used to address instance-level target search issues.

Similar to the task of person reidentification, the research for vehicle reidentification is divided into two aspects. Firstly, a model is built based on information of the appearance, and vehicles are distinguished according to their appearance characteristics. The second aspect is to use the distance measurement method that utilizes samples to train a distance measurement model to carry out vehicle reidentification through the principle of reducing and expanding the differences within intraclasses and between interclasses, respectively. However, vehicle reidentification faced greater challenges. Compared to person reidentification, the information of different vehicle images has higher similarities, and the same vehicle will display appearance disparities due to external factors such as lighting and weather. In addition, vehicles of the same make and model are identical, and it is difficult to distinguish the different IDs of the vehicles visually. As such, a robust method for vehicle reidentification is required to distinguish target vehicles from nontarget vehicles.

In the process of image identification, the image information is stored on the graph structure with strong description capability and high validity. The vertices in the graph represent the image feature elements, while the edges represent the relationship between these feature elements. Although graph

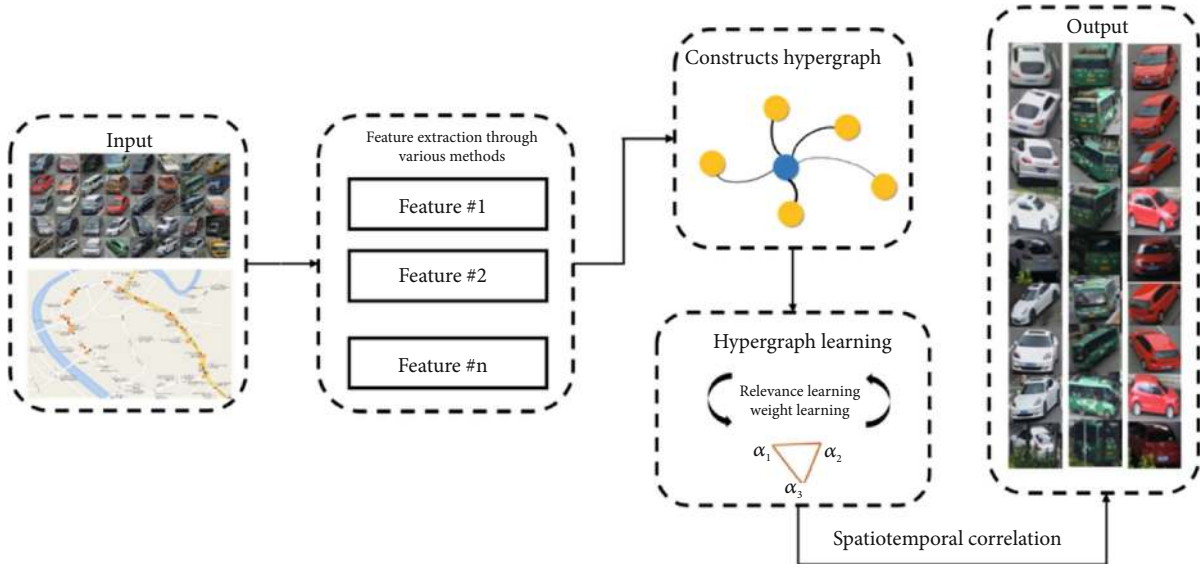

Figure 1: Overview of vehicle reidentification methods based on multifeature hypergraph fusion.

matching is often used to address image feature matching issues, these graphs only represent the binary relations between vertices. Comparatively, hypergraphs can describe multiple and higher-order relations of objects as the edges can contain multiple vertices.

This paper uses various feature extraction methods, constructs a hypergraph for each feature, and combines these hypergraphs. Through hypergraph learning, the query image is matched with image sets and, combined with spatiotemporal information of the vehicle image, achieves multitarget identification from multichannel traffic surveillance videos.

Unlike the traditional method that only calculates pairing distance or similarities for matching, this method utilizes hypergraphs optimization to learn about the similarities between the query image and library of images. In this way, the pairing and higher-order relation between query objects and image library objects are utilized, thereby improving the similarity measurement method as compared to direct matching. In addition, complementary information of different feature descriptors can be effectively utilized through learning multiple hypergraphs and feature descriptors. The fusion of multiple hypergraphs is realized through regularising a completely connected graph with each vertex representing the weight of a hypergraph. Our method is compatible with any feature description methods. Figure 1 illustrates an overview of the method.

There are three main contributions of this paper:

(1) We use different algorithms to extract different features of vehicles and construct hypergraphs, respectively, to represent their relationships. In the form of hypergraphs, we can express their deeper relationships

(2) We explored hypergraph learned algorithms to integrate multiple traditional feature extraction methods. This contribution combines the advantages of multiple methods to improve the accuracy of apparency matching

(3) The trickiest challenge of vehicle re-ID problem is the high similarity of the vehicles. It is known that the same production has the same appearance in most cases. However, the camera time of similar-looking cars passing through different regions may be different; there is a spatiotemporal correlation. Thus, we applied the spatiotemporal features of the vehicle as additional constraints to improve the matching accuracy of similar vehicles

2. Related Work

Common reidentification systems such as facial recognition [1, 2] and person recognition [3, 4], and vehicle reidentification has gradually become a popular research topic in recent years. Researchers in China and other countries have made numerous attempts on tracking vehicles. Currently, feature extraction methods are commonly used, combined with an SVM classifier, background modeling [5], etc. to build a tracking model for the target. Building a classifier requires certain initial samples. It also requires the target vehicle to have the same feature distribution when it appears between multiple cameras in order to have an improved tracking outcome. The background modeling method is based on the correlation of pixels between frames and robust real-time performance, but it is sensitive to light changes and camera movements and is unable to track targets across cameras.

The common vehicle feature classifications include contour and length. However, with the distance between cameras and the placement of cameras, the scale and angle changes and is not suitable for cross-camera tracking of vehicles [6]. Previous research utilizes a combination body. SIFT feature is invariant to rotation, scaling, changes in brightness, etc., which is a highly stable local feature [7]. Researchers ensured the stability by updating SIFT feature of the target in real-time [8]. However, the intermediate frame information is lacking after the target crosses between cameras as a SIFT

feature is unable to describe the target information after it reappears, resulting in the loss of tracking of the target and is unable to guarantee the accuracy of cross-camera tracking.

Since 2012, major breakthroughs have been achieved in object detection due to deep learning, among which the most critical technologies are Convolution Neural Network (CNN) [9–11] and region proposal algorithm [12]. In 2012, Hinton et al. successfully improved the accuracy of the Top 5 classification of 1000 object categories from 75% to 85% through the use of the deep learning AlexNet network. Since then, the convolutional neural network has been widely recognized by academics and in the industrial industry for its excellent performance. Apart from its application in vehicle verification [13], vehicle classification [14], vehicle driving safety [15, 16], and attribute prediction [17, 18], it has also been continuously applied to the fields of artificial intelligence such as computer vision and language recognition. During this period, the architecture and performance of convolutional neural network have continuously improved. Literature [19, 20] studies the comprehension and codification of contextual information to improve the performance of neural networks. In the paper, Liu et al. [21–23] combined texture, color, and depth features and, through the fusion of low-level and high-level features for vehicle reidentification, achieved positive results for the "VeRi" data set. Xu et al. [24] proposed RepNet, a multitask learning framework that can be used to learn about the characteristics of both coarse-grained and fine-grained vehicles. The paper also uses the bucket search method to improve retrieval speed.

Based on the theory of deep learning, this paper proposes a cross-camera vehicle tracking method based on feature matching and hypergraph fusion. The algorithm extracts multidimensional features of vehicle images, combines a spatiotemporal distribution model of the monitoring points, and utilizes a multifeature hypergraph fusion method for vehicle reidentification. This ensures efficiency and improves the accuracy of vehicle reidentification.

3. Multifeature Hypergraph Fusion

In this section, we first outline the concept of hypergraphs and why we need hypergraph representation to model the vehicle Re-ID problem. In the second part, we describe how we construct hypergraphs for vehicle features extracted by different methods. In the third part, we show the multiple hypergraphs learning processes.

3.1. Problem Definition. For graph-based image identification, the image information is stored on the graph structure with strong description capability and high validity. But traditional graphs only consider the pairing relationship represented by the edge that connects the two vertices. Comparatively, hypergraphs can describe multiple and higher-order relations of objects as the edges can contain multiple vertices.

A hypergraph consists of sets of vertices and hyperedges. The vertices in the hypergraph represent the image feature elements, while the hyperedges represent the relationship between these feature elements. Each hyperedge can connect to more than two vertices. For hypergraph, the correlation matrix H of size $|V| \times |E|$ is defined, whereby $|V|$ and $|E|$ represents the cardinality of vertex set and hyperedge set, respectively. The value of each element of H is

$$h(v,e) = \begin{cases} 1, \text{if } v \in e \\ 0, \text{if } v \notin e \end{cases}. \tag{1}$$

Indicates if vertex v is connected to edge e.

The degree of vertex $v \in V$ in the hypergraph is defined as

$$d(v) = \sum_{e \in E} w(e) h(v,e), \tag{2}$$

$w(e)$ is the total weight of all related hyperedges of vertex v and the degree of hyperedge $e \in E$ is defined as

$$\delta(e) = \sum_{v \in V} h(v,e). \tag{3}$$

Represents the number of vertices connected to hyperedge e.

The regularisation framework for classification can be expressed as

$$\underset{f}{\operatorname{argmin}} \left\{ \lambda R_{\text{emp}}(f) + \Omega(f) \right\} \tag{4}$$

f is the classification function,$R_{\text{emp}}(f)$is the empirical loss function, λ is the regular parameter, and regular classification is defined as $\omega(f)$

$$\Omega(f) = \frac{1}{2} \sum_{e \in E} \sum_{u,v \in V} \frac{w(e) h(u,e) h(v,e)}{\delta(e)} \left(\frac{f(u)}{\sqrt{d(u)}} - \frac{f(v)}{\sqrt{d(v)}} \right)^2. \tag{5}$$

By solving formula (4), prediction results, such as classification labels, can be obtained.

3.2. Hypergraph Structure. For vehicle reidentification task, there are huge differences within the class, often a single type of feature cannot well describe the vehicle information. In order to match a vehicle, different feature extraction methods are used to describe the features of an image. We set hypergraphs for the features extracted by each method. A hypergraph $G_i = (V_i, E_i)$ is built for each feature description and each vertex represents an image.

The construction of the hyperedge adopts a star expansion method. It selects a vertex as the center and connects with its nearest vertices to form a hyperedge. The strength of the connection is determined by the similarity between the center and the connected vertices. Under such circumstances, the binary connection of only 0 and 1 in formula (1) cannot be used. Each element of the correlation matrix H_i is defined as

$$h_i(v,e) = \begin{cases} s(v,e), & \text{if } s(v,e) \le \eta \\ 0 & \text{if } s(v,e) > \eta \end{cases}, \tag{6}$$

where $s(v, e)$ represents the connection strength of vertex v and hyperedge e defined as

$$s(v, e) = \exp\left(-\frac{\text{dist}(v, c)^2}{\sigma^2}\right), \tag{7}$$

where $\text{dist}(v, c)$ is the distance between vertex v and center c, η is the predefined threshold value; σ is the control parameter.

The i hypergraph can be expressed as

$$\begin{aligned}\Omega_i(f) &= \frac{1}{2}\sum_{e\in E_i}\sum_{u,v\in V_i}\frac{w_i(e)h_i(u,e)h_i(v,e)}{\delta_i(e)}\times\left(\frac{f(u)}{\sqrt{d_i(u)}}-\frac{f(v)}{\sqrt{d_i(v)}}\right)^2\\ &= \sum_{e\in E_i}\sum_{u,v\in V_i}\frac{w_i(e)h_i(u,e)h_i(v,e)}{\delta_i(e)}\times\left(\frac{f^2(u)}{\sqrt{d_i(u)}}-\frac{f(u)f(v)}{\sqrt{d_i(u)d_i(v)}}\right)\\ &= \sum_{u\in V_i}f^2(u)\sum_{e\in E_i}\frac{w_i(e)h_i(u,e)}{d_i(u)}\sum_{v\in V_i}\frac{h_i(v,e)}{\delta_i(e)}\\ &\quad-\sum_{e\in E_i}\sum_{u,v\in V_i}\frac{f(u)h_i(u,e)w_i(e)h_i(v,e)f(v)}{\sqrt{d_i(u)d_i(v)}\delta_i(e)}\\ &= \mathbf{f}^\top(\mathbf{I}-\Theta_i)\mathbf{f},\end{aligned} \tag{8}$$

where $\Theta_i = \mathrm{D}_{v,i}^{-1/2}\mathrm{H}_i\mathrm{W}_i D_{e,i}^{-1}H_i^\top\mathrm{D}_{v,i}^{-1/2}$,$D_{v,i}$ is a diagonal matrix whose diagonal element is the degree of the vertex calculated with formula (2). $D_{e,i}$ is also a diagonal matrix whose diagonal element is the degree of the hyperedge calculated using formula (3). W_i is the diagonal matrix whose diagonal element is the hyperedge weight. Since different connectivity strength is considered in formula (6), it is set as a unit matrix in the experiment. $\mathbb{f}$ is a learned correlation vector that contains the similarity between the query image and other images. The regularisation of the hypergraph structure under the constraint of (8) indicates that the more the two vertices are connected by the hyperedge, the higher the probability that they share similar labels. In addition, $\Delta_i = \mathbf{I} - \Theta_i$ is the Laplacian of the hypergraph, then the regularisation of the graph becomes

$$\Omega_i(f) = \mathbf{f}^\top\Delta_i\mathbf{f}. \tag{9}$$

3.3. Hypergraph Fusion

3.3.1. Regularisation of Graphs. To construct a hypergraph for each feature description, we obtain n_f hypergraphs $G_1 = (V_1, E_1), G_2 = (V_2, E_2), \cdots, G_{n_f} = (V_{n_f}, E_{n_f})$, these graphs are fused based on different weights and every hypergraph has weight α_i, and the sum of the weights of all hypergraphs is 1; then, the regularisation terms of multiple hypergraphs as a whole are expressed as

$$\begin{aligned}\Omega(\mathbf{f}) = \frac{1}{2}\sum_{i=1}^{n_f}\alpha_i\sum_{e\in E_i}\sum_{u,v\in V_i}&\frac{w_i(e)h_i(u,e)h_i(v,e)}{\delta_i(e)}\\ &\times\left(\frac{f(u)}{\sqrt{d_i(u)}}-\frac{f(v)}{\sqrt{d_i(v)}}\right)^2.\end{aligned} \tag{10}$$

By combining formula (8), (10) can be expressed as

$$\Omega(\mathbf{f}) = \sum_{i=1}^{n_f}\alpha_i\mathbf{f}^\top(\mathbf{I}-\Theta_i)\mathbf{f} = \mathbf{f}^\top\sum_{i=1}^{n_f}\alpha_i(\mathbf{I}-\Theta_i)\mathbf{f} = \mathbf{f}^\top\widehat{\Delta}\mathbf{f}, \tag{11}$$

where $\widehat{\Delta} = \sum_{i=1}^{n_f}\alpha_i(\mathbf{I}-\Theta_i)$.

3.3.2. Regularisation of Graph Weight. The weights of different hypergraphs are correlated. Hypergraphs with similar structure (i.e., similar correlation matrices) are expected to have similar weight. The similarity between hypergraphs G_i and G_j is defined as

$$\gamma(G_i, G_j) = \exp\left(-\frac{\left\|\mathrm{H}_i - \mathrm{H}_j\right\|_F^2}{\sigma^2}\right). \tag{12}$$

The similarity of two hypergraphs is measured by the F norm of the two correlation matrices. The correlation matric Γ between hypergraphs is expressed as $\Gamma(i, j) = \gamma(G_i, G_j)$.

The weights of multiple hypergraphs are defined as $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \cdots, \alpha_{n_f}]$; the cost function of the weights of the hypergraphs is

$$J(\boldsymbol{\alpha}) = \boldsymbol{\alpha}^\top(\mathbf{I}-\mathbf{S})\boldsymbol{\alpha}, \tag{13}$$

where $\mathbf{S} = \mathbf{D}^{-1/2}\Gamma\mathbf{D}^{-1/2}$, $\mathbf{D}$ is a diagonal matrix and $\mathbf{D}(i, i) = \sum_j\Gamma(i, j)$. In addition, the hypergraphs weights are L2 regularised to avoid the hypergraphs being too large.

3.3.3. Empirical Loss Function. The empirical loss function $R_{\text{emp}}(f)$ in formula (4) can be expressed as

$$R_{\text{emp}}(\mathbf{f}) = \|\mathbf{f}-\mathbf{y}\|^2 = \sum_{u\in V}(\mathbf{f}(u)-\mathbf{y}(u))^2. \tag{14}$$

$\mathbf{y}$ is the vector of the binary tags. In this paper, connecting the query image and image to be matched can be regarded as a classification issue. The label vector $\mathbf{y}\in\mathbf{R}^{N+1}$, N is the total number of images in the image library to be matched, and the first element corresponding to the query image is set to 1, while the rest are set to 0.

3.3.4. Objective Function. Combining the target loss function $R_{\text{emp}}(\mathbf{f})$ in formula (14), regularisation of graphs $R_{\text{emp}}(f)$ in formula (14), hypergraph weight regularisation $J(\alpha)$ in formula (13), and L2 weight regularisation in α, the overall objective function to be optimized is

$$\text{`}\underset{\mathbf{f},\boldsymbol{\alpha}}{\text{argmin}}\left\{\Omega(\boldsymbol{f})+\lambda_1 R_{\text{emp}}(\mathbf{f})+\lambda_2 J(\boldsymbol{\alpha})+\lambda_3\boldsymbol{\alpha}^\top\boldsymbol{\alpha}\right\},\ \text{s.t.}\boldsymbol{\alpha}^\top\mathbf{1}=1. \tag{15}$$

The purpose of introducing $\lambda_i (i = 1, 2, 3)$ is to balance all parts of the cost function and provide an optimal cost function. Specifically, λ_1 is the parameter controlling the loss of experience, λ_2 is the parameter that controls the weight of the graph defined by cost function in formula (13), and λ_3

is the L2 regularisation parameter of the hypergraph weight in α. Formula (15) can be expressed as

$$\underset{\mathbf{f},\boldsymbol{\alpha}}{\operatorname{argmin}}\left\{\mathbf{f}^{\top}\widehat{\Delta}\mathbf{f}+\lambda_1\|\mathbf{f}-\mathbf{y}\|^2+\lambda_2\boldsymbol{\alpha}^{\top}(\mathbf{I}-\mathbf{S})\boldsymbol{\alpha}+\lambda_3\boldsymbol{\alpha}^{\top}\boldsymbol{\alpha}\right\},\ \text{s.t.}\boldsymbol{\alpha}^{\top}\mathbf{1}=1. \tag{16}$$

3.3.5. Optimisation. In order to understand the optimized formula (16), the iterative method is used to update $\mathbf{f}$ and $\boldsymbol{\alpha}$.

When $\boldsymbol{\alpha}$ is fixed, the optimization of $\mathbf{f}$ becomes

$$\underset{\mathbf{f},\boldsymbol{\alpha}}{\operatorname{argmin}}\left\{\mathbf{f}^{\top}\widehat{\Delta}\mathbf{f}+\lambda_1\|\mathbf{f}-\mathbf{y}\|^2\right\}. \tag{17}$$

Formula (17) derivate $\mathbf{f}$ and the analytical solution of $\mathbf{f}$ is

$$\mathbf{f}=\left(\mathbf{I}+\frac{1}{\lambda_1}\Delta\wedge\right)^{-1}\mathbf{y}. \tag{18}$$

Because an inverse matrix is involved in the matrix, it is impossible to calculate the inverse of $\widehat{\Delta}$ directly when $\widehat{\Delta}$ becomes larger. As such, the solution of $\mathbf{f}$ can be obtained iteratively with the following formula.

$$\mathbf{f}^{(t+1)}=\frac{1}{1+\lambda_1}\left(\mathbf{I}-\widehat{\Delta}\right)\mathbf{f}^{(t)}+\frac{\lambda_1}{1+\lambda_1}\mathbf{y}. \tag{19}$$

t is the number of iterations to ensure the convergence of the iteration process in formula (19).

When $\mathbf{f}$ is fixed, the optimization of α becomes

$$\underset{\boldsymbol{\alpha}}{\operatorname{argmin}}\left\{\mathbf{f}^{\top}\widehat{\Delta}\mathbf{f}+\lambda_2\boldsymbol{\alpha}^{\top}(\mathbf{I}-\mathbf{S})\boldsymbol{\alpha}+\lambda_3\boldsymbol{\alpha}^{\top}\boldsymbol{\alpha}\right\},\ \text{s.t.}\boldsymbol{\alpha}^{\top}\mathbf{1}=1. \tag{20}$$

The Lagrangian function of formula (20) can be written as

$$\begin{aligned}L(\boldsymbol{\alpha},\gamma)&=\mathbf{f}^{\top}\widehat{\Delta}\mathbf{f}+\lambda_2\boldsymbol{\alpha}^{\top}(\mathbf{I}-\mathbf{S})\boldsymbol{\alpha}+\lambda_3\boldsymbol{\alpha}^{\top}\boldsymbol{\alpha}+\gamma\left(\boldsymbol{\alpha}^{\top}\mathbf{1}-1\right)\\&=\mathbf{f}^{\top}\sum_{i=1}^{n_f}\alpha_i(\mathbf{I}-\Theta_i)\mathbf{f}+\lambda_2\boldsymbol{\alpha}^{\top}(\mathbf{I}-\mathbf{S})\boldsymbol{\alpha}+\lambda_3\boldsymbol{\alpha}^{\top}\boldsymbol{\alpha}+\gamma\left(\boldsymbol{\alpha}^{\top}\mathbf{1}-1\right)\\&=\sum_{i=1}^{n_f}\alpha_i\mathbf{f}^{\top}(\mathbf{I}-\Theta_i)\mathbf{f}+\lambda_2\boldsymbol{\alpha}^{\top}(\mathbf{I}-\mathbf{S})\boldsymbol{\alpha}+\lambda_3\boldsymbol{\alpha}^{\top}\boldsymbol{\alpha}+\gamma\left(\boldsymbol{\alpha}^{\top}\mathbf{1}-1\right)\\&=\boldsymbol{\alpha}^{\top}\mathbf{P}+\lambda_2\boldsymbol{\alpha}^{\top}(\mathbf{I}-\mathbf{S})\boldsymbol{\alpha}+\lambda_3\boldsymbol{\alpha}^{\top}\boldsymbol{\alpha}+\gamma\left(\boldsymbol{\alpha}^{\top}\mathbf{1}-1\right),\end{aligned} \tag{21}$$

where $\mathbf{P}=[\mathbf{f}^{\top}(\mathbf{I}-\Theta_1)\mathbf{f},\cdots,\mathbf{f}^{\top}(\mathbf{I}-\Theta_{n_f})\mathbf{f}]^{\top}$. Formula (21) derivate α and sets it to 0, and α can be solved using formula (22).

$$\boldsymbol{\alpha}=\frac{1}{2}(\lambda_2\mathbf{S}-(\lambda_2+\lambda_3)\mathbf{I})^{-1}(\gamma\mathbf{1}+\mathbf{p}). \tag{22}$$

Because $\boldsymbol{\alpha}^{\top}\mathbf{1}=\mathbf{1}^{\top}\boldsymbol{\alpha}=1$, γ can be solved with the following formula.

$$\frac{1}{2}\mathbf{1}^{\top}(\lambda_2\mathbf{S}-(\lambda_2+\lambda_3)\mathbf{I})^{-1}(\gamma\mathbf{1}+\mathbf{p})=1. \tag{23}$$

In addition, $\mathbf{Q}=\lambda_2\mathbf{S}-(\lambda_2+\lambda_3)\mathbf{I}$, the solution of γ can be derived from formula (23) as shown in the following formula.

$$\gamma=\frac{2-\mathbf{1}^{\top}\mathbf{Q}^{-1}\mathbf{p}}{\mathbf{1}^{\top}\mathbf{Q}^{-1}\mathbf{1}}. \tag{24}$$

Substituting the solution of γ back into formula (22), the hypergraph weight vector α can be obtained with the following formula.

$$\boldsymbol{\alpha}=\frac{1}{2}\mathbf{Q}^{-1}\left(\frac{2-\mathbf{1}^{\top}\mathbf{Q}^{-1}\mathbf{p}}{\mathbf{1}^{\top}\mathbf{Q}^{-1}\mathbf{1}}\mathbf{1}+\mathbf{P}\right). \tag{25}$$

Iteratively, update $\mathbf{f}$ and α until convergence and the reidentification result is obtained by ranking the learned correlation of the matching results in correlation vector $\mathbf{f}$.

4. Spatiotemporal Correlation

In the process of vehicle reidentification, apart from utilizing the feature extraction method for image matching, we also need to consider the spatiotemporal relationship when the vehicle is moving. The data set proposed in this paper has low image clarity, and it is difficult to achieve the optimal results for vehicle reidentification by using only feature matching. At the same time, this data set has abundant and comprehensive spatiotemporal information compared to other public vehicle reidentification data sets, including vehicle location, speed, and elapsed time. Therefore, the spatiotemporal correlation model built based on this data set can provide targeted and efficient extraction of spatiotemporal information.

Vehicle reidentification differs from person reidentification. People have strong particularities and the degree of uniqueness of each person is made up of appearance, dressing, hairstyle, and other personal characteristics, making it easy to determine whether two images are of the same person. Vehicles have a high degree of similarity. Disregarding license plate information and external car modification details, vehicles of the same brand, make and model has a high degree of similarity. With the low vehicle image resolution of this data set, license plate information cannot be used as the basis for the judgment and other unclear details such as window stickers, determination of whether it is the same vehicle based on appearance characteristics will have a great impact on the accuracy.

When this data set is built, the comprehensive traffic flow parameters are saved as reference factors of spatiotemporal relationship. Each vehicle image has a corresponding monitoring point, the time it passes this point, and the speed. The distance between the monitoring points is measured and stored in the distance matrix. In addition, apart from recording the current speed of the vehicle, it also calculates the average speed of the vehicle passing through two monitoring points as a reference, thereby improving the accuracy of spatiotemporal correlation calculation.

The spatiotemporal relationship can be expressed in many ways, but it follows the same principle. The same

vehicle will not exist in two distant locations within a short time range, and the same vehicle appearing in two locations will be within a short period of time. This paper calculates the conformity of the average speed of the vehicle passing through two points and the actual situation as the spatiotemporal distance of two vehicles. The calculation method is

$$\mathrm{Dis}(i,j) = \frac{\left|\lambda_1\left(V_i + V_j/2\right) + \lambda_2 V_{\mathrm{ave}}(i,j) + \lambda_3 V_{\mathrm{lim}} - S(i,j)/T_j - T_i\right|}{S(i,j)/T_j - T_i}, \tag{26}$$

where i and j represent two vehicle images, V_i and V_j is the current speed of two vehicles, and $V_{\mathrm{ave}}(i,j)$ represents the average speed of the vehicles between the two points where the two vehicles are located, respectively. V_{lim} is the maximum speed limit of 60 km/h, and $\lambda_1, \lambda_2, \lambda_3$ represent the weight parameters of the average speed of two points, the average speed of the road section, and the maximum speed limit of the road. $S(i,j)/T_j - T_i$ indicates the actual average speed of the two vehicles passing through two points, $S(i,j)$ is the distance between two points, and $T_j$andT_i refer to the time recorded when two vehicles travel to the current position. The spatiotemporal distance $\mathrm{Dis}(i,j)$ calculates the probability if the two vehicles are the same vehicle in terms of spatiotemporal correlation. The larger the value, the farther the spatiotemporal distance is, the less likely they are the same vehicle. The smaller the value, the closer the spatiotemporal distance is, the more likely they are the same vehicle. After obtaining the similarity vector calculated with the multifeature hypergraph fusion method, the spatiotemporal distance between the query image and each matching image is calculated, and the matching sequence is reordered by combining the similarity and spatiotemporal distance to obtain a more accurate reidentification result.

5. Dataset

In practical application scenarios, vehicle image quality is often poor, making it impossible to apply methods using existing HD data sets. Therefore, we propose a vehicle rerecognition data set based on real traffic monitoring with low resolution. And our data set contains the space-time relationship of the vehicle.

Our data set is based on the high-speed monitoring of Beijing Airport and the monitoring of Tongzhou G103. The camera points are shown in the Figures 2 and 3, and we have clearly recorded these points. Faster R-CNN and Hungary algorithm were used to detect and track the vehicle in the traffic surveillance video to propose the vehicle type, speed, and time and store the vehicle ID and the video point ID together with the vehicle image to the local area.

Our data set consists of three parts: 606 vehicles entering Beijing on Airport Expressway, 662 vehicles leaving Beijing on Airport Expressway, and 638 vehicles in the direction of Tongzhou G103 entering Beijing. The three parts of the data set are relatively independent, and the shooting angle of the same part of the data is relatively consistent. The resolution of some airport expressway images is 80×80. The angle of the camera in the direction entering Beijing shows the rear of the vehicle, and the camera in the direction of leaving Beijing shows the front of the vehicle. The image quality of Tongzhou G103 is relatively high, with a resolution of 150×150. Some examples are shown in Figure 4.

6. Experimental Results and Analysis

6.1. Evaluation Indicator. The evaluation indicator, mean Average Precision (mAP), is the sum of the average precision (AP) in multiclassification tasks. To calculate the average accuracy, we need to first obtain the accuracy rate and recall rate. Precision is the proportion of the image with the same ID as the query image in the query result, that is, how many of the matching vehicle images are correct. Recall if the ratio of the number of images with the same ID as the query image in the query results to the total number of images in the search database, that is, how many of the correct results have been retrieved. If the number of positive samples with correct prediction is TP, the number of positive samples with incorrect prediction is FP, the number of negative samples with correct prediction is TN, and the number of negative samples with prediction error is FN, the calculation method of the accuracy rate P and the recall rate R is as shown in the following formula.

$$\begin{aligned} P &= \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \\ R &= \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}. \end{aligned} \tag{27}$$

Although the accuracy rate and recall rate seem to have a certain significance, they cannot be used directly to evaluate the performance of the ReID model. Therefore, a PR graph is derived and the number of query results is gradually increased. From the first to the last query results given by the system, the accuracy and recall rate corresponding to the middle of each point are plotted on the graph to obtain a curve graph.

The area enclosed by this PR curve and the coordinate axis, or AP, can reflect the performance of the model to a certain extent. By sorting AP, the formula can be expressed as

$$\mathrm{AP} = \frac{\sum_{k=1}^{n} P(k) \times gt(k)}{N_{gt}} \tag{28}$$

Where n is the number of vehicles in the test set, N_{gt} represents the number of vehicle images that are actually related to the query image at other locations, $P(k)$ represents the accuracy of matching in the first k results, and $gt(k)$ is a symbolic function. If the kth image and the query image belong to the same ID, return to 1, otherwise, return to 0. AP is the average accuracy of query images, as such, the average accuracy of all query images mAP can be expressed as

$$\mathrm{mAP} = \frac{\sum_{q=1}^{Q} \mathrm{AP}(q)}{Q}, \tag{29}$$

Figure 2: Points distribution and vehicle orientation in the direction of entering Beijing.

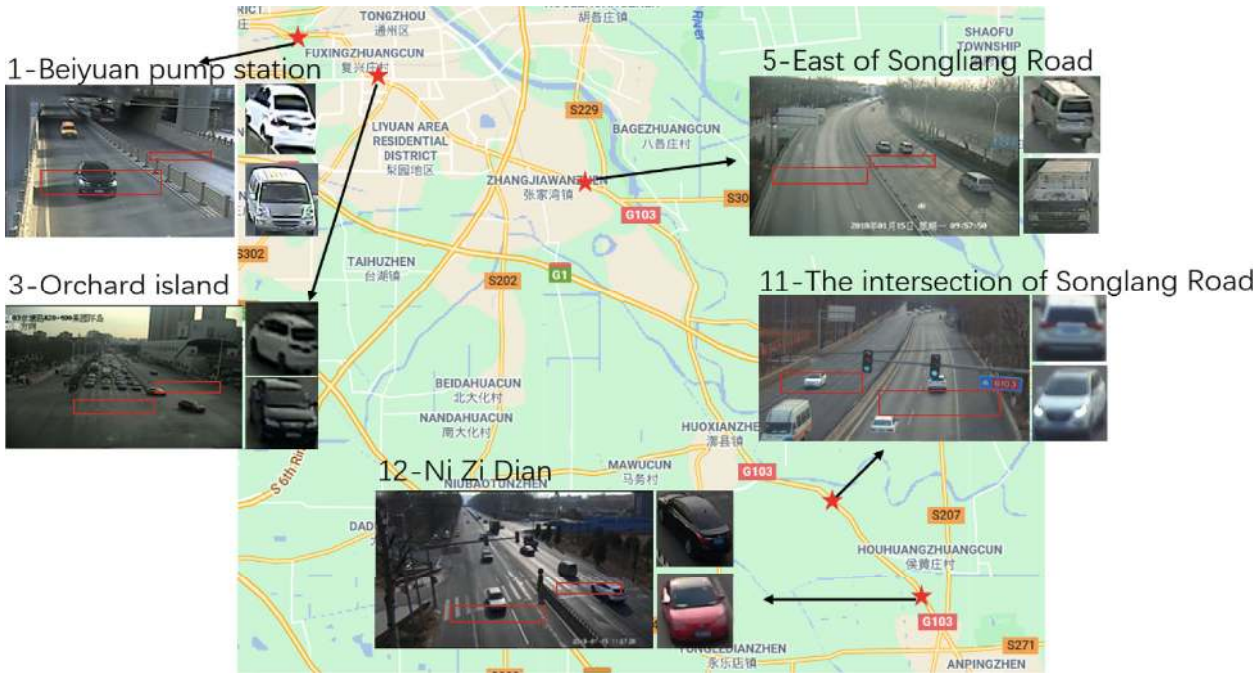

Figure 3: Points distribution and vehicle orientation in the direction of leaving Beijing.

where Q is the total number of query images. In addition to mAP, this paper also selects rank-1 and rank-5 as the auxiliary criteria for performance evaluation. Rank-1 represents the probability that the image ranked first in the matching result sequence is the correct result, and rank-5 represents the probability of a correct result in the first 5 digits.

6.2. Algorithm Performance and Result Analysis. In this paper, the data set selected for the vehicle reidentification experiment is the vehicle reidentification image library based on traffic video surveillance constructed in Chapter 5. The image pairs were randomly and uniformly selected as the training and test set. The experiment was conducted multiples times, and the average value is taken as the final experiment result.

This method selects three feature extraction algorithms, namely, SIFT, ORB, and HSV. On this basis, multifeature hypergraph fusion and spatiotemporal correlation constraints are added, and the accuracy is significantly improved. Tables 1–4 show the specific experimental results.

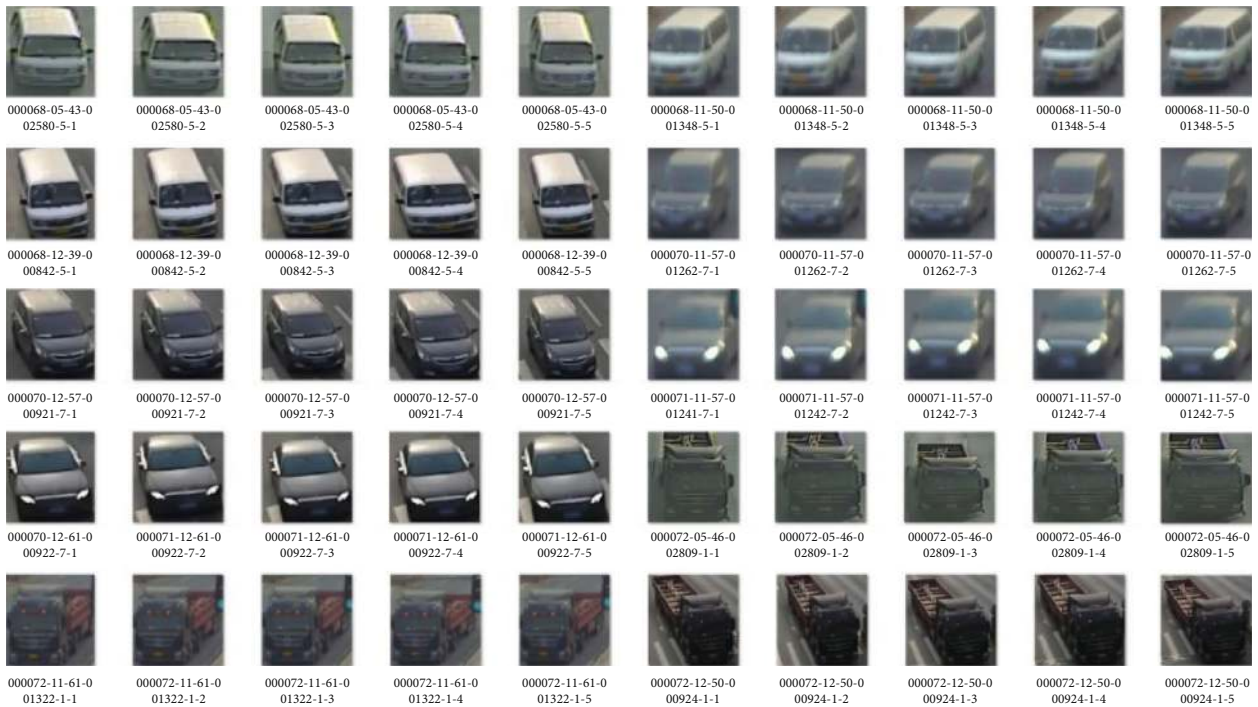

Figure 4: Dataset of five points of Tong Zhou G103 entering Beijing.

Table 1: ReID accuracy of Airport Expressway vehicles entering Beijing.

Method	mAP (%)	Rank-1 (%)	Rank-5 (%)
SIFT+ORB+HSV	14.34	33.54	46.87
SIFT+ORB+HSV+HG	18.36	42.71	51.32
SIFT+ORB+HSV+HG+S-T	**23.46**	**51.11**	**68.63**

Table 2: ReID accuracy of Airport Expressway vehicles leaving Beijing.

Method	mAP (%)	Rank-1 (%)	Rank-5 (%)
SIFT+ORB+HSV	16.15	35.64	48.24
SIFT+ORB+HSV+HG	19.73	44.54	57.95
SIFT+ORB+HSV+HG+S-T	**24.79**	**52.68**	**72.15**

Table 3: ReID accuracy of Tong Zhou G103 vehicles entering Beijing.

Method	mAP (%)	Rank-1 (%)	Rank-5 (%)
SIFT+ORB+HSV	22.21	47.25	66.95
SIFT+ORB+HSV+HG	26.52	56.24	81.31
SIFT+ORB+HSV+HG+S-T	**30.46**	**75.11**	**88.63**

Tables 1–3 show the reidentification results of the three data sets of Beijing inbound and outbound vehicles on the Airport Expressway and on Tongzhou G103. SIFT+ORB +HSV are the three feature extraction methods selected in this paper through the simple fusion of setting weight parameters. HG is the added multifeature hypergraph fusion method, and S-T is the added spatiotemporal correlation constraint. The results of Table 1—Airport Expressway vehicles entering Beijing—are relatively poor. The shooting angle of some data sets is the rear of the vehicle, and the characteristics are not obvious compared to the front of the vehicle. The shooting angle in Table 2—Airport Expressway vehicles leaving Beijing—is the front of the vehicle and the identification accuracy slightly improved. However, the clarity of the expressway is poor, and the image quality of the data set is low, which has a great impact on vehicle reidentification, thereby lowering the reidentification accuracy on the part of the airport expressway. Compared to the 80×80 image pixels of the airport expressway, the image resolution of the partial data set of Tongzhou G103 is 150×150, and the clarity of the vehicle significantly improved. Therefore, the reidentification experiment results of partial data sets of Tongzhou G103 significantly improved compared to the airport expressway. It can be found in these three tables that the addition of the multifeature hypergraph fusion method has significantly improved the accuracy of reidentification and with the adding of spatiotemporal correlation constraints, the accuracy further improved, proving the effectiveness of combining multifeature hypergraph fusion with spatiotemporal correlation. Table 4 compares the experimental results

Table 4: Accuracy comparison of ReID methods.

Method	mAP (%)	Rank-1 (%)	Rank-5 (%)
AlexNet	9.69	42.39	55.09
GoogLeNet	17.88	58.87	74.10
FACT	18.49	50.95	73.48
FACT+Plate+STR	27.77	61.44	78.78
SIFT+ORB+HSV+S-T+HG	**30.46**	**75.11**	**88.63**

of this method with the other four methods on the VeRI data set and proves to a certain extent the advanced nature of the method proposed in this paper.

Although the reidentification methods proposed in this paper have significantly improved the results compared with some other methods, its accuracy has not reached the best and optimal state. There are many factors affecting the accuracy of the vehicle reidentification method in this paper. Firstly, some cameras are installed at the intersection. When the vehicle arrives at the intersection, it will slow down, stop, or start gradually, affecting the accuracy of speed measurement. Secondly, the spatiotemporal correlation model does not take into account the number of traffic lights between the monitoring points, which has different waiting times, causing the use of spatiotemporal features to be inaccurate. Lastly, the video definition lacks clarity, saved vehicle images are fuzzy, vehicle contour is unclear, and there is a certain degree of color disparity in some vehicle images. All these factors cause the reduction of reidentification accuracy.

Due to our data set is derived from real traffic footage, its video resolution is low and so are the images saved. In addition, the image disparity of the same vehicle in this data set under different cameras is far greater than that of the same make and model but different vehicles under the same camera. The vehicle images in the data sets used by other methods are clear and display sharp details, which makes the data set in this paper unsuitable to be used for other reidentification methods. In addition, this data set contains a relatively comprehensive spatiotemporal correlation information, including position, speed, and time, and considers vehicle speed as an important parameter, while other data sets do not store information such as vehicle speed, making other public data sets unbefitting to be applied to this method.

7. Conclusion

We used three feature extract methods to extract features from vehicle images, constructed a hypergraph for each feature, and through hypergraph learning, fused multiple hypergraphs to match the query image and the image set. In addition, the spatiotemporal correlation between vehicles constrains the degree of image matching and improves the accuracy of vehicle reidentification. The complementary information of different feature descriptors can be effectively utilized through learning and using multiple hypergraphs and feature descriptors, respectively. The experiments conducted on the image library constructed in this paper proves the effectiveness of using multifeature hypergraph fusion and the spatiotemporal correlation model to address issues in vehicle reidentification.

Data Availability

The image data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] Y. Sun, D. Liang, X. Wang, and X. Tang, *DeepID3: Face Recognition with Very Deep Neural Networks*, Computer Vision and Pattern Recognition, 2015, http://arxiv.org/abs/1502.00873.

[2] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 815–823, Boston, MA, USA, June 2015.

[3] T. Xiao, S. Li, B. Wang, L. Lin, and X. Wang, "Joint detection and identification feature learning for person search," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3415–3424, Honolulu, HI, USA, July 2017.

[4] H. Zhao, M. Tian, S. Sun et al., "Spindle net: person re-identification with human body region guided feature decomposition and fusion," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1077–1085, Honolulu, HI, USA, July 2017.

[5] R. J. Wood, D. Reed, J. Lepanto, and J. M. Irvine, "Robust background modeling for enhancing object tracking in video," in , Article ID 908902*Geospatial InfoFusion and Video Analytics IV; and Motion Imagery for ISR and Situational Awareness II*, vol. 9089, Baltimore, MD, USA, 2014.

[6] H. Liu, Y. Tian, Y. Yang, L. Pang, and T. Huang, "Deep relative distance learning: tell the difference between similar vehicles," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2167–2175, Las Vegas, NV, USA, June 2016.

[7] Y. Zhou and L. Shao, "Aware attentive multi-view inference for vehicle re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6489–6498, Salt Lake City, UT, USA, 2018.

[8] Y. Zhou, L. Liu, and L. Shao, "Vehicle re-identification by deep hidden multi-view inference," *IEEE Transactions on Image Processing*, vol. 27, no. 7, pp. 3275–3287, 2018.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097–1105, 2012.

[10] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, http://arxiv.org/abs/1409.1556.

[11] C. Szegedy, W. Liu, Y. Jia et al., "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–9, Boston, MA, USA, 2015.

[12] B. Alexe, T. Deselaers, and V. Ferrari, "Measuring the objectness of image windows," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2189–2202, 2012.

[13] L. Yang, P. Luo, C. Change Loy, and X. Tang, "A large-scale car dataset for fine-grained categorization and verification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3973–3981, Boston, MA, USA, June 2015.

[14] J. Krause, M. Stark, J. Deng, and L. Fei-Fei, "3d object representations for fine-grained categorization," in *Proceedings of*

the IEEE international conference on computer vision workshops, pp. 554–561, Sydney, NSW, Australia, 2013.

[15] S. Patra, D. Van Hamme, P. Veelaert et al., "Detecting vehicles' relative position on two-lane highways through a smartphone-based video overtaking aid application," *Mobile Networks and Applications*, vol. 25, no. 3, pp. 1084–1094, 2020.

[16] S. Patra, W. Zamora, C. T. Calafate, J. C. Cano, P. Manzoni, and P. Veelaert, "Using the smartphone camera as a sensor for safety applications," in *Proceedings of the 5th EAI International Conference on Smart Objects and Technologies for Social Good*, pp. 84–89, 2019.

[17] K. Ramnath, S. N. Sinha, R. Szeliski, and E. Hsiao, "Car make and model recognition using 3d curve alignment.," in *IEEE winter conference on applications of computer vision*, pp. 285–292, Steamboat Springs, CO, USA, March 2014.

[18] Y. Xiang, W. Choi, Y. Lin, and S. Savarese, "Data-driven 3d voxel patterns for object category recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1903–1911, Boston, MA, USA, June 2015.

[19] M. Roccetti, G. Delnevo, L. Casini, and P. Salomoni, "A cautionary tale for machine learning design: why we still need human-assisted big data analysis," *Mobile Networks and Applications*, vol. 25, no. 3, pp. 1075–1083, 2020.

[20] G. Delnevo, M. Roccetti, and S. Mirri, "Intelligent and good machines? The role of domain and context codification," *Mobile Networks and Applications*, pp. 1–9, 2019.

[21] X. Liu, W. Liu, T. Mei, and H. Ma, "A deep learning-based approach to progressive vehicle re-identification for urban surveillance," in *European conference on computer vision*, pp. 869–884, Springer, 2016.

[22] X. Liu, W. Liu, H. Ma, and H. Fu, "Large-scale vehicle re-identification in urban surveillance videos," in *2016 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6, Seattle, WA, USA, July 2016.

[23] X. Liu, W. Liu, T. Mei, and H. Ma, "Provid: progressive and multimodal vehicle reidentification for large-scale urban surveillance," *IEEE Transactions on Multimedia*, vol. 20, pp. 645–658, 2017.

[24] Q. Xu, K. Yan, and Y. Tian, "Learning a repression network for precise vehicle search," 2017, http://arxiv.org/abs/1708.02386.