sciendo

Transport and Telecommunication, 2020, volume 21, no. 1, 15–31
Transport and Telecommunication Institute, Lomonosova 1, Riga, LV-1019, Latvia
DOI 10.2478/ttj-2020-0002

THE USE OF MULTI-SENSOR VIDEO SURVEILLANCE SYSTEM TO ASSESS THE CAPACITY OF THE ROAD NETWORK

Vladimir Shepelev[1], Sergei Aliukov[1]*, Kseniya Nikolskaya[1], Arkaprava Das[2], Ivan Slobodin[1]

[1]Institute of Engineering and Technology, South Ural State University
Chelyabinsk 454080, Russia
shepelevvd@susu.ru (V.S.); nikolskaiaki@susu.ru (K.N.); slobodinis@yandex.ru (I.S.)
**Correspondence: alysergey@gmail.com (S.A.)*
[2] Atmospheric Plasma Division, Institute for Plasma Research
Gandhinagar 382428, Gujarat, India
arkapravadas222@gmail.com (A.D.)

Currently, in many cities around the world there is a significant increase in the number of vehicles, which leads to an aggravation of problems and contradictions in the road and transport system. This is especially true of traffic congestion, since the presence of the congestion leads to a number of negative consequences: an increase in travel time, additional fuel consumption and vehicle wear, stress and irritation of drivers and passengers, environmental poisoning and others. To solve the problem of congestion, it is necessary to have a reliable system for collecting information about the situation on the roads and a well-developed method for analyzing the collected information. The paper discusses the possibilities of collecting the required information using multi-touch video cameras and ways to improve them. A distinctive feature of this study is the registration of pedestrians crossing the road at the intersection. The aim of the work is to develop methods for collecting information using road sensor video surveillance systems in a traffic congestion and data processing using statistical methods such as: multiple regression analysis, cluster analysis, multidimensional scaling methods and others. The tasks were set: 1) to identify the most significant factors affecting the intensity of movement of vehicles at intersections in a congestion; 2) divide congestion into clusters with the identification of their characteristics; 3) to give a visual representation of multidimensional statistical information obtained with the help of multi-touch road video cameras.

Keywords: traffic congestion, multi-touch video cameras, vehicle detection, pedestrians, factors, road situation analysis

1. Introduction

The number of photos and video cameras that record situation on the roads of the countries of the world is increasing every year. The cameras reduce the emergency situation on the relevant sections of the roads, they are assistants in the work of traffic inspectors, because the patrol cars can not fix all the violators. Cameras definitely make the roads safer. Automatic analysis of vehicle activity in monitoring traffic in cities is an important and urgent problem due to the large number of traffic violations and their adverse impact on the daily traffic management. Today, when traveling in European countries, it is almost impossible to meet a policeman on the roads, but, oddly enough at first glance, almost all drivers try not to violate the rules of the road. The secret is quite simple – the Europeans have been using very successfully video recording cameras for violations for a long time. Recently, in Russia on the roads more and more such devices can be found. In recent years, the number of fixed cameras on Russian roads has increased in tens times.

There are various types of camcorders. The main ones are mobile, portable and stationary. Mobile cameras are installed directly in a car of a traffic inspector. Portable cameras require daily installation and configuration. Stationary cameras are very convenient and practical. Stationary cameras (Figure 1) are constantly located on the same place of the road. They need to be configured only once. Stationary cameras operate automatically and can control movement simultaneously in several lanes, including on the opposite lanes. The most common and frequently used video surveillance systems in Russia are “Strelka”, “Kordon”, “Rapira”, and some others.

As a rule, traffic video cameras are used to record violations of traffic rules. However, we should not forget that the data obtained using video surveillance systems can be successfully and effectively used to analyze the traffic situation in order to improve the road and transport infrastructure. This is especially true of crossings, since it is on these sections of the road that the greatest number of problem cases occur. Of great interest is the analysis of road congestion. In the absence of traffic congestion, traffic occurs normally and there are no specific reasons for changing the status quo. The presence of congestion due to the large number of negative consequences requires immediate intervention in the organization of the road process, finding causes of congestion and ways to eliminate these causes in the future. An invaluable service in solving such a problem can be provided by the analysis of information obtained from multi-touch video cameras.

The detection and classification of vehicles are important parts of intelligent transport systems. They help to monitor and calculate traffic, which is necessary to track the effectiveness of road operations.

Despite the widespread occurrence of traffic congestion as a phenomenon, there is still no universally accepted definition of the notion of "congestion." According to one of the definitions, a congestion or a traffic jam is an accumulation of vehicles moving at an average speed on the road much lower than the normal speed for the given road segment. Such a definition is not without controversial points, namely what “normal speed” and what “significantly less” mean. Therefore, it is necessary to determine a criterion that allows to speak about the presence or absence of congestion. In addition, it is necessary to develop a methodology that allows using the information obtained from video cameras to analyze the traffic situation, to identify the main factors of the road infrastructure that most affect the causes of traffic congestion. The solution of such a problem requires the widespread use of statistical and computer methods. These and other issues are considered in this article.

Figure 1. Stationary video cameras for automatic recording of road traffic

Our paper is structured as follows. In Section 2, we briefly discuss the relevant works of video recording of road users and analysis of the collected data. In Section 3 we focus on the aim and scope of our paper. In Section 4, we describe our research methodology, the process of variable selection with some examples and the data collection system in our case with an explanation of its advantages over existing vehicle detection systems. Besides in this Section we describe the use of statistical data analysis methods. In Section 5, we process the camera data using the methods of factor, cluster analysis, methods of multidimensional scaling and multiple regression. In the same section, we identify the main factors affecting throughput in congestion conditions. In addition, we carry out segmentation of congestions in accordance with their characteristics. Section 6 presents the results of the study and brief conclusions on the work.

2. Literature Review

At the moment, the calculation and categorization of road transport is an important task, as evidenced by many publications on this subject.

Yiren Zhou *et al.* (2019) have taken as the basis of the Deconvolutional Neural Network. The average recognition accuracy was 86%. However, the authors used cameras removing the car from the rear view. On cameras with a different review, their accuracy dropped significantly.

Debojit Biswas *et al.* (2017) developed and implemented two algorithms BSM and OverFeat Framework for automatic counting of cars based on the Convolution Neural Network. Accuracy is assessed by comparison with manual counting. The average recognition accuracy was 96.55%. The system was trained on 3698 images in which 6 classes were allocated. Authors used ready dataset from Stanford Image-net library [Image Net Library. Available online: http://imagenet.stanford.edu/].

The paper (Zhang *et al.*, 2019) provides a different approach. The authors did not begin to develop their own system, they took trained networks and made an application for counting road transport based on convolutional neural networks. The studies showed a good result. However, as can be seen from the examples given in the paper, if the camera was taken for tests from a different angle, then the recognition percentage fell. Despite the availability of ready-made datasets, as well as a variety of ready-made solutions, a number of unsolved problems remain. For example, the quality of the images from the cameras differs from each other, and this leads to the fact that the neural network learn on the existing dataset, but it does not recognize the image from the necessary cameras other than those on which the dataset was going. Another problem is the location of the camera. If the camera hangs from a different angle, the same problem arises as with the image quality. Therefore, until a universal algorithm has been developed for detecting and classifying automobile transport, it is necessary to select your own tools for each task and create your own datasets.

Currently, there are methods for detecting vehicles in the deep CNN area based on suggesting areas. These methods first create candidate areas in the image, and then classify each one of them. The original R-CNN (Girshick *et al.*, 2014) applies high-throughput CNN to the ascending candidate areas proposed using the selective search algorithm (Uijlings *et al.*, 2013), which effectively improves accuracy. Although R-CNN combines regional offerings with CNN, this leads to high computational costs without sharing of convolutional layers.

Existing methods use various types of information to detect and classify vehicles, including acoustic signature (Wang *et al.*, 2013), radar signal (Kim and Song, 2013), frequency signal (McKay *et al.*, 2012), and image presentation (Mishra and Banerjee, 2013). The development of image processing methods, along with the widespread use of road cameras, facilitates the detection and classification of vehicles based on the images.

Tianyu Tang *et al.* (2017) propose a through convolutional neural network for the direct generation of randomly oriented detection results. Their approach, called Oriented_SSD (Single Shot MultiBox Detector, SSD), uses a set of default blocks with different scales at each object map location to create bounding detection blocks. At the same time, offsets are predicted for each block by default, which better corresponds to the shape of the object. This method can determine both the location and orientation of the vehicle with high accuracy and high speed. For test images in the DLR ship antenna dataset with a size of 5616 × 3744, the method reaches 76.1% of average accuracy (AP) and 78.7% of correct direction classification at 5.17 s on the NVIDIA GTX-1060.

Convolutional neural networks (CNN) have achieved a breakthrough in computer vision tasks and have achieved great success in the classification of road signs. Jianming Zhang *et al.* (2017) presented the Chinese algorithm for detecting road signs, based on a deep convolutional network. To detect Chinese road signs in real time, the authors proposed an end-to-end convolutional network based on YOLOv2. To efficiently detect small road signs, they divided the input images into dense grids in order to get more accurate feature maps. All the experimental results, evaluated in accordance with the Chinese Advanced CTSD and the German Road Sign Detection Standard (GTSDB), have shown that the proposed method is fast enough and reliable. The highest detection rate was 0.017 seconds per image.

You can note the methods of detection of objects based on regression. These methods attract much attention as they significantly increase the speed of detection using a single neural network. For example, the YOLO regression method (Redmon *et al.*, 2016) splits the input image into several grids and predicts the bounding box and confidence directly in each grid. This method works fairly quickly. However, YOLO makes numerous localization errors compared to algorithms based on a region proposal. An improved model, called YOLOv2, removes fully bonded layers and uses an anchor cell to offer a trade-off between speed and accuracy (Redmon and Farhadi, 2016).

To ensure good performance in vehicle detection, an algorithm based on sensors has been proposed (Tang *et al.*, 2017). This algorithm provides significant improvements in accuracy over existing methods. The paper (Deng, 2017) proposed the associated R-CNN method, which combines a network of exact vehicle offers and a network of vehicle attribute studies in order to detect vehicles quickly and accurately. A vehicle detection algorithm is also proposed that uses convolutional neural networks based on a spatial pyramidal pool (Qu *et al.*, 2017), which can better adapt to input images of different sizes to study the multi-scale characteristics of objects.

Buch *et al.* (2012) presented a brief overview of intelligent traffic monitoring systems using road cameras. Daigavane and Bajaj (2010) developed a method for background recording and segmentation using a morphological operator. In this study, a system was developed for the dynamic detection and counting of objects on motorways. The system effectively combines the knowledge of the subject area about classes of objects with statistical indicators in the time domain and identifies the target objects in the presence of partial occlusions and ambiguous positions. Chen *et al*. (2001) addressed issues related to uncontrolled image segmentation and object modeling using multimedia inputs to describe the spatial and temporal behavior of an object to monitor traffic. Gupte *et al*. (2002) showed the algorithms for detecting and classifying vehicles based on monocular images of motion scenes, which are recorded by a stationary camera.

A new single-shot object detection method, called RefineDet, is presented in article (Zhang *et al*., 2018). This method simultaneously optimizes the anchor refinement module and the object detection module for their effective detection. In the paper (Liu, 2018), a structural logical network (SIN) is described, which considers the scene contextual information and the relationships in a single image. At the same time, it is proposed to consider object detection as a problem of displaying the structure of a graph and get the desired result. To solve the scaling problem when an object is detected, the paper (Zhou, 2018) presents a scalable network for the detection of a multi-scale object based on a dense convolutional network.

Faster R-CNN architecture has been used in many vehicle detection works. Suichan Li (2018) proposed to process several adjacent frames in order to better cope with blurriness and short-term occlusions. Xiaoliang Wang *et al*. (2018) explored the use of focal loss. Tsung-Yi Lin *et al*. (2017) monitored vehicles and showed that, being a relatively simple technique, focal loss provides a significant improvement in performance. Xiaowei Hu *et al*. (2019) focused on increasing the reliability of Faster R-CNN scaling and suggested a context-sensitive association RoI (CARoI), which uses deconvolution with bilinear cores to accurately represent functions for small objects. In addition, the CARoI pool runs on top of several levels to use high and low semantic information to further improve performance.

3. Objective and Scope

This study is aimed at collecting, processing and analyzing information about the traffic situation and the movement of vehicles at street intersections in a traffic congestion. Information is collected by using sensory stationary video cameras with a reasonable level of accuracy and precision. A distinctive feature is the consideration of pedestrians, crossing the intersections in different directions. When processing the collected information, such statistical methods are used as: multiple regression analysis, cluster analysis, multidimensional scaling methods, principal component method, correlation analysis and others. Information processing is performed using the SPSS computer program. When processing the information, the characteristic of the traffic flow is as follows: "The actual number of passing cars." This variable is the dependent one (the last one in the Table 1). The independent variables are all the others indicated in the Table 1. The purpose of this study in this article is to develop a methodology that allows us to identify the main factors affecting the throughput of intersections in a congested environment. Analysis of the results of information processing allows us to do this. In addition, the work is clustering intersections and provides a visual representation of the source data. Achieving this goal will allow taking the right steps to improve road and transport logistics. It is important to note that the results obtained make it possible to do predictions about throughput of an intersection under given conditions. Examples of such a prediction in the work are given.

4. Research Methodology

The research methodology in the article includes the selection of variables for which data is collected, the method of data collection, the use of statistical methods for processing the collected information and the interpretation of the results obtained.

4.1. Variable Selection

In the study, the data were collected for various intersections in Chelyabinsk (Russia). At each intersection, lanes were selected that meet the following requirements: turn right and conflict with pedestrians in a jam state. Twenty-five such bands were selected. The data was collected over a certain period of time from "date" to "date". The data was collected using software developed in South Ural State University (Chelyabinsk). The set of the variables from the video stream used in the study is presented in Table 1.

Table 1. Variable description

Variable	Units
Duration of the resolving signal of a traffic light	second
The number of pedestrians in the direction of the vehicle (right)	one unit
The number of pedestrians in the direction of the vehicle (left)	one unit
The duration of the 1st free window in the pedestrian stream for driving	second
Number of vehicles driven in the 1st window	one unit
The duration of the 2nd free window in the pedestrian flow for driving	second
Number of vehicles driven in the 2nd window	one unit
The duration of the 3rd free window in the pedestrian stream for driving	second
Number of vehicles driving in the 3rd window	one unit
Driving time through the free window in the pedestrian flow, taking into account the distance of 1 meter to the pedestrian crossing and its release	second
Number of vehicles in the queue due to waiting for pedestrians to pass	one unit
t1 – time of movement of the 1st vehicle from the stop line to the beginning of rounding	second
t2 – time of movement of the 1st vehicle in an arc (until the exit from the turn)	second
t3 – the time of movement of the 1st vehicle on the segment of approach to the pedestrian crossing after exiting from the turn	second
t4 – time of leaving the 1st vehicle of the pedestrian crossing, taking into account the distance of 1 meter to the pedestrian crossing and its release	second
Number of vehicles completing the passage to the red signal of the traffic light	one unit
Sampling for the maximum possible number of vehicles driving without pedestrians	one unit
L1 – the distance from the stop line to the border of the intersection with the conflicting direction	meter
L2 – the curvature of the carriageway when turning right	meter
L3 – the distance from the end point of the curvature of the carriageway (intersection border) to the pedestrian crossing when turning right	meter
The actual number of passing cars	one unit

A crossroad is the place of intersection, junction or fork of roads at the same level, limited by imaginary lines connecting the opposite, farthest from the center of the intersection, the beginning of curvature of the carriageway. The geometry of the intersection, especially how the arc is long when turning right, the place of marking (stop lines, crosswalks), largely determines its throughput. Pedestrian exit times and the start of vehicles at regulated intersections are usually the same. This does not take into account the time during which the vehicle travels from the stop line to the pedestrian crossing and the parameters of pedestrian traffic. Studies have shown that pedestrian and car traffic are heterogeneous at different intersections. In our investigation we tried to establish the effect of the length of the arc, the location of the marking and the parameters of pedestrian traffic on the traffic capacity of the intersection. L1 is the distance from the stop line to the border of the intersection with the conflicting direction, m; L2 is the curvature of the carriageway when turning right, m; L3 is the distance from the end point of the curvature of the carriageway (intersection border) to the pedestrian crossing when turning right, m.

Data, such as the duration of the permissive signal of the traffic light, are open data of the work of traffic lights. These data were compared with a video timer. Data L1, L2, L3 are static characteristics of the intersection.

4.2. Data Collection

A feature of the data acquisition system is that it is designed specifically for stationary multi-sensor outdoor cameras. The study was conducted in Chelyabinsk city, Russia. The system works with Intersvyaz company's outdoor cameras (https://www.is74.ru/home/service/#streets_online). The system is a software that works on the basis of neural networks. For recognition of road transport and pedestrians, the system uses the Mask R-CNN, or its variation Faster R-CNN to be exact. The R-CNN mask is a combination of the faster R-CNN that performs object detection (class + bounding box) and FCN (Fully Convolutional Network), which creates a pixel border. The R-CNN mask is conceptually simple: the faster R-CNN has two exits for each candidate object, a class label and an offset bounding box. To this is added a third branch, which displays the object mask — this is a binary mask that indicates the pixels in which the object is in the bounding box. But the additional mask output differs from the output of the class and block, which requires the extraction of a much more accurate spatial location of the object. For this, the R-CNN mask uses the fully collapsed k (FCN) network.

The FCN is a popular algorithm for semantic segmentation. This model uses various convolution blocks and maximum pool layers to reduce the image to 1/32 to its original size. Then this algorithm makes a class prediction at this level of detail. Finally, the algorithm uses the sample and deconvolution layers to resize the image to its original size.

The mask R-CNN unites two networks, namely: the faster R-CNN and the FCN into one big architecture. The loss function for the model is the total loss when performing classification, creating the bounding box, and creating the mask. For the training of the system, a dataset was assembled, in which about one thousand marked images were collected for one intersection. The images were taken at different times of the day, under different weather conditions. This allows us to receive data under any conditions without loss of quality.

Into Faster R-CNN, the image is provided as an input to a convolutional network which provides a convolutional feature map. Instead of using selective search algorithm on the feature map to identify the region proposals, a separate network is used to predict the region proposals. The predicted region proposals are then reshaped using a RoI pooling layer which is then used to classify the image within the proposed region and predict the offset values for the bounding boxes (Gandhi, 2018).

Another feature of the system can be considered that it works with poor quality cameras. Some examples of the images from the road cameras with markup can be seen in Figures 2 and 3.

Examples of images with observation marks are given in Figure 4. These images provided data for the studies conducted in this paper.

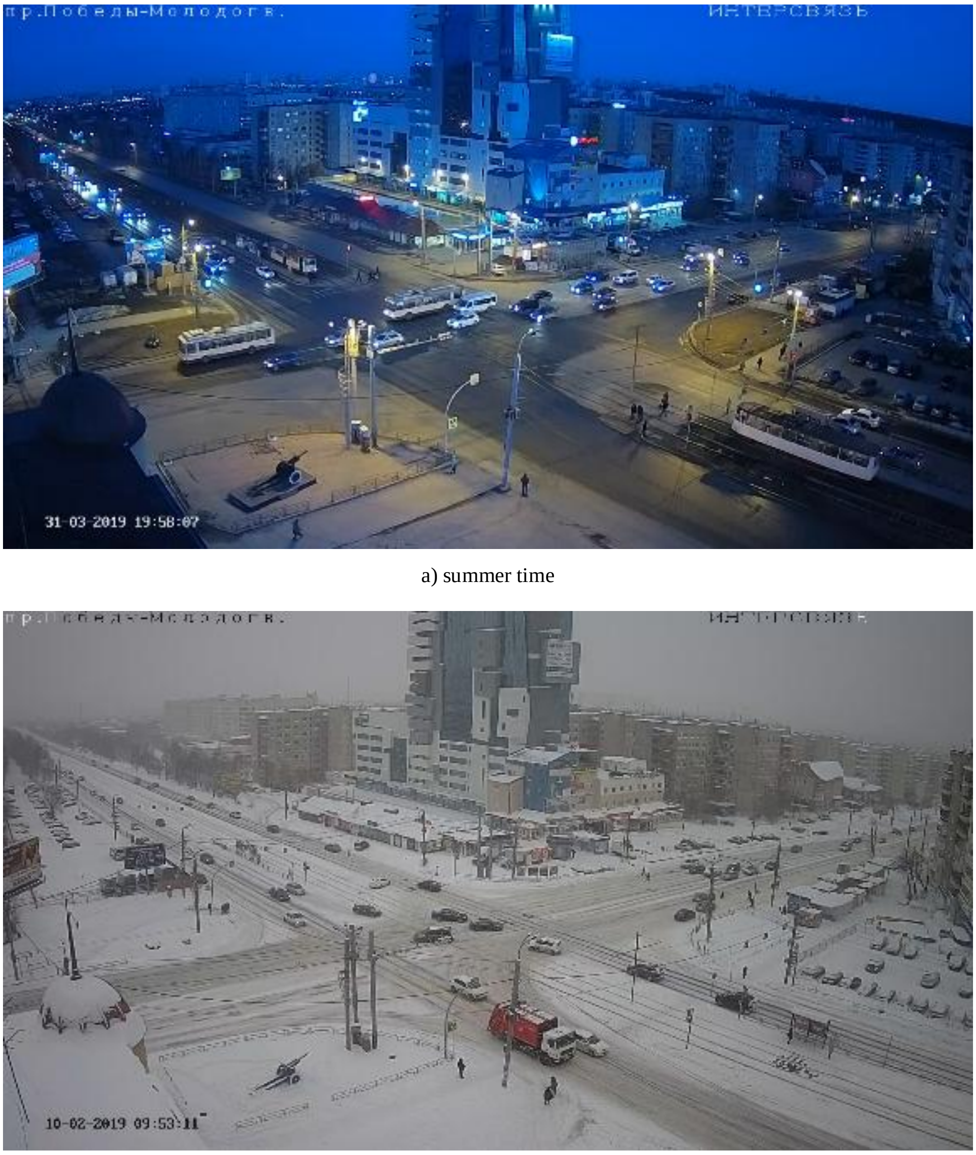
A feature of our information gathering system is that the tags of observations (Figure 4) are placed on the images obtained from video cameras directly, and not from photographs from video surveillance, as it is done in the vast majority of similar studies.

4.3. Statistical Methods

In our paper, we widely use statistical methods for processing and analyzing information obtained from road sensory surveillance cameras for road situations at intersections. These methods characterize the quantitative laws of transport flows in close connection with their qualitative content.

The problems of statistics in our study are most closely related to real life and are associated with the detection of trend characteristics of road traffic at intersections under traffic congestion conditions. In the paper we use such modern statistical analysis methods as: multiple regression analysis, cluster analysis, factor analysis, methods of multidimensional scaling, and others.

The use of statistical methods in our work is due to our desire to show that in the study of traffic flows it is important not only to collect data from video cameras quickly and accurately, but also to be able to process the collected information using appropriate statistical methods. Nowadays, under conditions of the wide distribution of high-performance sensor systems, the collection of information does not present significant difficulties. The foreground is the ability to process the received information properly. We process information using the SPSS computer program. The use of statistical data contributes to familiarization of specialists in transport logistics with the situation on the roads, provides adaptation to changing conditions and making the right management decisions.



a) summer time

b) winter time

Figure 2. Examples of images obtained from road cameras in different seasons: a) summer time; b) winter time

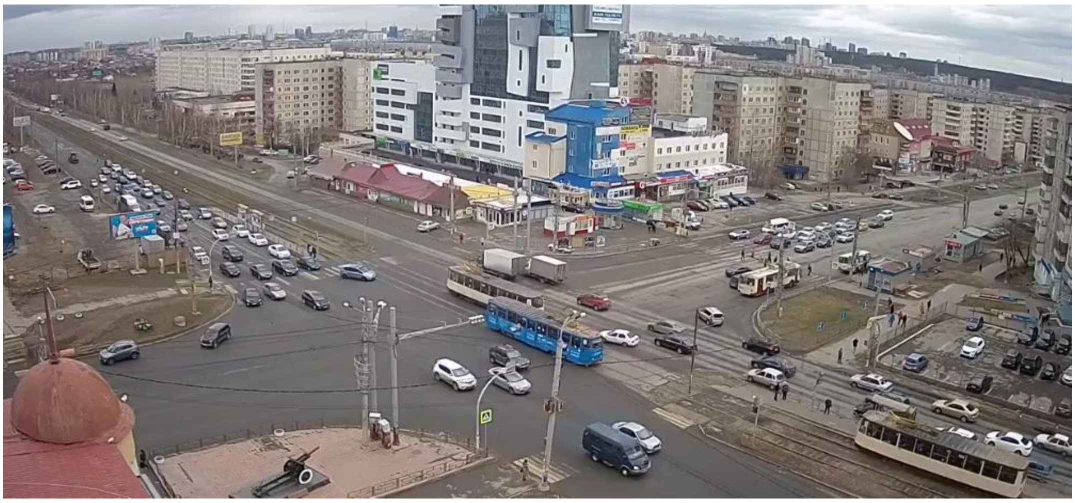
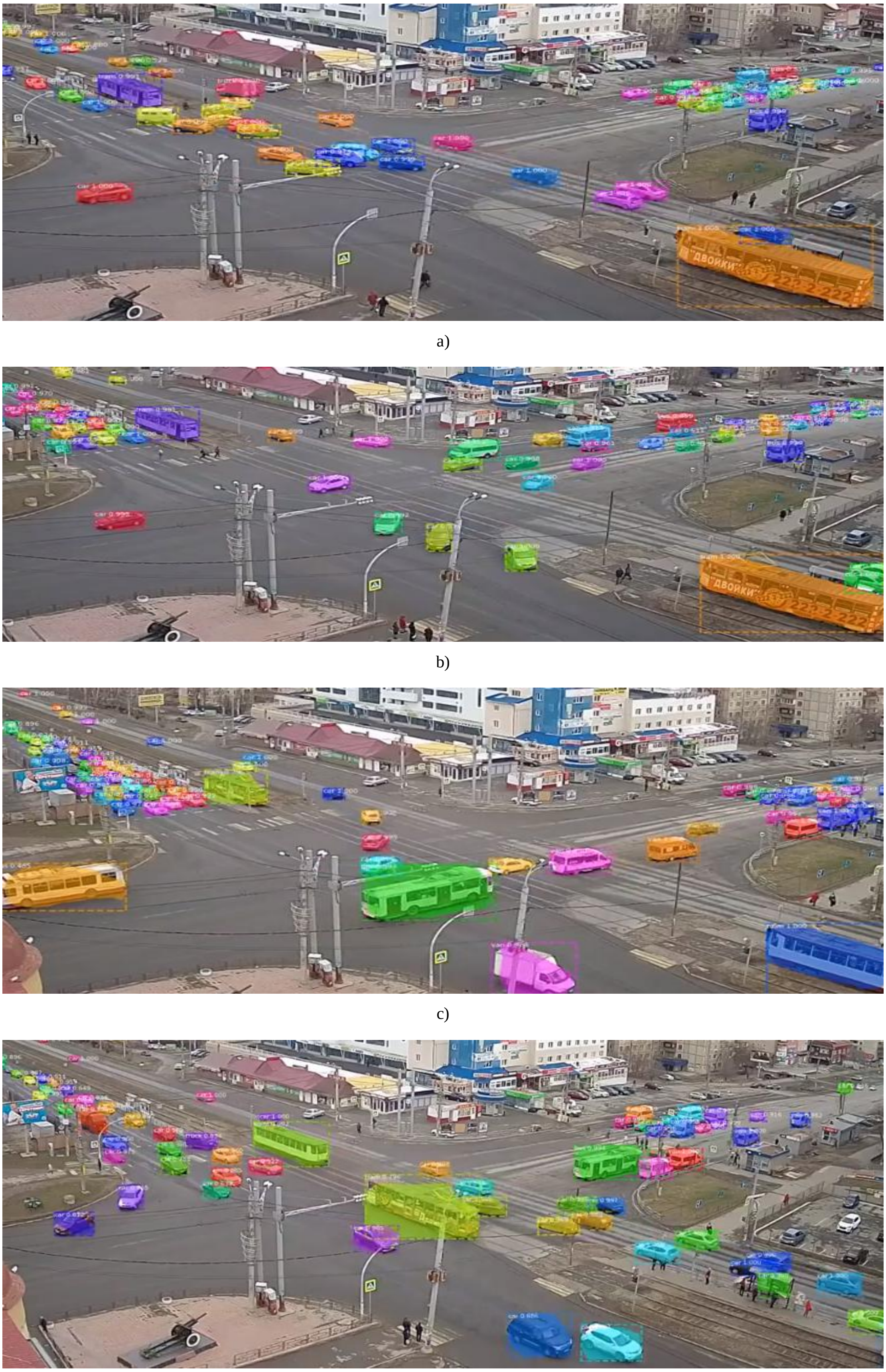


Figure 3. The images obtained from a road camera



a)

b)

c)

d)

Figure 4. Images from road video cameras with observation marks; a)-d) different moments of time

5. Data Post-Processing

In this section, we will carry out statistical processing of the data obtained from the multi-touch video surveillance systems in conditions of traffic congestion. The processing is carried out using the computer program SPSS. The applied statistical methods include factor analysis, construction of multidimensional regression, cluster analysis. For a visual representation of the results of the study we use the method of multidimensional scaling.

5.1. Factor Analysis

Factor analysis allows us to establish for a twenty-one number of source variables a relatively narrow set of "properties" characterizing the relationship between the groups of these variables and called factors. There are various methods of conducting factor analysis. In our study we use the principal component analysis.

In order to obtain a simpler structure for interpretation of the selected factors, which corresponds to a large value of the load of each variable only for one factor and a small one for all other factors, we carry out the procedure of rotating factors. The most popular rotation option is the Varimax orthogonal method (Basto and Pereira, 2012). We use this method in our research.

One of the hardest things to determine when conducting a factor analysis is how many factors (components) to settle on. The scree plot (Figure 5) shows the eigenvalues on the y-axis and the number of factors on the x-axis. It displays the downward curve. The point where the slope of the curve is "leveling off" indicates the number of factors that should be generated by the analysis. Based on the scree plot we can take five factors. The eigenvalues for them is bigger than 1.5. Therefore, the graph of eigenvalues (scree plot), shown in Figure 5, allows us to conclude that five factors can be distinguished.

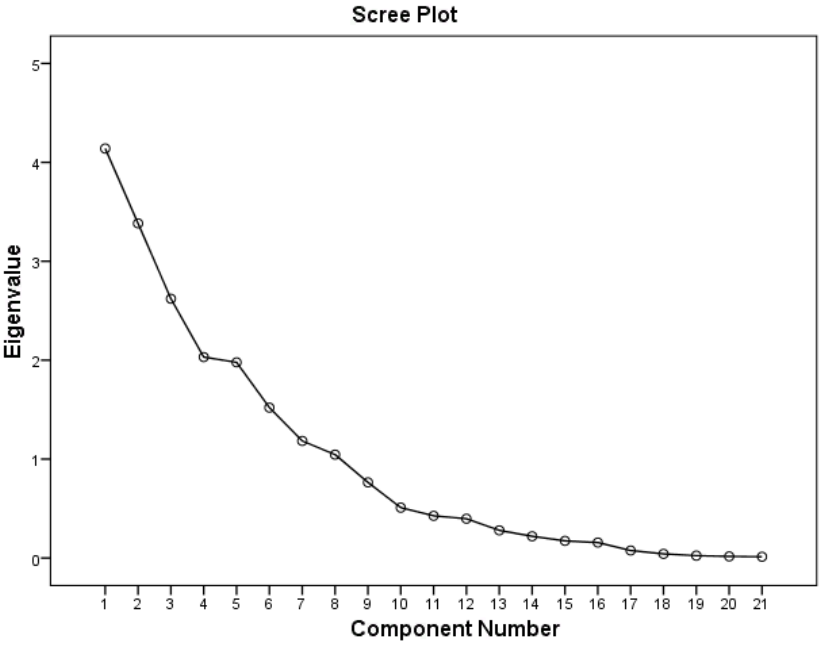


Figure 5. Graph of eigenvalues

The results of the factor analysis in the form of a matrix of rotated components are presented in Table 2.

The performed factor analysis allowed us to switch to latent variables, while reducing the number of variables from twenty one only to five. The sharp decrease in the number of variables will make it possible in the future to greatly facilitate the study of the behavior of traffic flows under traffic congestion.

5.2. Multiple Linear Regression

Multiple regression analysis allows us to select from the totality of the initial variables those that have the most significant impact on the throughput of intersections under traffic congestion conditions. In addition, this analysis makes it possible to rank the selected variables according to the degree of their influence on the throughput of intersections and to quantify the degree of this influence. The multiple regression, constructed as a result of the analysis, makes it possible to predict the throughput of the intersection in terms of specific values of its initial characteristics. It is very important from a practical point of view.

As the dependent variable, we take "The actual number of passing cars," since this variable is the criterion of the intersection capacity. The remaining variables from Table 1 are taken as independent ones.

For the analysis in the package of statistical computer programs SPSS, we choose the option "Multiple linear regression analysis".

Table 2. Rotated Component Matrix

	Component				
	1	2	3	4	5
The duration of the 1st free window in the pedestrian stream for driving	0.881				
t4 – time of leaving the 1st vehicle of the pedestrian crossing, taking into account the distance of 1 meter to the pedestrian crossing and its release	0.684				
Number of vehicles in the queue due to waiting for pedestrians to pass	0.619				
t3 – the time of movement of the 1st vehicle on the segment of approach to the pedestrian crossing after exiting from the turn	-0.593				
Driving time through the free window in the pedestrian flow, taking into account the distance of 1 meter to the pedestrian crossing and its release	0.583				
Number of vehicles driven in the 1st window	0.558				
t1 – time of movement of the 1st vehicle from the stop line to the beginning of rounding	0.466				
The number of pedestrians in the direction of the vehicle (left)		0.803			
The number of pedestrians in the direction of the vehicle (right)		0.779			
L3 – the distance from the end point of the curvature of the carriageway (intersection border) to the pedestrian crossing when turning right		-0.645			
The actual number of passing cars		-0.609			
Number of vehicles driven in the 2nd window			0.870		
The duration of the 2nd free window in the pedestrian flow for driving			0.842		
L1 – the distance from the stop line to the border of the intersection with the conflicting direction			0.709		
Duration of the resolving signal of a traffic light				0.876	
Sampling for the maximum possible number of vehicles driving without pedestrians				0.865	
Number of vehicles completing the passage to the red signal of the traffic light				-0.577	
t2 – time of movement of the 1st vehicle in an arc (until the exit from the turn)				-0.409	
The duration of the 3rd free window in the pedestrian stream for driving					0.870
Number of vehicles driving in the 3rd window					0.762
L2 – the curvature of the carriageway when turning right					-0.469

Table 3 shows the multiple linear regression model summary and overall fit statistics. The coefficient of multiple correlation R reflects the connection of the dependent variable "The actual number of passing cars" with a set of the independent variables and is equal to .958. We find that the adjusted R² of our model is .409 with the coefficient of multiple determination R² = .902. This means that the linear regression explains 90.2% of the variance in the data, which is a very good result. The Durbin-Watson d = 1.533, which is between the two critical values of 1.5 < d < 2.5. Therefore, we can assume that there is no first order linear auto-correlation in our multiple linear regression data.

Table 3. Variable description

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	0.950	0.902	0.409	1.881	1.533

The absence of a negative phenomenon of heteroscedasticity is confirmed by the residual diagram (Figure 6). There is no pattern in the scatter. This scatterplot of standardized residuals against predicted values is a random pattern centered on the line of zero standard residual value. From the scatterplot we can see no clear relationship between the residuals and the predicted values which is consistent with the assumption of linearity. The dispersion of residuals over the predicted value range between –1 and 1 looks constant. The assumption of homoscedasticity has been met.

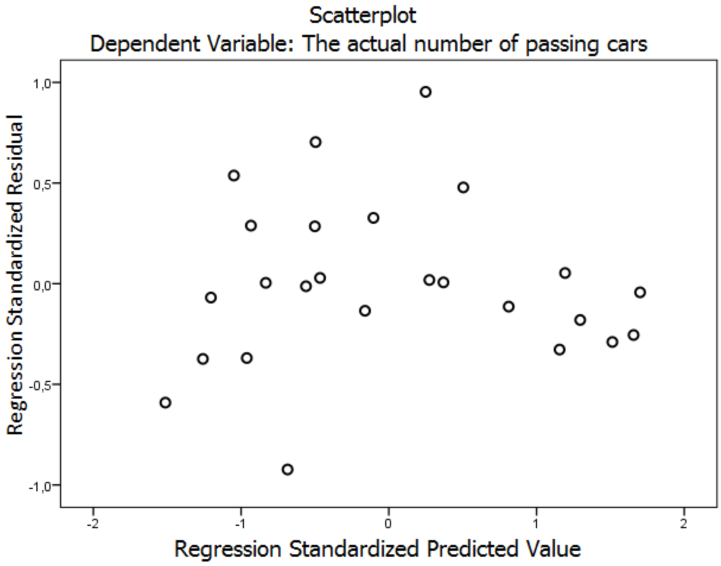


Figure 6. Scatterplot of standardized residuals against predicted values

The absence of pairing correlations with values greater than 0.9 suggests that there is no any multicollinearity. This is also evidenced by the diagnostics of multicollinearity.

The results of the regression analysis are presented in Table 4 and they are ordered by the absolute value of the standardized coefficients.

Table 4. Values of the coefficients of the multiple regression

Coefficients		
	B	Beta
	Unstandardized Coefficients	Standardized Coefficients
(Constant)	0.614	
Duration of the resolving signal of a traffic light	0.303	1.275
Sampling for the maximum possible number of vehicles driving without pedestrians	-0.360	-0.752

Coefficients		
	B	**Beta**
	Unstandardized Coefficients	**Standardized Coefficients**
L2 – the curvature of the carriageway when turning right	0.189	0.573
t4 – time of leaving the 1st vehicle of the pedestrian crossing, taking into account the distance of 1 meter to the pedestrian crossing and its release	0.358	0.522
Number of vehicles driven in the 2nd window	0.461	0.426
The number of pedestrians in the direction of the vehicle (left)	-0.280	-0.417
L1 – the distance from the stop line to the border of the intersection with the conflicting direction	-0.184	-0.396
The duration of the 1st free window in the pedestrian stream for driving	-0.099	-0.268
t2 – time of movement of the 1st vehicle in an arc (until the exit from the turn)	0.242	0.250
Number of vehicles driven in the 1st window	0.272	0.232
The duration of the 3rd free window in the pedestrian stream for driving	-0.392	-0.210
The duration of the 2nd free window in the pedestrian flow for driving	0.102	0.166
The number of pedestrians in the direction of the vehicle (right)	-0.117	-0.159
Number of vehicles driving in the 3rd window	0.531	0.106
Driving time through the free window in the pedestrian flow, taking into account the distance of 1 meter to the pedestrian crossing and its release	0.056	0.098
Number of vehicles in the queue due to waiting for pedestrians to pass	0.099	0.098
L3 – the distance from the end point of the curvature of the carriageway (intersection border) to the pedestrian crossing when turning right	0.049	0.062
t1 – time of movement of the 1st vehicle from the stop line to the beginning of rounding	-0.068	-0.046
t3 – the time of movement of the 1st vehicle on the segment of approach to the pedestrian crossing after exiting from the turn	0.110	0.041
Number of vehicles completing the passage to the red signal of the traffic light	0.013	0.005

With help of the standardized regression coefficients Table 4 allows us to identify the most significant independent variables (factors) that affect the actual number of passing cars. From the table it follows that the variable Duration of the resolving signal of a traffic light has the greatest effect on the dependent variable. Further, in a descending order, the variables follow such as: Sampling for the maximum possible number of vehicles driving without pedestrians, L2 – the curvature of the carriageway when turning right, t4 – time of leaving the 1st vehicle of the pedestrian crossing, taking into account the distance of 1 meter to the pedestrian crossing and its release, Number of vehicles driven in the 2nd window etc. By the ratio of the corresponding standardized coefficients, one can judge the strength of this influence of one variable compared to another one.

In addition, the constructed regression allows us to make predictions for the dependent variable. For example, suppose we have the following set of values for the independent variables:: Duration of the resolving signal of a traffic light = 49 s, L1 = 12, L2 = 15, L3 = 4, The number of pedestrians in the direction of the vehicle (right) = 7, The number of pedestrians in the direction of the vehicle (left) = 8, The duration of the 1st free window in the pedestrian stream for driving = 10 s, Number of vehicles driven in the 1st window = 3, The duration of the 2nd free window in the pedestrian flow for driving = 5 s, Number of vehicles driven in the 2nd window = 1, The duration of the 3rd free window in the pedestrian stream for driving = 2 s, Number of vehicles driving in the 3rd window = 1, Driving time through the free window in the pedestrian flow, taking into account the distance of 1 meter to the pedestrian crossing and its release = 7 s, Number of vehicles in the queue due to waiting for pedestrians to pass = 4, t1 – time of movement of the 1st vehicle from the stop line to the beginning of rounding = 5 s, t2 – time of movement of the 1st vehicle in an arc (until the exit from the turn) = 6 s, t3 – the time of movement of the 1st vehicle on the segment of approach to the pedestrian crossing after exiting from the turn = 2 s, t4 – time of leaving the 1st vehicle of the pedestrian crossing, taking into account the distance of 1 meter to the pedestrian crossing and its release = 12 s, Number of vehicles completing the passage to the red signal of the traffic light = 8, Sampling for the maximum possible number of vehicles driving without pedestrians = 22. Then the value of the dependent variable (The actual number of passing cars) will be 13.

5.3. Hierarchical Cluster Analysis

On the basis of the cluster analysis, we will divide the set of the studied intersections, characterized by a set of features (variables), into groups (clusters) that are homogeneous in the corresponding understanding. In other words, we solve the problem of classifying the intersections under consideration and identifying the corresponding structure in it. In this case, it is assumed that compact removed groups of intersections are separated from each other or that a “natural” partitioning of the aggregate of these intersections into areas of gathering is found.

As a measure of proximity, we take the Squared Euclidean distance. As the clustering method, we choose the Ward's method, which allows forming clusters with minimal dispersion. The number of clusters is determined using the dendrogram (Figure 7), which shows the optimal number of clusters is equal to six.

Of the 25 intersections under consideration, 2 are in the first cluster, 10 in the second, 2 in the third, 5 in the fourth, 3 in the fifth, and 3 in the sixth.

Table 5 shows the average values of the initial variables depending on the cluster, which allows us to characterize each cluster. So, considering the most interesting for us variable. The actual number of passing cars, we can see that its highest average values are observed for the third and the sixth clusters. The following variables L1 and Number of vehicles variables driven in the 2nd window have the higher mean values for these clusters. The variable Number of vehicles completing the passage to the red signal of the traffic light, on the contrary, has the lowest average values for these clusters. Similarly, you can consider the other clusters.

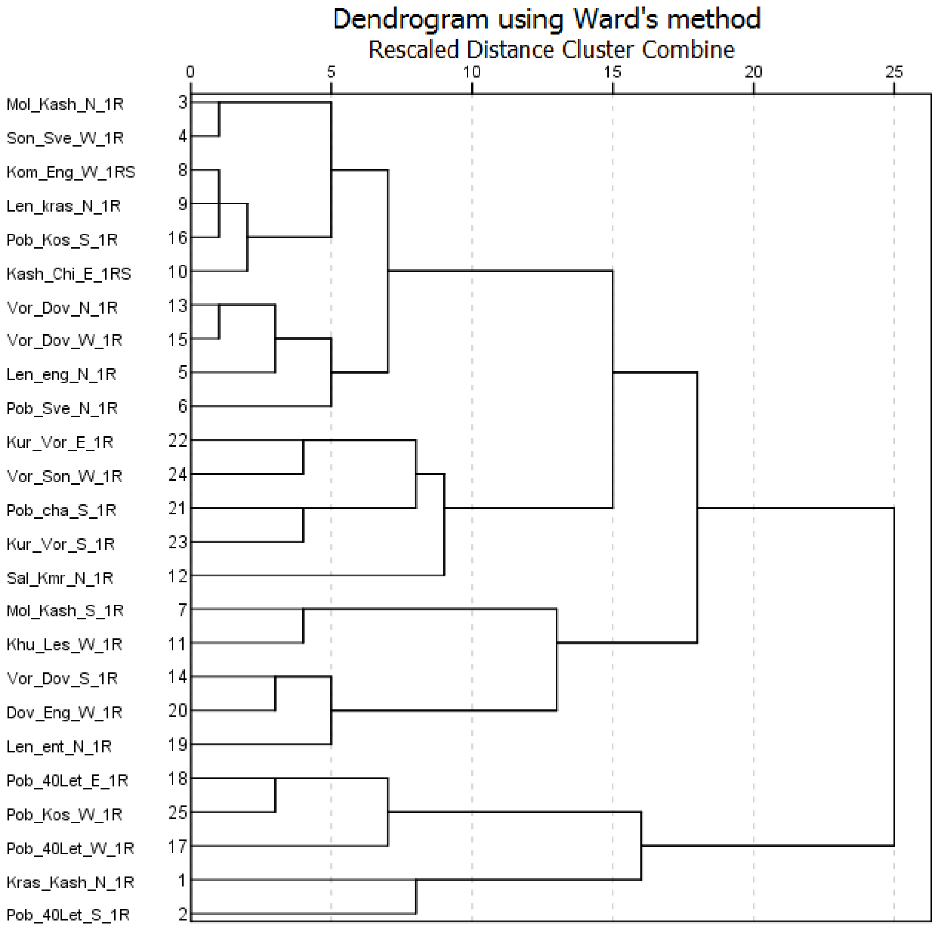


Figure 7. Dendrogram

Table 5. Average values of the source variables by the clusters

Variable	Clusters					
	1	2	3	4	5	6
The duration of the 1st free window in the pedestrian stream for driving	0.00	8.36	13.00	17.02	7.70	0.00
t4 – time of leaving the 1st vehicle of the pedestrian crossing, taking into account the distance of 1 meter to the pedestrian crossing and its release	0.00	4.10	2.15	5.52	4.00	0.00
Number of vehicles in the queue due to waiting for pedestrians to pass	0.15	2.82	6.50	4.20	4.43	3.00
t3 – the time of movement of the 1st vehicle on the segment of approach to the pedestrian crossing after exiting from the turn	2.65	0.40	0.00	0.74	0.00	1.95
Driving time through the free window in the pedestrian flow, taking into account the distance of 1 meter to the pedestrian crossing and its release	0.00	3.28	3.15	8.54	2.33	0.00
Number of vehicles driven in the 1st window	5.65	2.18	4.50	5.72	2.20	2.23
t1 – time of movement of the 1st vehicle from the stop line to the beginning of rounding	3.90	5.04	5.70	5.04	5.33	4.27
The number of pedestrians in the direction of the vehicle (left)	2.65	6.01	3.50	6.12	10.67	3.90
The number of pedestrians in the direction of the vehicle (right)	3.25	6.38	2.00	6.06	5.43	4.50
L3 – the distance from the end point of the curvature of the carriageway (intersection border) to the pedestrian crossing when turning right	4.50	1.00	0.00	2.60	0.00	5.33
Number of vehicles driven in the 2nd window	0.15	0.97	6.65	1.40	2.77	4.67
The duration of the 2nd free window in the pedestrian flow for driving	0.00	1.90	10.80	2.74	8.23	4.00
L1 – the distance from the stop line to the border of the intersection with the conflicting direction	11.00	8.45	18.00	10.00	8.33	12.00
Duration of the resolving signal of a traffic light	19.00	33.50	36.00	48.20	42.67	48.67
Sampling for the maximum possible number of vehicles driving without pedestrians	8.35	17.40	15.50	24.40	21.33	24.00
Number of vehicles completing the passage to the red signal of the traffic light	1.50	1.58	0.80	1.10	1.13	0.00
t2 – time of movement of the 1st vehicle in an arc (until the exit from the turn)	11.25	5.66	3.50	4.64	6.67	7.07
The duration of the 3rd free window in the pedestrian stream for driving	0.00	0.00	0.00	0.40	3.57	0.00
Number of vehicles driving in the 3rd window	0.00	0.00	0.00	0.14	1.10	0.00
L2 – the curvature of the carriageway when turning right	13.50	18.30	16.50	14.00	11.33	19.67
The actual number of passing cars	7.35	7.57	11.00	9.46	7.23	11.43

5.4. Multidimensional Scaling

Multidimensional scaling is used as a tool for visual presentation (visualization) of source data. It allows presenting complex data in a visual form, which facilitates their perception and interpretation in comparison with a tabular form. In our case, the intersections under study are described by twenty-one variables. However, using the ALSCAL multidimensional scaling procedure (https://mondi.web.elte.hu/spssdoku/spsspro.pdf), it is possible to compress the dimension of the original space to two and present the totality of the intersections being studied as points on a plane. In this case, there is some distortion of information, but the methods of multidimensional scaling are designed so that these distortions are minimal. The results of the multidimensional scaling are shown in Figure 8.

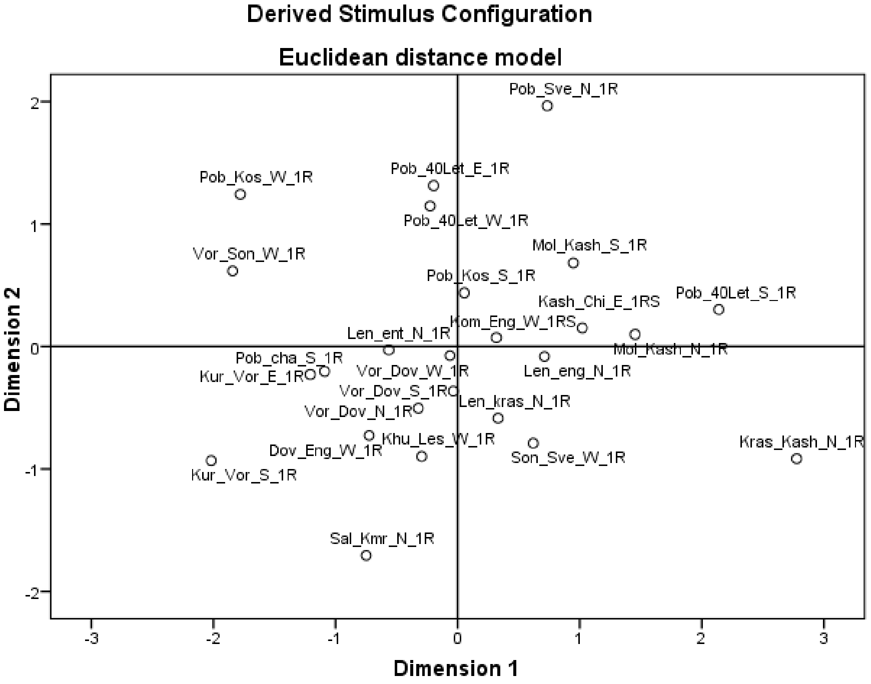


Figure 8. Final configuration of the compared intersections

Proximity of the points in Figure 8 indicates the degree of similarity of the corresponding intersections across the entire set of source variables. For example, the intersections of Pob_40Let_E_1R and Pob_40Let_W_1R, Son_Sve_W_1R and Len_kras_N_1R are quite close in their characteristics. However, Pob_Sve_N_1R, Sal_Kmr_N_1R and Kras_Kash_N_1R intersections are very different from the other ones. Besides, you can see that these intersections are very different from each other. A detailed analysis of the results will allow us to make the right management decisions to improve the transport and logistics infrastructure.

6. Conclusions

In this article, we have developed a data collection system for detecting vehicles in urban traffic monitoring. The system works with outdoor cameras of “Intersvyaz” company and it is a software that works on the basis of neural networks. For recognition of road transport and pedestrians in the system Mask R-CNN is used. The R-CNN Mask is a combination of the faster R-CNN that performs object detection (class + bounding box) and FCN (Fully Convolutional Network), creating a pixel border.

A large number of experiments with data sets shows that our method is superior to some existing algorithms in accuracy and provides real-time detection. The method provides more accurate detection of small vehicles. In addition, in contrast to existing methods, our method tracks not only vehicles, but pedestrians as well. This advantage of our method allows us to drastically reduce the number of potential traffic accidents and improve road safety. The additional mask output is different from the output of the class and block, which requires the extraction of a much more accurate spatial location of the object. For this, the R-CNN mask uses the fully collapsed k (FCN) network.

For the training of the system, a dataset was assembled, in which about 1 000 labeled images were collected for each intersection. The images were taken at different times of the day, under different weather conditions. This allows us to receive data under any conditions without loss of quality. Another feature of the system can be considered that it works with low quality cameras, providing the necessary accuracy of information collection.

The statistical analysis of the collected information using factor, cluster, regression methods and methods of multidimensional scaling made it possible to identify the most important characteristics of the intersections that affect their throughput under traffic congestion conditions. They are the following: Duration of the resolving signal of a traffic light has the greatest effect on the dependent variable, Sampling for the maximum possible number of vehicles driving without pedestrians, The curvature of the carriageway when turning right, Time of leaving the 1st vehicle of the pedestrian crossing, taking into account the distance of 1 meter to the pedestrian crossing and its release, Number of vehicles driven in the 2nd window. The analysis allowed us make predictions of throughput depending on the initial parameters of the intersections, provided the implementation of segmentation of the intersections by initial characteristics and visualization of the results obtained.

In the planned further studies, we assumed using the data from sensory traffic cameras and discriminant analysis to identify the main parameters of traffic at intersections that determine traffic with traffic congestion and without traffic congestion. Such studies will improve the road transport infrastructure of urban networks.

Acknowledgments

The work was supported by the Government of the Russian Federation (Resolution No. 211 of 16 March 2013), contract No. 02.A03.21.0011.

References

1. Zhou, Y., Nejati, H., Do, T., Cheung, N., Cheah, L. (2019) Image-based Vehicle Analysis using Deep Neural Network: A Systematic Study. Available online: https://arxiv.org/pdf/1601.01145.pdf (accessed on 05 May 2019).
2. Biswas, D., Su, H., Wang, C., Blankenship., J., Stevanovic, A. (2017) An Automatic Car Counting System Using OverFeat Framework. *Sensors (Basel)*. 2017 July, 17(7): 1535. Published online 2017 June 30. DOI: 10.3390/s17071535.
3. Zhang, F., Li, C., Yang, F. (2019) Vehicle Detection in Urban Traffic Surveillance Images Based on Convolutional Neural Networks with Feature Concatenation. *Sensors* 2019, 19(3), 594, https://doi.org/10.3390/s19030594.
4. Girshick, R., Donahue, J., Darrell, T., Malik, J. (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2014, pp. 580–587.
5. Uijlings, J.R.R., Sande, K.E.A., Gevers, T., Smeulders, A.W.M. (2013) Selective Search for Object Recognition. *Int. J. Comput.* Vis. 2013, 104, 154–171.
6. Wang, K., Wang, R., Feng, Y., Zhang, H., Huang, Q., Jin, Y., Zhang, Y. (2014) Vehicle recognition in acoustic sensor networks via sparse representation, in *IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*. IEEE, 2014, pp. 1–4.
7. Kim, H., Song, B. (2013) Vehicle recognition based on radar and vision sensor fusion for automatic emergency braking, in *13th International Conference on Control, Automation and Systems (ICCAS)*. IEEE, 2013, pp. 1342–1346.
8. McKay, T., Salvaggio, C., Faulring, J., Salvaggio, F., McKeown, D., Garrett, A., Coleman, D., Koffman, L. (2012) Passive detection of vehicle loading, in *IS&T/SPIE Electronic Imaging. International Society for Optics and Photonics*, 2012, pp. 830511–830511.
9. Mishra, P., Banerjee, B. 2013) Multiple kernel based KNN classifiers for vehicle classification, *International Journal of Computer Applications*, Vol. 71, No. 6, 2013.
10. Tang, T., Thou, S., Dag, Z., Lei, L., Zou, H. (2017) Arbitrary-Oriented Vehicle Detection in Aerial Imagery with Single Convolutional Neural Networks. *Remote Sens.* 2017, *9*(11), 1170, https://doi.org/10.3390/rs9111170.
11. Zhang, J., Huang, M., Jin, X., Li, X. (2017) A Real-Time Chinese Traffic Sign Detection Algorithm Based on Modified YOLOv2. *Algorithms* 2017, *10*, 127.
12. Redmon, J., Divvala, S., Girshick, R., Farhadi, A. (2016) You Only Look Once: Unified, Real-Time Object Detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* Las Vegas, NV, USA, 26 June–1 July 2016.
13. Redmon, J., Farhadi, A. (2016) YOLO9000: Better, Faster, Stronger. *arXiv*, 2016, arXiv:1612.08242v1.
14. Tang, T., Zhou, S., Deng, Z., Zou, H., Lei, L. (2017) Vehicle Detection in Aerial Images Based on Region Convolutional Neural Networks and Hard Negative Example Mining. *Sensors* 2017, 17, 336.
15. Deng, Z., Sun, H., Zhou, S., Zhao, J., Zou, H. (2017) Toward Fast and Accurate Vehicle Detection in Aerial Images Using Coupled Region-Based Convolutional Neural Networks. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens*. 2017, 10, 3652–3664.
16. Qu, T., Zhang, Q., Sun, S. (2017) Vehicle detection from high-resolution aerial images using spatial pyramid pooling-based deep convolutional neural networks. *Multimedia Tools Appl.* 2017, 76, 21651–21663.

17. Buch, N., Velastin, S.A., Orwell, J. (2011) A review of computer vision techniques for the analysis of urban traffic. *IEEE Trans. Intell. Transp. Syst.* 2011, 12:920–939. DOI: 10.1109/TITS.2011.2119372.
18. Daigavane, P.M., Bajaj, P.R. (2010) Real Time Vehicle Detection and Counting Method for Unsupervised Traffic Video on Highways. *Int. J. Comput. Sci. Netw. Secur.* 2010, 10:112–117.
19. Chen, S.C., Shyu, M.L., Zhang, C. (2001) An Intelligent Framework for Spatio-Temporal Vehicle Tracking. *Proceedings of the 4th IEEE Intelligent Transportation Systems*, Oakland, CA, USA. 25–29 August 2001.
20. Gupte, S., Masoud, O., Martin, R.F., Papanikolopoulos, N.P. (2002) Detection and Classification of Vehicles. *IEEE Trans. Intell. Transp. Syst.* 2002, 3:37–47. DOI: 10.1109/6979.994794.
21. Zhang, S., Wen, L., Bian, X., Lei, Z., Li, S.Z. (2018) Single-Shot Refinement Neural Network for Object Detection. *arXiv*, 2018, arXiv:1711.06897.
22. Liu, Y., Wang, R., Shan, S., Chen, X. (2018) Structure Inference Net: Object Detection Using Scene-Level Context and Instance-Level Relationships. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, 18–22 June 2018.
23. Zhou, P. (2018) Scale-Transferrable Object Detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, 18–22 June 2018.
24. Li, S. (2018) 3D-DETNet: a Single Stage Video-Based Vehicle Detector. Computer Science: Computer Vision and Pattern Recognition. *arXiv*, 2018, arXiv:1801.01769.
25. Wang, X., Cheng, P., Liu, X., Uzochukwu. (2018) Focal loss sensitive detectors for vehicle surveillance. In: *2018 International Conference on Intelligent Systems and Computer Vision (ISCV)*. Vol. 2018-May, pp. 1–5.
26. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollar, P. (2017) Focal Loss for Dense Object Detection. In: *Proceedings of the IEEE International Conference on Computer Vision 2017-Octob*, pp. 2999–3007.
27. Hu, X., Xu, X., Xiao, Y., Chen, H., He, S., Qin, J., Heng, P. (2019) A Scale-Insensitive Convolutional Neural Network for Fast Vehicle Detection. *IEEE Transactions on Intelligent Transportation Systems,* 20(3). Mar 2019, pp. 1010–1019.
28. Gandhi, R. (2018) R-CNN, Fast R-CNN, Faster R-CNN, YOLO Object Detection Algorithms. July 9, 2018. https://towardsdatascience.com/r-cnn-fast-r-cnn-faster-r-cnn-yolo-object-detection-algorithms-36d53571365e.
29. Basto, M, Pereira, J. (2012) An SPSS R-Menu for Ordinal Factor Analysis. *Journal of Statistical Software,* 46(4), pp. 1–29.