Archives of
Transport System
Telematics

Volume 5

Issue 1

February 2012

Performance of video detectors working with lossy compressed video streams

W. PAMUŁA
FACULTY OF TRANSPORT, Silesian University of Technology, Krasińskiego 8, 40-019 Katowice, Poland
EMAIL: wieslaw.pamula@polsl.pl

ABSTRACT
Complex traffic control systems are equipped with a range of cameras for traffic surveillance, road traffic measurements. On many sites the different cameras cover the same observation areas but provide different quality streams to the system, usually compressed for surveillance and raw for vehicle detection. Elimination of duplicate cameras especially high quality devices is desired for enhancing the performance of systems. Vehicle detectors based on image processing are sensitive to the quality of input video streams. The paper presents results from tests of using lossy data compression for delivering video streams to vehicle detectors for traffic control. The limit of data loss is determined for assuring correct vehicle detection. The recommendations can be used for optimising traffic vision systems.

KEYWORDS: video detectors, image sequence, lossy data compression, road traffic parameters

1. Introduction

The growing number of cameras in the urban space brings about the issue of exploiting their potential for devising redundant systems performing various monitoring tasks. On many sites different cameras cover the same or similar observation areas, but are used to provide different content or feed data to diverse users. Although multiple streams duplicate information some are altered to enable efficient transfer, which mostly leads to loss of data making them irreversibly distorted. Predominantly video streams are lossy compressed.

Monitoring traffic and more precisely measuring it's parameters is fundamental for traffic control and management systems. Video based measurements provide a wider spectrum of traffic information than traditional road surface mounted detectors eg. vehicles may be tracked along road lanes and through junctions [1],[2].

Another trend in camera deployment is the proliferation of IP based devices with direct Internet connectivity. Network operation due to its universal and widespread nature requires rationalization of transmission bandwidths, thus coercing the implementation of highly efficient video stream compression algorithms [3].

Enhancing the reliability of providing video for measuring traffic becomes a task of integrating compressed streams for feeding video detectors.

In [4] it is shown that standard compression algorithms retain image fidelity in terms of perceptual quality without preserving spectrally significant information for target detection. It is important to identify the sensitive processing steps of the detection algorithms in order to adapt these to fidelity loss. Correlation based target detection algorithms are investigated. The performance may be enhanced by emphasizing middle and high frequency components and discarding low frequency components.

The tracking of objects with correlation based filters, which handle scale distortions, is proved to be much degraded, when using highly compressed video streams [5]. Such filters are suitable for efficient detection of objects using limited computing resources.

The problem of detection and tracking in compressed video is investigated by many researchers in the case of surveillance of human behaviour. There are correspondences to congested traffic flow as traffic scenes are highly changing with occluding objects which is alike to crowd monitoring.

The main contribution of the paper is the discussion of errors introduced by compression artefacts for the performance of vehicle detection algorithms. An investigation programme is proposed for determining the levels of admissible distortion, due to compression losses, for attaining expected detection performance. Conclusions of the discussion enable the determination of recommendations for employing IP based cameras for the use with video detectors.

This paper is divided into five sections. Firstly features of compression artefacts are discussed. In the next section main processing tasks of video detection are outlined with an emphasis on error tolerance. Section four presents the experimental setup and achieved results. The concluding section summarizes the results and sets goals for further examination of the problem of using compressed streams for measuring traffic parameters.

2. Compression

Video monitoring systems use a diverse spectrum of compression methods to reduce the data flow and comply with different demands of the users. Raw streams and losslessly compressed streams are rarely used and mainly for direct feed to video processing devices, because of high bandwidth demands requiring costly wiring.

The prevailing systems make use of methods exploiting interframe block-oriented motion-compensation-based coding such as MPEG1, MPEG2, MPEG4.

Since the introduction of IP based devices the H.264/MPEG4-AVC standard becomes dominating [6]. This is a refined version of MPEG4 with flexible sized blocks and more sophisticated motion compensation. The standard may be enhanced with Context-Based Adaptive Binary Arithmetic Coding (CABAC). Context modelling enables the exploitation of neighbouring pixel characteristics for further streamlining of the coding efficiency.

This version of the coder will be used for investigating the changes of performance of video detectors.

2.1. Compression artefacts

Compression algorithms attain high compression ratios by reducing the information content of the compressed data. The loss of information is manifested by artefacts, which may significantly hinder the functioning of devices processing video.

How much is the functioning degraded depends on the characteristics of artefacts and processing goals. An artefact may render a ghost object or mask an existing one, misleading the monitoring system and reducing the detection quality.

Fig. 1. presents examples of degradation of video content as the compression ratio rises. Artefacts distinctly appear on highly compressed images. The artefacts may be broadly classified into groups describing the way the stream is distorted: ringing, blurring and motion blocking.

Ringing. Ringing artefacts appear as rippling edges around high contrast contours and may also appear as displaced echoes of objects. Some of the temporal fluctuations of pixel values give the impression of flying mosquitoes so called mosquito noise. Fig. 1. e) shows a displaced echo of the side view mirror.

This distortion is the source of ghost objects and thus erratic occupancy of detection fields

Fig.1. Video frames recovered from compressed stream – a), b) original content, c) 150, d) 500, e) 1400 compression ratios

Blurring. The blurring artefacts cause the loss of detail of image objects. The number of distinguishable pixel values is reduced. Objects loose texture features which are used for differentiating them. Fig. 1. c). illustrates this cartoon like effect, the depicted car seems to be made up of flat surfaces.

This artefact may have a positive effect on detection as it filters out noise from the image. On the other hand reducing the resolution of the pixel values may hider operations discerning objects features.

Motion blocking. Motion blocking arises from using GOP (group of pictures) processing. As the compression ratio is raised the size of GOP is enlarged even up to 160 frames. A large GOP means fewer intra coded frames and higher dependence on frames with motion block prediction. In the case of variably moving objects these blocks follow the movement with a significant error. Intensity discontinuities appear on the movement boundaries.

This artefact may be very troublesome for determination of edges of moving objects for finding whether an object entered a detection field.

2.2. Removal of artefacts

Although compression errors are irreversible, efforts are done to recover some of the fidelity loss by artificial reconstruction. A technique based on a forward diffusion function diminishes motion blocking artefacts by exploiting temporal and spatial correlations for smoothing out discontinuities in pixel values [7].

The proposed unsupervised estimation of diffusion parameters requires little computational resources. It can be used prior to compression to improve it by removing spurious noise and less significant features from the original data.

Corrections of artefacts give a visual improvement and in some cases may also give better detection results, but additional computation is required which may be difficult to ensure in real time processing systems.

3. Video based detection

The problem of detecting vehicles in a sequence of images is solved using several approaches. The approaches may be broadly divided into two classes: segmentation based and feature-based methods. The first class covers algorithms that divide frame pixels into sets having similar characteristics, thus representing presumably objects. The division criteria depend on local characteristics of the pixel values and their neighbourhood context.

Feature-based methods use spectral, temporal and spatial characteristics of image patches as features. The features in turn are grouped as objects. The feature based processing may incorporate operations that are common with compression steps.

The idea of extracting moving objects uses the notions of background and foreground representing respectively elements that are permanent fixtures of the observed scene and elements of interest that change locations. Foreground elements constitute sought moving objects.

The detection of moving vehicles in video streams is generally performed by analysis of the differences between the background and the current stream content. This is done by matching object models, extracting and clustering of the features of objects or else by using various filtering methods.

Subtraction of the background from the current contents of the image frame gives moving objects [8]. As the background changes, mostly due to ambient light changes, a reliable model is required to represent the changes.

Techniques, such as: average, median filtering, were the first applied to derive a background. The reference background is the result of filtering a number of consecutive frames [9]. The median filter may be approximated using the sigma-delta filter which significantly reduces memory requirements in implementation solutions.

To cope with complex light changes, pixel intensities are modelled using a mixture of Gaussian distributions [10], [11]. An arbitrarily chosen number of parameterized distributions represent pixel intensity. Each incoming image frame contents updates the parameters of the distributions.

The proposed investigation programme uses video detectors, which principles of operation are based on algorithms relying on background modelling. Two solutions were chosen for investigation. The first solution uses a spatio-temporal model of the background whereas the other a feature based model. Objects are detected by subtracting the background model from the current frame or the objects are detected in feature domain by subtracting feature based representation of the background and of the current frame. Both types of devices determine the presence of objects in defined detection fields, that is the occupancies of virtual detection fields are regarded as signs of objects. The detection fields are configured to be of the size slightly smaller than an average vehicle.

Solutions based on similar processing are prevailing in commercially available devices installed for collecting traffic data for traffic control systems.

Video detector Autoscope RackVision Terra ("auto") developed by Image Sensing Systems [12] and ZIR-WD made by ZIR-SSR Bytom were used for tests [16]. These detectors are commonly used in polish road traffic control systems for providing traffic information to controllers.

3.1. Background modelling

In both tested solutions statistical characteristics of the frame pixels are tracked. Fig. 2. shows an excerpt from the patent description of the method of vehicle detection used in "auto".

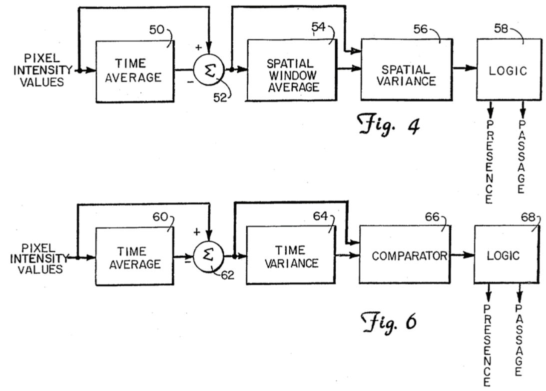

Fig. 2. Block diagram of processing steps for modelling background, spatial models (*Fig.4*), temporal models (*Fig.6*)

Characteristic are blocks for calculating averages and variances of the frame contents. These are used for representing background and moving vehicles. Careful choice of threshold levels enables the determination of the slowly varying background description. Pixels with values of variance smaller than a threshold are classified as background.

In the case of the ZIR-WD video detector the statistics are calculated over image patches $p_{ij}(n)$ (features).

The statistics are calculated using running averages $A_{ij}(n)$ and median approximations $M_{ij}(n)$ [13], [14]:

$$A_j\,(n) = A_j\,(n-1) + \frac{p_j\,(n) - A_j\,(n-1)}{k+1}, \qquad (1)$$
$$M_j\,(n) = M_j\,(n-1) + \operatorname{sgn}(p_j\,(n) - M_j\,(n-1)),$$

p_{ij} – mean value of pixels in a patch *ij* on frame *n*,
k – averaging period (number of passed frames).

Patches of the size 3x3 pixels are used. Each pixel surrounded by four patches is described by attributes, which indicate the relations between values of the pixel and patches. Similarly as in "auto" slowly changing values are attributed to the background. Additionally statistics are not updated when temporal changes in the detection fields are less than a "movement" threshold.

3.2. Detection of objects

In the first investigated solution a careful choice of threshold levels enables the differentiation of highly changing features of moving objects and slowly varying background description. Objects are detected when the sum of spatial and temporal variances of pixels in a detection field exceed these thresholds, which are continuously updated to track ambient changes.

In ZIR-WD each image patch has 5 attributes expressed using one of three values (black, grey, white) [15]. One attribute denotes colour, four other denote a contrast measure describing the difference of the centre pixel to the mean values of patch pixels in the neighbourhood.

Contrast attribute values are determined using thresholding. The thresholds are determined using the histogram of the image frame and running averages of image patches. The histogram is calculated over a number of past frames to avoid sudden changes due to dazzling or other high speed light phenomena.

Attribute values are accumulated for detection fields for each frame and only when the field previously was marked as empty. The values are compared to the running averages and when they exceed the detection field is marked as occupied. A comparison with hysteresis is applied to avoid erratic behaviour when the attribute values frequently fluctuate.

The main operations of the detection algorithms are calculation of statistics of pixel values. These are:

- computing average values in a frame over a number of consecutive frames – temporal average;
- computing average values: in a window, in a patch – spatial average;
- computing variance values in a frame over a number of consecutive frames;
- computing spatial variance in a window.

Crucial for the determination of occupancy of detection fields are threshold based operations.

The operations of computing spatial averages and variances are sensitive to blurring artefacts. Operations processing consecutive frame content are sensitive to ringing and motion blocking.

4. Detection performance

The detection performance was evaluated using a database of road traffic video streams. Traffic data was collected during different weather condition at one camera site. Traffic was registered on daily basis and encompassed morning and afternoon peaks as well as evening lows and midday plateau.

The nominal length of the films was 11 hours. This way of registering was chosen to facilitate the functioning of video detectors as close to real world as possible. Vehicle detection algorithms model backgrounds using up to several thousand frames. Long video streams eliminate the necessity of artificially initialising background models.

The database contains films registered:

- in sunny weather, with dazzling effects,
- during rain, containing road surface reflections,
- in cloudy conditions, which are regarded as standard observation conditions – contain no shadow effects, no significant ambient light changes.

This vast database of traffic material was reviewed in search for films containing sequences potentially generating compression artefacts. As pointed out in section 2 ringing is evoked by high contrast edges of objects, blurring is visible for small objects and motion blocking is acute when the movement dynamics is high.

These characteristics may be ascribed to forming and discharging of vehicle queues in changing ambient light conditions. A representative film registered on a rainy day was chosen. The film contains a few tens of queue formations and satisfies the changing light condition.

Important problems were the choice of a performance index and the elaboration of a way to compare performance. The common function of video detector solutions is the capability to register the occupancy of defined detection fields (zones). This function gives the schedule of vehicle arrivals in the fields.

The self evident way of performance evaluation is the comparison of registered schedules of arrivals. If arrivals can be superimposed on a reference dataset with no errors such a detection is perfect. This approach is not suitable in practice, because the moment of arrival can randomly fall within a range of few frames. Slightly late or early arrivals do not par with the reference dataset. A modification of the reference by introducing time periods of arrivals complicates the preparation of the reference dataset.

Instead of comparing vehicle arrivals it is proposed to compare traffic flow intensity. Traffic flow intensity $Q(n)$ for consecutive measurement moments n is calculated in a way to retain dynamics features for traffic control. This assumption requires the limitation of the vehicle counting period T to a size which complies with the traffic control cycles lengths as:

$$Q(n) = \frac{k}{T} \sum_{i=n}^{n+T} O(i) \qquad (2)$$

where: lengths are expressed in the number of video frames, 40 ms elapses between frames, T – vehicle counting period, s – measurement step, $O(i)$ – occupation index at frame i, k – conversion factor used to express Q in units of vehicles per hour.

Traffic control cycles at the measurement site vary in the range of 60 to 120 seconds during the day. The vehicle counting period T was set to 300 sec. (3750 frames), whereas the measurement step s was set to 10 sec. (125 frames). A 10 sec. step spans at most one vehicle arrival so the tracking error for mean flows of about 360 veh/hour is about 3% for the counting period.

Vehicles were detected in a detection field configured at the stop line of the traffic junction being monitored. Fig. 3 shows the setup for ZIR-WD with the overlaid setup for "auto". The size of the detection field is defined, in a standard way, as slightly smaller than an average passenger car. In the case of "auto" it is declared as the sum of occupancy of three strips covering the corresponding area of the detection field.

Experiments of vehicle detections were carried out in parallel with both detectors. A PC based video player was the source of video. The video stream was duplicated using two optically separated buffer amplifiers. This solution provides identical streams and prevents interactions between devices using them. The tests were repeated several times to assure reliable results.

Fig. 3. Camera view with a marked detection field and "auto" detection stripes

Detection results were recorded using the detectors build in logging functionality. In the case of "auto" the log is a text file which contains occupancy indices of strips with millisecond tick marks indicating the moment of vehicles entry and exit, to and from the detection strips. The conversion to data suitable for processing requires a simple data filtering using a spreadsheet.

ZIR-WD detector has a binary file with detection results. Each occupancy entry contains a number of auxiliary variables such as: the level of occupancy, values of logical functions resolved with occupancy indices of the detection fields. No time marks are noted. The ordinal of the entry in the file indicates time in 80 millisecond increments. This file is directly used by the traffic flow calculation programme.

Fig. 4. shows graphs of traffic flow intensity calculated using raw video data. This is only a part of the processed occupation data, it covers a period of 10000 seconds of traffic, roughly a part of the morning traffic peak.

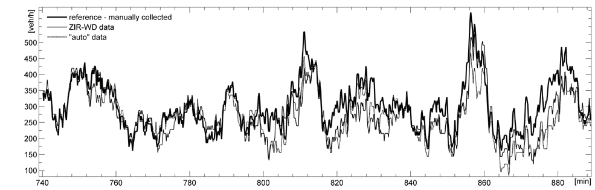

Fig. 4. Traffic flow intensity measured using uncompressed video stream

Traffic flow values calculated for both detectors coincide with reference data for periods of calm traffic. Traffic congestion in periods of 815-840 min and past 860 min causes large differences in measurements. Due to high density of vehicles the detection fields are constantly occupied preventing correct vehicle detection, which is manifested by lower flow values.

A preliminary study of the detection degradation has shown that for a gradual increase of compression the degradation rate is consistent with the increase and rises monotonically but very slowly. Taking this result into account it was proposed to carry out tests for exponentially changing compression rates.

The standard MPEG4/AVC does not use the notion compression rate but expresses the degree of compression as the bitrate of the compressed stream. The bitrate directly indicates the required bandwidth for transmission of the compressed stream.

The series of bitrates starting from 4 Mb/s was chosen for tests. These are: 4 Mb/s, 2Mb/s, 1Mb/s, 500 kb/s, 250 kb/s, 150 kb/s, 100 kb/s and 50 kb/s. At the low end the bitrates were chosen more densely as visual inspection of the film contents has shown significant distortions.

The set of 8 compressed versions of the chosen film were prepared. This series of films was put on the playlist of the video player and the tests were started in the same manner as detection tests using raw video.

The collected data was processed using Matlab which enabled an efficient analysis of the detection performance. ZIR-WD detector stores video data with detection marks assigned to frame numbers. This file was used to locate different artefacts distorting the detection.

Sixteen data sets of traffic flow intensities corresponding to the set of bitrates were prepared on the basis of the logging files from the detectors. These results constitute the database for analysis of detection errors. Matching reference data with these results gives an approximate classification of the detection performance.

High bitrates 4 Mb/s, 2 Mb/s, 1 Mb/s and 500 kb/s give results similar to raw video based traffic flow data. Inspection of the streams shows that the discrepancies may be accounted for blurring and ringing artefacts.

At low bitrates movement blocking effects dominate.

4.1. Deterioration due to blurring

Blurring and ringing artefacts cancel noise in the video stream and enhance edges of objects with high contrast, relative to background, contours. Noise reduction helps the segmentation of the frame content.

Calculation of patch attributes in the case of ZIR-WD and window statistics in "auto" remains stable thus the objects are correctly detected.

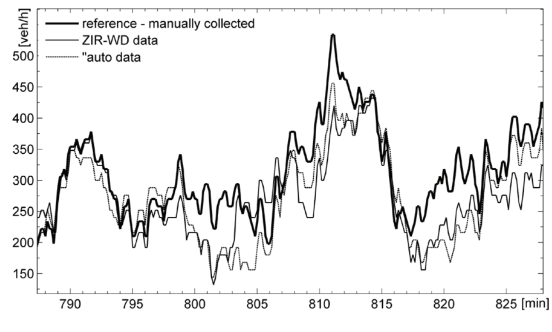

Fig. 5. Traffic flow intensity at 4 Mb/s

Fig. 5 shows traffic flow graphs obtained from both detectors that almost cover each other. This signifies that distinguishing vehicles in a detection field is of comparable accuracy. The traffic flow values are lower although the shapes of the graphs follow the reference.

As noted in the case of raw video tests, the highest errors coincide with congestion periods especially with forming of queues. Blurring artefacts prevent the correct detection of vehicles with less distinct contours eg. dark vehicles.

4.2. Deterioration due to blocking

Bitrates lower than 500 kb/s bring about a considerable change in the appearance of traffic flow graphs. Again discrepancies are distinct in times of congestion flow.

As in fig. 6 “auto” graphs lie much lower than the reference whereas ZIR-WD appear to match closer the reference data.

This behaviour may be accounted mostly for motion blocking which divides the frame into patches related to objects movement. Together with increasing blurring this artefact diminishes the resolution of pixel values making the threshold operations more sensitive. Detection algorithm used in “auto” relies to a large extent on the accuracy of thresholds, so there are more errors in distinguishing objects from the background.

Determination of pixel attributes using statistics of patches in its neighbourhood works with wide margins. Blocking introduces local discontinuities which helps in keeping these.

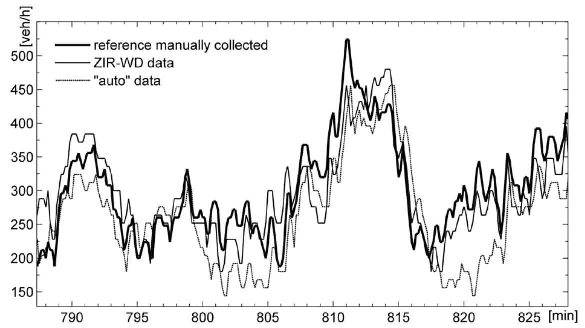

Fig. 6. Traffic flow intensity at 100 kb/s

The use of attributes for the description of pixels proves robust to the loss of fidelity due to compression. A slight improvement of detection rate is observed for bitrates from 500kb/s down to 100 kb/s.

4.3. Error rates

The comparison of traffic flow graphs for different bitrates shows variable behaviour highly influenced by such traffic phenomena as: queuing, congestion and queue discharges. Related to these is the changing character of video content bringing about artefacts when compressed.

A measure which averages the differences between the reference data and the calculated flow based on video processing is proposed. This measure – an error rate is expressed as:

$$E = \frac{1}{N}\sum_{i=1}^{N}\frac{|Q(i) - Q_R(i)|}{Q_R(i)} \qquad (3)$$

where: $Q(i)$, $Q_R(i)$ – traffic flow intensities calculated, reference, N – number of calculated samples of Q.

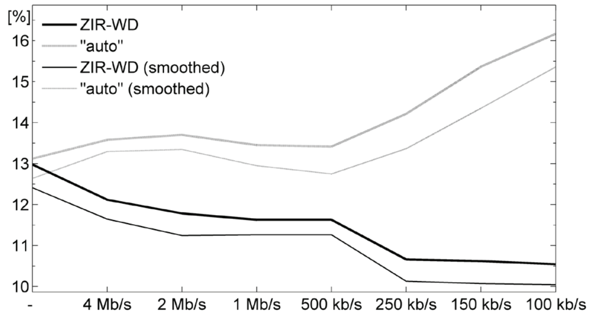

Fig. 7. Error rates

Error rates were computed for all collected results. Additionally results were smoothed by averaging 3 consecutive values to achieve 15 minute counting periods usually used for managing traffic in networks.

Fig. 7 shows the graphs of error rates. Graphs for ZIR-WD and “auto” run in parallel till the bitrate 500 kb/s. Lower bit rates give diverged graphs. The values of errors for 50 kb/s are omitted as they were tested only once.

5. Conclusion

Lossy compression deteriorates the performance of video detection algorithms based on tracking mean values and variances of pixels for the determination of changes of frame contents. Blurring distorts the values of the statistic measures lowering the margins for correct detection of objects. Ringing and especially motion blocking at high compression ratios bring about errors in determination of thresholds which become sensitive to the intensity of traffic flow.

Algorithms utilising object features are more robust to compression artefacts. The use of compressed streams surprisingly gives a reduction of the detection errors with growing compression ratios.

Both types of detection algorithms have a near constant error rate, comparable to raw stream processing error, till the 500 kb/s rate. This means that compression ratios up to about 140 times do not significantly influence the performance of the algorithms.

The results prove that IP cameras providing MPEG4/AVC compressed video streams are suitable for feeding video to video detectors commonly used in traffic control applications.

Investigation of the use of compressed video for tracking objects is proposed as a follow up to the current research. Devices for tracking vehicles are being introduced to traffic control practice to cope with tracking complex traffic flow at junctions. The prepared measurement site and acquired experience during tests constitutes a good foundation to carry out such an investigation.

Bibliography

[1] WU Y., SHEN J., DAI M.: Traffic object detections and its action analysis, Pattern Recognition Letters, vol. 26, pp. 1963-1984, 2005.

[2] CHITTURI M. V., MEDINA J.C., BENEKOHAL R.F.: Effect of shadows and time of day on performance of video detection systems at signalized intersections, Transportation Research Part C 18, pp. 176–186, 2010.

[3] CHEN M., FOWLER M.L., ZHANG S.: Standards-compatible compression for automated image recognition in sensor networks and surveillance systems, Proc. SPIE vol. 5208, pp. 46-57, 2004.

[4] KEREKES R. A., KUMARA V., SIMS S. R.: Estimating performance limits for automatic target recognition in compressed video, Proc. SPIE vol. 5807, pp. 296-307, 2005.

[5] BUGEAU A., PÉREZ P.: Detection and segmentation of moving objects in complex scenes, Computer Vision and Image Understanding vol. 113 pp. 459-476, 2009.

[6] ITU-T Rec. H.264 | ISO/IEC 14496-10 AVC 2004

[7] BOURDON P., AUGEREAU B., OLIVIER C. CHATELLIER C.: MPEG-4 compression artefacts removal on colour video sequences using 3d nonlinear diffusion. Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP ,04), pp.729-733, 2004

[8] PICCARDI M.: Background subtraction techniques: A review, in Proc. IEEE Int. Conf. Syst., Man, Cybern., vol. 4, pp. 3099-3104, 2004.

[9] LI L., HUANG W., GU I.Y.H., TIAN Q.: Statistical modeling of complex backgrounds for foreground object detection, IEEE Trans. on Image Processing, vol. 13 pp. 1459-1472, 2004.

[10] VARGAS M., MILLA J. M., TORAL S. L., BARRERO F.: An Enhanced Background Estimation Algorithm for Vehicle Detection in Urban Traffic Scenes. IEEE Transactions On Vehicular Technology, Vol. 59, No. 8, pp. 3694-3709, 2010.

[11] STAUFFER C., GRIMSON W.: Learning patterns of activity using realtime tracking," IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 8, pp. 747-757, 2000.

[12] MICHALOPOULOS P.G., FUNDAKOWSKI R.A., FITCH R.C.: Vehicle detection through image processing for traffic surveillance and control. US Patent no. 4847772 Jul. 11. 1989.

[13] ZIR-WD videodetector. Project Report: Modules of Video Traffic Incidents Detectors ZIR-WD for Road Traffic Control and Surveillance. WKP-1/1.4.1/1/2005/14/14/231/2005, vol.1-6 Katowice, Poland 2007.

[14] PAMUŁA W.: Vehicle Detection Algorithm for FPGA Based Implementation, in Computer Recognition Systems Eds. M. Kurzyński, M. Woźniak, Springer Verlag Berlin, pp. 586-592, 2009.

[15] PLACZEK, B.: A Real Time Vehicle Detection Algorithm for Vision-Based Sensors. Bolc L. et al, (Eds.) Computer Vision and Graphics, LNCS 6375. Springer-Verlag, pp. 211-218, 2010.

[16] Autoscope RackVision Terra datasheet. Image Sensing Systems Inc USA, 2009.