

# Quality assessment for a visual and automatic license plate recognition

Lucjan Janowski · Piotr Kozłowski · Remigiusz Baran ·  
Piotr Romaniak · Andrzej Glowacz · Tomasz Rusc

Published online: 30 August 2012

© The Author(s) 2012. This article is published with open access at Springerlink.com

**Abstract** Video transmission and analysis is often utilized in applications outside of the entertainment sector, and generally speaking this class of video is used to perform specific tasks. Examples of these applications include security and public safety. The Quality of Experience (QoE) concept for video content used for entertainment differs significantly from the QoE of surveillance video used for recognition tasks. This is because, in the latter case, the subjective satisfaction of the user depends on achieving a given functionality. Recognizing the growing importance of video in delivering a range of public safety services, we focused on developing critical quality thresholds in license plate recognition tasks based on videos streamed in constrained networking conditions. Since the number of surveillance cameras is still growing it is obvious that automatic systems will be used to do the tasks. Therefore, the presented research includes also analysis of automatic recognition algorithms.

**Keywords** Quality of experience · License plate recognition ·  
Automatic plate recognition · Compression influence

---

L. Janowski (✉) · P. Kozłowski · A. Glowacz  
Department of Telecommunications, AGH University of Science and Technology,  
al. Mickiewicza 30, 30-059 Krakow, Poland  
e-mail: janowski@kt.agh.edu.pl

R. Baran  
Faculty of Electrical Engineering, Automatics and Computer Science,  
Kielce University of Technology, al. 1000-lecia P.P. 7, 25-314 Kielce, Poland

P. Romaniak  
Academic Computer Centre CYFRONET AGH, AGH University of Science and Technology,  
ul. Nawojki 11, 30-950 Kraków, Poland

T. Rusc  
Institute of Physics, Jan Kochanowski University, ul. Swietokrzyska 15,  
25-406 Kielce, Poland

## 1 Introduction

The transmission of video is often used for various applications outside of the entertainment sector, and generally this class of video is used to perform specific tasks. Examples of these applications are security, public safety, remote command and control, and sign language. Monitoring of public urban areas (traffic, intersections, mass events, stations, airports, etc.) for safety threats using transmission of video content has become increasingly important because of a general increase in crime and acts of terrorism (e.g. attacks on the World Trade Center and the public transportation systems in London and Madrid). Nevertheless, video surveillance is also viewed with concern by numerous civil right organizations, which see foremost the need for the protection of citizens against Orwellian-like “permanent surveillance”. Among these, we should mention the Liberty Group (dedicated to human rights), an Open Europe organization, the Electronic Frontier Foundation, and the Ethics Board of the FP7-SEC INDECT (INDECT is intelligent information system supporting the observation, search and detection of suspicious or criminal activity in order to protect the security of citizens in an urban environment) [9]. This matter was also one of the main themes (“Citizens Security Needs Versus Citizens Integrity”) of the Fourth Security Research Conference organized by the European Commission (September 2009) [11]. Despite this, many studies suggest that public opinion about CCTV is becoming more favorable [14]. This trend intensified after September 11, 2001. Furthermore, methods do exist that partially protect privacy. They are based on the selective monitoring of privacy information like faces or license plates. Than automatic erasing of faces/license plates not related to the investigation is possible. One of possible technique to achieve privacy protection is private part of image hiding using digital watermarking.

Protecting privacy is important aspect of CCTV systems but we need a working system i.e. system which makes it possible, for authorized person, to access the full signal. Unfortunately in some cases the signal produced by a camera is “protected” from seeing by anyone. Such situation is caused by artefacts. Anyone who has experienced artefacts or freezing play while watching an action movie on TV or at a live sporting event, knows the frustration accompanying sudden quality degradation at a key moment. However, for practitioners in the field of public safety the usage of video services with blurred images can result in much more severe consequences. The above-mentioned facts convince us that it is necessary to ensure adequate quality of the video. The term “adequate” quality means quality good enough to recognize objects such as faces or cars.

In this paper, recognizing the growing importance of video in delivering a range of public safety services, we have attempted to develop critical quality thresholds in license plate recognition tasks, based on video streamed in constrained networking conditions. The measures that we have been developing for this kind of task-based video provide specifications and standards that will assist users of task-based video to determine the technology that will successfully allow them to perform the function required.

Even if license plate algorithms are well known, practitioners commonly require independent and stable evidence of the accuracy of human or automatic recognition for a given circumstance. Re-utilization of the concept of QoE (Quality of

Experience) for video content used for entertainment is not an option as this QoE differs considerably from the QoE of surveillance video used for recognition tasks. This is because, in the latter case, the subjective satisfaction of the user depends on achieving a given functionality (event detection, object recognition). Additionally, the quality of surveillance video used by a human observer is considerably different from the objective video quality used in computer processing (Computer Vision). In the area of entertainment video, a great deal of research has been performed on the parameters of the contents that are the most effective for perceptual quality. These parameters form a framework in which predictors can be created, so objective measurements can be developed through the use of subjective testing. For task-based videos, we contribute to a different framework that must be created, appropriate to the function of the video—i.e. its use for recognition tasks, not entertainment. Our methods have been developed to measure the usefulness of degraded quality video, not its entertainment value. Assessment principles for task-based video quality are a relatively new field. Solutions developed so far have been limited mainly to optimizing the network QoS parameters, which are an important factor for QoE of low quality video streaming services (e.g. mobile streaming, Internet streaming) [22]. Usually, classical quality models, like the PSNR [10] or SSIM [26], were applied. Other approaches are just emerging [12, 13, 25].

The remainder of this paper is structured as follows. Section 2 presents a license plate recognition test-plan. Section 3 describes source video sequences. In Section 4, we present the processing of video sequences and a Web test interface that was used for psycho-physiological experiments. Section 5 offers a short description of the automatic algorithms used in the experiments. In Section 6, we analyze the results obtained from the human experiment. In Section 7 we analyzed results obtained for the automatic systems. Section 8 draws conclusions and plans for further work, including standardization.

## 2 Licence plate recognition test-plan

The purpose of the tests was to analyze people's ability to recognize car registration numbers on video material recorded using a CCTV camera and compressed with the H.264/AVC codec. In order to perform the analysis, we carried out a subjective experiment. The intended outcome of this experiment was to be data regarding human recognition capabilities. Non-expert testers rated video sequences influenced by different compression parameters. The video sequences used in the test were recorded at a parking lot using a CCTV camera. We adjusted the video compression parameters in order to cover the recognition ability threshold.

The recognition task was threefold: 1) type in the license plate number, 2) select car color, and 3) select car brand. Testers were allowed to control playback and enter full screen mode. The experiment was performed using diverse display equipment. Each subject scored 30 sequences each of a different source scene. The subjects were not paid. The experiment was limited since we had only 30 different source scene. We did not want to repeat the same source scene in order to be sure that subjects did not memorize any license plates from any prior viewings of the same source.

The tests were conducted using a web-based interface connected to a database. In the database both information about the video samples and the answers received from the testers were gathered. The interface is presented in Fig. 1.

A common problem of CCTV systems is continuous observation. It is impossible to watch feeds from all available CCTV cameras in metropolitan area, e.g. Chicago, where the police CCTV system has 17 thousand cameras. Therefore, automatic systems are needed. Having already planned the subjective experiment, we decided to use exactly the same sequences to check the performance of two different automatic systems. As we expected the quality needed for an automatic system is much higher than is needed for humans. The automatic systems were not build to work the best for this particular compression algorithm. They are developed to recognize number plates in general.

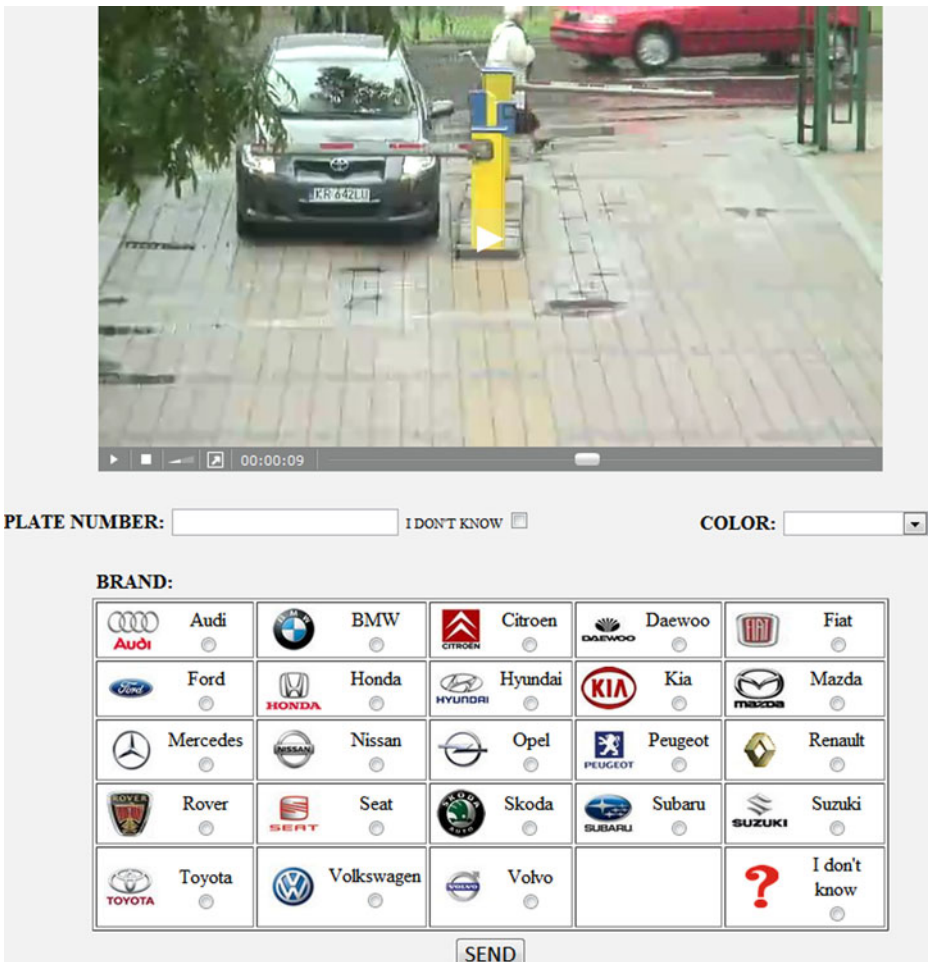


Fig. 1 Test interface

### 3 Source video sequences

Source video sequences were collected at AGH University of Science and Technology by filming a car parking lot during a time of high traffic volume. In this scenario, the camera was located 50 m from the parking lot entrance in order to simulate typical video recordings. Using ten-fold optical zoom, a 6 m by 3.5 m field of view was obtained. The camera was placed statically without changing the zoom throughout the recording time, which reduced global movement to a minimum.

Acquisition of video sequences was conducted using a two mega-pixel camera with a CMOS sensor. The recorded material was stored on an SDHC memory card inside the camera.

All the video content collected in the camera was analyzed and cut into 20 second shots including cars entering or leaving the car park. Statistically, a license plate was visible for 17 s on average in each sequence. The length of the sequences was dictated mostly by the need to capture the vehicles not only when they were stopped at the entrance barrier but also in motion. The parameters of each source sequence are as follows:

- resolution: 1280 by 720 pixels (720p)
- frame rate: 25 frames/s
- average bit-rate: 5.6 – 10.0 Mb/s (depending on the local motion amount)
- video compression: H.264/AVC in Matroska Multimedia Container (MKV)

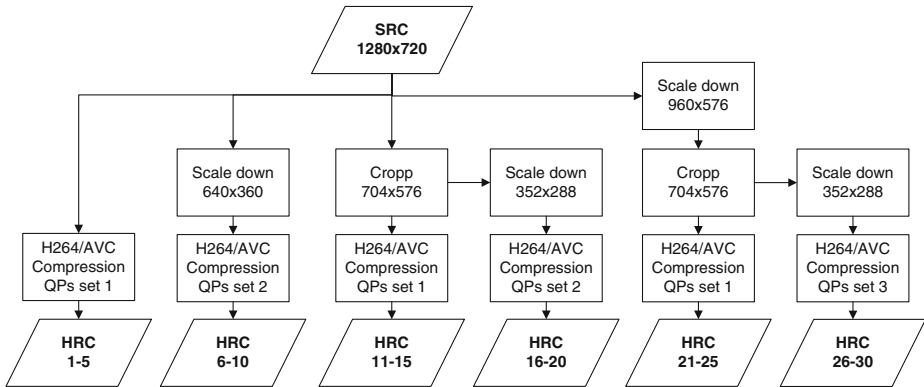
The owners of the vehicles filmed were asked for their written consent, which allowed the use of the video content for testing and publication purposes.

### 4 Processed video sequences

If picture quality is not acceptable, the question that naturally arises is: what is the cause of the degradation? Sources of potential problems are located throughout the end-to-end video delivery chain. The first group of distortions (1) can be introduced at the time of image acquisition. The most common problems are noise, lack of focus or improper exposure. Other distortions (2) appear as a result of further compression and processing. Problems can also arise when scaling video sequences in the quality, temporal and spatial domains, as well as, for example, the introduction of digital watermarks. Then (3), for transmission over the network, there may be some artefacts caused by packet loss. At the end of the transmission chain (4), problems may relate to the equipment used to present video sequences.

Considering this, all source video sequences (SRC) were encoded with a fixed quantization parameter QP using the H.264/AVC video codec, x264 implementation. Prior to encoding, some modifications involving resolution change and crop were applied in order to obtain diverse aspect ratios between car plates and video size (see Fig. 2 for details related to processing). Each SRC was modified into six versions and each version was encoded with five different quantisation parameters (QP). Three sets of QPs were selected:

- {43, 45, 47, 49, 51},
- {37, 39, 41, 43, 45},
- {33, 35, 37, 39, 41}.



**Fig. 2** Generation of HRCs

Selected QP values were adjusted for different video processing paths in order to provide a wide range of plate number recognition ability thresholds. Frame rates have been kept intact as, due to inter-frame coding, their deterioration does not necessarily result in bit-rates savings [16]. Furthermore, network streaming artefacts have been not considered as the authors believe in numerous cases they are related to excessive bit-streams, which had been already addressed by different QPs. Reliable video streaming solutions should adjust the video bit-stream according to the available network resources and prevent packet loss. As a result, 30 different hypothetical reference circuits (HRC) were obtained.

HRCs are described in terms of QP parameters nevertheless in the rest of the paper we use bit-rate as a more network driven parameter. Note that the same QP for different views will result in different bit rates.

Based on the above parameters, it is easy to determine that the whole test set consists of 900 sequences (each SRC 1–30 encoded into each HRC 1–30).

## 5 Automatic number plate recognition

Two different automatic license plate recognition (ALPR) algorithms were considered. The first one is called Labeling and Artificial Neural Networks (LANN) and the second one is called Periodic Walsh Piecewise-Linear Descriptors (PWPLD). In this section both of them are shortly presented.

During early prototyping and design stages this custom built LANN algorithm was extensively compared with more general approaches like well-known Haar cascades. As a result, LANN algorithm outperformed other approaches and research was focused on it. Our early approaches were compared on raw (not distorted by compression artifacts) video sequences. LANN algorithm was able to perform at mean performance of 92 % correctly recognized characters, while Haar cascades achieved only 58 %. Furthermore, custom built algorithm enabled us to fine-tune parameters like license plate dimensions ratio or space between license plate signs and its bounding box. These calibrations yielded performance improvements.

The second PWPLD algorithm is also a custom built ALPR approach. However, it completely differs from LANN algorithm with regard to the classification technique. Because recognizing of license plate characters in the PWPLD algorithm is based on contour features, it should be categorized as a feature based OCR tool, as for instance String Reader, a part of Matrox Imaging software [20]. Taking this into consideration we examined the PWPLD approach in comparison to the String Reader. For row (not distorted by compression artifacts) video sequences we obtained that effectiveness of both methods is comparable. However, as show the results depicted in Section 7, recognition accuracy of the PWPLD approach is much better than of String Reader, in case of video sequences distorted by compression artifacts. Due to the fact that the PWPLD algorithm is still in improvement, we assume above results as promising.

### 5.1 Labeling and Artificial Neural Networks

The LANN algorithm follows a typical chain of a license plate recognition i.e. it includes four steps: preprocessing, plate localization, segmentation for obtaining individual characters, and recognition.

Preprocessing consists of binarization using Otsu method and noise removal using morphological opening and closing [21]. If video signal resolution is too high, it could be downsampled for performance reasons. In the localization step each of connected components in the image are labeled and assigned a set of parameters, such as bounding box, fill factor and size [24]. This parameters are then analyzed to find candidates for number plate. The segmentation process is designed in similar way—each object in the plate candidate region is labeled and parametrized with a bounding rectangle and an area.

The recognition process is based on the Artificial Neural Network with 150 inputs, 73 neurons in the hidden layer and 35 outputs. Input data, i.e. objects located in the license plate, are prescaled to  $10 \times 15$  pixels resolution. This particular size was chosen empirically as the smallest size that enables the network to recognize characters correctly. Clearly, increasing the size would result in better recognition results, but will also extend learning time and processing complexity during the recognition phase.

Moreover, feedbacks from segmentation to localization and from segmentation to preprocessing are used to improve accuracy. Our novel approach with feedbacks in the process enables to correct camera miscalibrations (such as over- and under exposure), and too low or too high number of characters in a particular region obtained after segmentation. Moreover, if all candidate regions were analyzed with no result, the entire process is repeated, but with a binarization threshold modified. Modifications are in a range of 0.8 to 1.2 value of the Otsu threshold.

Data set for training purposes was built using static images, entirely different from video sequences analyzed later on. 30 specimens of each character were split into training, testing and cross-validation sets (respectively 50 %, 25 %, 25 %). The network was trained using scaled conjugate gradient backpropagation with following parameters: performance goal:  $1e-6$ , maximum number of epochs to train: 5000, minimum performance gradient:  $1e-9$ .

Labeling of connected components and ANN approach was chosen as computationally efficient and robust methods. Complexity of labeling could be expressed



as  $O(n)$ , where  $n$  is the number of pixels in single frame and ANN computation is constant time  $O(1)$ . Using one thread on Intel Core i3 CPU the algorithm could process more than 50 frames (resolution  $640 \times 480$  pixels) per second.

## 5.2 Periodic Walsh Piecewise-Linear Descriptors

The PWPLD method was developed according to the Periodic Walsh Piecewise-Linear (PWL) transform descriptors [5]. PWPLD, like LANN, is divided into four main steps. The first step, preprocessing, includes leveling of vehicle position on the processed image. Vehicle leveling is achieved in relation to the car symmetry axis which is drawn as an extension of the longest line detected using both the Canny Edge Detector [6] and the Hough Transform [15].

The second step is the license plate detection (LPD). Because all the analyzed license plates have a standard form—the characters on the plate are dark characters on a light background, a relatively easy LPD method based on looking for regular changes in image brightness was applied.

The third—segmentation and contour transformation step, starts from separation of the individual license plate symbols. This separation is achieved based on the capture window. Capture window width is returned by the LPD procedure and it varies depending on the sequence. In relation to pretty small sizes and low resolution of images separated by the capture window, the resolution and size enhancement operations are performed as the next ones. Those are followed by the Gaussian smoothing which is achieved according to empirically assigned parameters. The Otsu's binarization followed by the set of selected morphological operations (including removing of isolated pixels and dilation) are performed at the next turn [21]. The last performed operations includes contour extraction, normalization and transformation. Contour extraction is achieved using the Single Step Parallel Contour Extraction (SSPCE) algorithm [3, 8]. Extracted contours are uniformly sampled, scaled to predetermined size and shifted to the center of the image. Normalized in that way contours are transformed then using the PWL transform [4]. The PWL transform was selected according to the piecewise-linear character of its basis functions as well as its good approximation abilities. The first ten coefficients (number of spectrum coefficients was assigned empirically) from both X and Y spectrums (simply Cartesian—X,Y, contour representation method was used) are selected then as the individual license plate symbol descriptors. The fourth recognition step is performed according to rules which were constructed using the See5 freeware data mining tool [23].

See5 is a widely used system for informative patterns extraction from data sets. We decided to use this tool on early stage of the PWPLD implementation due to its easiness of use and interpretability (classifiers in See5 are expressed as decision trees or sets of “if-then” rules). However, according to the fact that our approach is still in improvement, we are going to replace See5 tool with dedicated neural network. We expect that neural network will increase the PWPLD recognition rate as well as meaningfully reduce its computational effort. Actually, regarding to the large computational costs of applied See5 “if-then” rules (training set provided as the input to the See5's classification algorithm consisted of almost 1000 PWL descriptors) and due to the fact that the PWPLD algorithm was implemented in Matlab environment (was not speed optimized), its performance was not examined.



## 6 Human recognition analysis

Thirty non-expert testers participated in this study and provided a total of 900 answers. The subjects average age was 23 years old with maximum 29 and minimum 17 years old. 11 female and 19 mails taken part in the study. Each answer obtained was interpreted as 1 or 0, i.e. correct or incorrect recognition. The goal of this analysis is to find the detection probability as a function of a certain parameter(s) i.e. the explanatory variables. The most obvious choice for the explanatory variable is bit-rate, which has two useful properties. The first property is a monotonically increasing amount of information, because higher bit-rates indicate that more information is being sent. The second advantage is that if a model predicts the needed bit-rate for a particular detection probability, it can be used to optimize the network utilization.

Moreover, if the network link has limited bandwidth the detection probability as a function of a bit-rate computes the detection probability, what can be the key information which could be crucial for a practitioner to decide whether the system is sufficient or not.

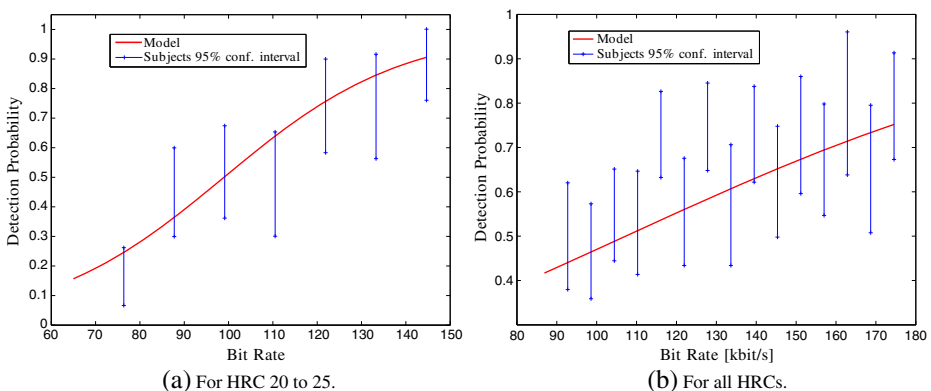
The Detection Probability (DP) model should predict the DP i.e. the probability of obtaining 1 (correct recognition). In such cases, the correct model is logit [1]. The simplest logit model is given by the following equation:

$$p_d = \frac{1}{1 + \exp(a_0 + a_1x)} \quad (1)$$

where  $x$  is an explanatory variable,  $a_0$  and  $a_1$  are the model parameters, and  $p_d$  is detection probability.

The logit model can be more complicated; we can add more explanatory variables, which may be either categorical or numerical. Nevertheless, the first model tested was the simplest one.

Building a detection probability model for all of the data is difficult, and so we considered a simpler case based on the HRCs groups (see Section 4). Each five HRCs (1–5, 6–10, etc.) can be used to estimate the threshold for a particular HRCs group. For example, in Fig. 3a we show an example of the model and the results obtained for HRCs 20 to 25.



**Fig. 3** Example of the logit model and the obtained detection probabilities

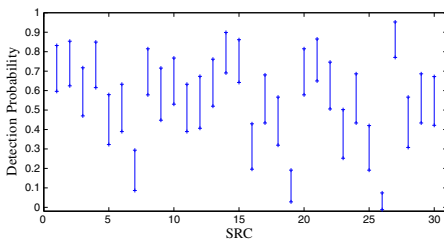
The obtained model it crosses all the confidence intervals for the observed bit-rates. The saturation levels on both sides of the plot are clearly visible. Such a model could successfully be used to investigate detection probability. We present an extension of the sequences analyzed to all HRCs results in the model drawn in Fig. 3b. The result obtained is less precise. Some of the points are strongly scattered (see results for bit-rate 110 to 130 kb/s). Moreover, comparing the models presented in Figs. 3a and b different conclusions can be drawn. For example, 150 kb/s results in around a 90 % detection probability for HRCs 20 to 25 and less than 70 % for all HRCs. It is therefore evident that the bit-rate itself cannot be used as the only explanatory variable. The question then is, what other explanatory variables can be used.

In Fig. 4a we show DP obtained for SRCs. The SRCs has a strong impact on the DP. It should be stressed that there is one SRC (number 26) which was not detected even once. The non-zero confidence interval comes from the corrected confidence interval computation explained in [2]. In contrast, SRC number 27 was almost always detected, i.e. even for very low bit-rates. A detailed investigation shows that the most important factors (in order of importance) are:

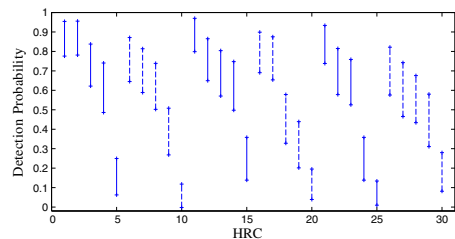
1. the contrast of the plate characters,
2. the characters, as some of them are more likely to be confused than others, as well as
3. the illumination, if part of the plate is illuminated by a strong light.

A better DP model has to include these factors. On the other hand, these factors cannot be fully controlled by the monitoring system, and therefore these parameters help to understand what kind of problems might influence DP in a working system. Factors which can be controlled are described by different HRCs.

In Fig. 4b we show the DP obtained for different HRCs. For each HRC, all SRCs were used, and therefore any differences observed in HRCs should be SRC independent. HRC behavior is more stable because detection probability decreases for higher QP values. One interesting effect is the clear threshold in the DP. For all HRCs groups two consecutive HRCs for which the DPs are strongly different can be found. For example, HRC 4 and 5, HRC 17 and 18, and HRC 23 and 24. Another



(a) For different SRCs with 90% confidence intervals.



(b) For different HRCs. The solid lines corresponds to QP from 43 to 51, and the dashedlines corresponds to QP from 37 to 45.

**Fig. 4** The detection probabilities obtained for human subjects

effect is that even for the same QP the detection probability obtained can be very different (for example HRC 4 and 24).

Different HRCs groups have different factors which can strongly influence the DP. The most important factors are differences in spatial and temporal activities and plate character size. The same scene (SRC) was cropped and/or re-sized resulting in a different output video sequence which had different spatial and temporal characteristics.

In order to build a precise DP model, differences resulting from SRCs and HRCs analysis have to be considered. In this experiment we found factors which influence the DP, but an insufficient number of different values for these factors was observed to build a correct model. Therefore, the lesson learned from this experiment is highly important and will help us to design better and more precise experiments in the future.

## 7 Automatic systems analysis

Our custom build LANN and PWPLD approaches were examined in comparison with three selected widely used and available as a trial version or under the terms acquired license ANPR products:

- JavaANPR—an ANPR approach created in 2007 at Brno University of Technology as a part of the research plan “security-oriented research in information technology, MSM 0021630528” [17],
- DTK ANPR SDK (DTK) library from DTK Software [7],
- String Reader—a part of Matrox Imaging Library (MIL) software [20].

Our first goal was to compare automatic systems with human subjects. Nevertheless, for the sequences seen by subjects the automatic systems recognized correctly only 44 and 15 (LANN and PWPLD respectively) or even less (JavaANPR, DTK, MIL) out of 900 different sequences. For human subjects we obtained 633 out of 900 correct answers. It clearly shows that automatic systems are much more susceptible to coding than human subjects.

Since the number of correctly recognized plate numbers is too small, the number of correctly recognized characters of the whole number plate was considered instead. One of popular string metrics for measuring the difference between two sequences:  $S_1$  and  $S_2$  (in our case original and recognized plate numbers respectively) is the Levenshtein Edit Distance (LED) [18]. The Levenshtein Edit Distance gives an information of the minimum number of character operations (insertions, deletions, substitutions) needed to transform one sequence ( $S_1$ ) to another ( $S_2$ ). In the paper [19] Dekang Lin stated that LED can be expressed as a similarity measure given by the following equation:

$$Sim(S_1, S_2) = \frac{1}{1 + LED(S_1, S_2)} \quad (2)$$

However, in regard of illustration needs (900 different sequences were examined and the total number of all analyzed characters is 6360) we total up the LED values we

obtained for each of the examined sequences and then we divided these totals by the total number of characters. Such a normalized LED value we used then as a substitute for LED value in (2).

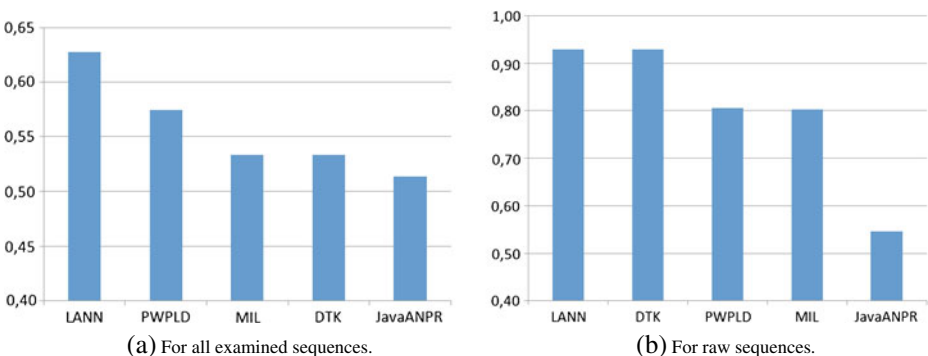
The obtained results are presented in Fig. 5. Figure 5a shows that number of incorrectly recognized characters is the least for LANN ( $Sim = 0.63$ ) and PWPLD ( $Sim = 0.57$ ) algorithms. Similarities for the rest of tested approaches are comparable, however less than 0.55 (0.53 for MIL and DTK and 0.51 for JavaANPR).

However, results presented in Fig. 5b shows that examined methods, according to their accuracies, can be divided into three main groups:

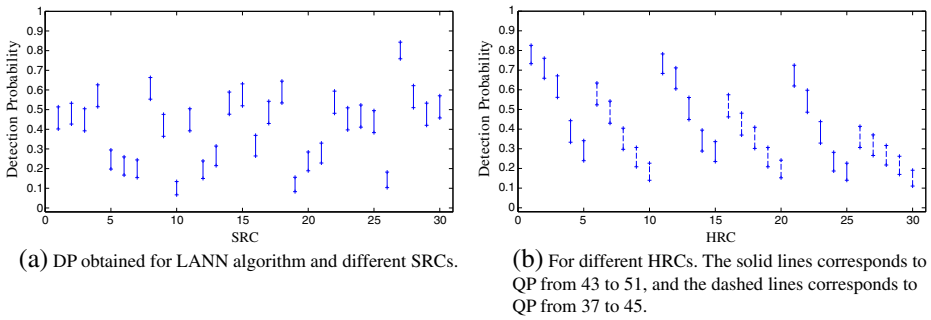
- the first which contains two comparable approaches: LANN and DTK ANPR SDK—similarity for these two methods come to 0.93,
- the second which includes PWPLD and String Reader (MIL) algorithms—similarities for these solutions are 0.81 and 0.80 respectively,
- the third which contains JavaANPR algorithm with similarity equal to 0.55.

It is noticeable that LANN algorithm (as well as DTK ANPR SDK) offers the best accuracy. As it was described in Section 5.1 LANN approach is related to recognition techniques based on dedicated neural networks. PWPLD, and String Reader (MIL) approaches however, as well as JavaANPR method, belong to the category of feature-based OCR tools. When this category only is under consideration, PWPLD algorithm can be indicated as the best solution. Since LANN and PWPLD approaches appeared the best in their categories, further and detailed analysis was related strictly to them both.

The results obtained for the LANN algorithm are presented in Fig. 6. The behavior of the LANN algorithm is similar to subject answers. Different resolutions and QP changes can be clearly seen in Fig. 6b. Each consecutive five points correspond to increasing QP and the DP drops the same as for human subjects. The resolution change influences the detection probability slightly different since in case of human subjects HRC6–8 have just slightly smaller DP than HRC1–2. In case of the LANN algorithm it is not a true.



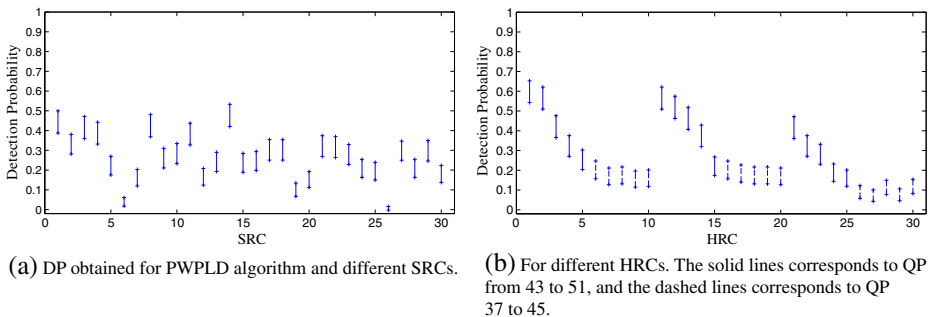
**Fig. 5** The recognition similarities



**Fig. 6** The characters detection probabilities obtained for algorithm LANN

A careful look at Fig.6a shows that the LANN algorithm finds different SRC difficult. For example it seems that SRC10 is particularly difficult for the LANN algorithm and it is not so difficult for human subjects. On the other hand, in case of SRC26 the algorithm outperformed human recognition, as on average one character was recognized for every HRC. It is not shown but in most of the cases human subjects did not recognized even a single character. This is due to the feedbacks and adaptive binarisation thresholds described in Section 5.1 which significantly improve local contrast and enable to extract characters. In case of the PWPLD algorithm the obtained results are shown in Fig. 7.

The results obtained for different HRCs in case of the PWPLD algorithm are not similar to those obtained by human subjects (see Fig. 7b). It is caused by different minimal character size which is assumed by PWPLD algorithm. The minimum size of the separated symbol which meets requirements of the the PWPLD algorithm, especially the requirements of its closed contour extraction procedure (SSPCE), is around 10 by 15 pixels. Obviously for sequences with lower resolution algorithm does not work as good as for the expected resolution. In cases were resolution was high enough QP changes influence DP similarly like in case of human subjects and the LANN algorithm (see HRC1–5, 11–15, 21–25).



**Fig. 7** The characters detection probabilities obtained for algorithm PWPLD

## 8 Conclusions and further work

The paper has answered the practical problem of a car plate recognition probability for video transmitted over a network link with a limited bandwidth. We have presented the results of the quality thresholds detection for a license plate recognition performed by human subjects and by two ALPR algorithms, based on a video streamed in a constrained networking conditions.

In case of human subjects the overall detection probability was much higher than in case of the ALPR algorithms. Therefore, detection model was build for human subject. We have shown that, for a particular view, a model of a detection probability based on a video bitrate can work well. Nevertheless, different views have very different effects on the results obtained. We have also learned that for these kinds of psycho-physiological experiments, license plate characteristics (such as illumination) are of a great importance, sometimes even prevailing over the distortions caused by bitrate limitations and compression.

One important conclusion is that for a bit rate as low as 180 kbit/s the detection probability is over 80 % even if the visual quality of the video is very low. Moreover, the detection probability depends strongly on the SRC (over all detection probability varies from 0 % to over 90 %.)

The performance of the ALPR algorithms is different from the performance of the human subjects in some cases. Nevertheless, at least in case of the LANN algorithm the behavior is close enough to draw a conclusion valid also for the human subjects. It is an important conclusion since the subjective experiments are expensive and strongly limited by a number of analyzed conditions. Another important conclusion is that in almost all cases human subjects outperform the ALPR algorithms. Therefore, a system planned to support an ALPR algorithm can be easily used by humans.

The ultimate goal of this research is to create standards for surveillance video quality. This achievement will include the coordination of research work being conducted by a number of organizations.

**Acknowledgements** The work presented was supported by the European Commission under Grant INDECT No. FP7-218086. Preparation of source video sequences and subjective tests was supported by European Regional Development Fund within INSIGMA project no. POIG.01.01.02-00-062/09. Analysis of video quality was financed by The National Centre for Research and Development (NCBiR) within project no. SP/I/1/77065/10.

The numerical experiments reported in this paper and related to Matrox Imaging Library software have been performed using computational equipment purchased in the framework of the EU Operational Programme Innovative Economy (POIG.02.02.00-26-023/09-00) and the EU Operational Programme Development of Eastern Poland (POPW.01.03.00-26-016/09-00).

**Open Access** This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

## References

1. Agresti A (2002) *Categorical data analysis*, 2nd edn. Wiley, New York
2. Agresti A, Coull BA (1998) Approximate is better than “exact” for interval estimation of binomial proportions. *Am Stat* 52(2):119–126 (1998)

3. Baran R (2011) Segmentation-based object-oriented image compression scheme. *Measurement Automation and Monitoring* 57(1):94–96
4. Baran R, Wiraszka D, Dziech W (2000) Scalar quantization in the pwl transform spectrum domain. In: *Proc. of the int. conf. of Mathematical Methods in Electromagnetic Theory, MMET2000*, pp 218–221
5. Belgassem F, Dziech A, Aboudeana M, Dziech W (1998) Improved fast algorithm for PWL transform. In: *Proc. of the Int. Workshop on Systems, Signals and Image Processing IWSSIP, Zagreb, Croatia*
6. Canny J (1986) A computational approach to edge detection. *IEEE Trans Pattern Anal Mach Intell PAMI* 8(6):679–698
7. DTKSOFT (2011) <http://www.dtksoft.com/index.php>. Accessed 13 Apr 2012
8. Dziech A, Besbas WS, Nabout A, Nour Eldin HA (1997) Fast algorithm for closed contour extraction. In: *Proc. of the 4th international workshop on systems, signals and image processing. Poznan, Poland*, pp 203–206
9. Dziech A, Derkacz J, Leszczuk M (2009) Projekt INDECT (intelligent information system supporting observation, searching and detection for security of citizens in urban environment). *Prz Telekomun Wiad Telekomun* 8–9:1417–1425
10. Eskicioglu AM, Fisher PS (1995) Image quality measures and their performance. *IEEE Trans Commun* 43(12):2959–2965
11. European Commission European Security Research Conference (SRC09) (2009) <http://www.src09.se/>. Accessed 9 Aug 2012
12. Ford C, Stange I (2010) Framework for generalizing public safety video applications to determine quality requirements. In: *3rd INDECT/IEEE international conference on multimedia communications, services and security. AGH University of Science and Technology, Krakow, Poland*, p 5
13. Ford CG, McFarland MA, Stange IW (2009) Subjective video quality assessment methods for recognition tasks. In: Rogowitz BE, Pappas TN (eds) *Human vision and electronic imaging. SPIE Proceedings*, vol 7240. SPIE, Bellingham, WA, p 72400
14. Honess T, Charman E (1992) Closed circuit television in public places: its acceptability and perceived effectiveness. *Tech. rep., Home Office Police Department, London*
15. Hough PV (1959) Machine analysis of bubble chamber pictures. In: Kowarski L (ed) *Proc. of the int. conf. on high energy accelerators and instrumentation, CERN*, pp 554–556
16. Janowski L, Romaniak P (2010) QoE as a function of frame rate and resolution changes. In: Zeadally S, Cerqueira E, Curado M, Leszczuk M (eds) *Future multimedia networking. Proceedings of the third international workshop, FMN 2010, Krakow, Poland, 17–18 June 2010. Lecture Notes in Computer Science*, vol 6157. Springer, New York, pp 34–45
17. JAVA ANPR (2006) <http://javaanpr.sourceforge.net/>. Accessed 13 Apr 2012
18. Levenshtein Edit Distance (LED) (2012) <http://www.levenshtein.net/>. Accessed 13 Apr 2012
19. Lin D (1998) An information-theoretic definition of similarity. In: *ICML '98 proceedings of the fifteenth international conference on machine learning. Morgan Kaufmann Publishers Inc, San Francisco, CA, USA*
20. Matrox Imaging software (2012) [http://www.matrox.com/imaging/en/press/white\\_papers/anpr/](http://www.matrox.com/imaging/en/press/white_papers/anpr/). Accessed 13 Apr 2012
21. Otsu N (1979) A threshold selection method from gray-level histograms. *IEEE Trans Syst Sci Cybern SMC* 9(1):62–66
22. Romaniak P, Janowski L (2010) How to build an objective model for packet loss effect on high definition content based on SSIM and subjective experiments. In: In: Zeadally S, Cerqueira E, Curado M, Leszczuk M (eds) *Future multimedia networking. Proceedings of the third international workshop, FMN 2010, Krakow, Poland, 17–18 June 2010. Lecture Notes in Computer Science*, vol 6157. Springer, New York, pp 46–56
23. See5 (2012) <http://www.rulequest.com/products.html>. Accessed 20 Aug 2012
24. Tadeusiewicz R, Korohda P (1997) *Komputerowa analiza i przetwarzanie obrazow*. Wydawnictwo Fundacji Postpu Telekomunikacji, Krakow
25. VQIPs (2011) Video quality in public safety working group. [http://www.pscr.gov/projects/video\\_quality/vqips/vqips.php](http://www.pscr.gov/projects/video_quality/vqips/vqips.php). Accessed 21 Mar 2012
26. Wang Z, Lu L, Bovik AC (2004) Video quality assessment based on structural distortion measurement. *Signal Process Image Commun* 19(2):121–132 (2004). doi:10.1016/S0923-5965(03)00076-6%20
27. Zeadally S, Cerqueira E, Curado M, Leszczuk M (eds) (2010) *Future multimedia networking. Proceedings of the third international workshop, FMN 2010, Krakow, Poland, 17–18 June 2010. Lecture Notes in Computer Science*, vol 6157. Springer, New York





**Lucjan Janowski** received his M.Sc. degree in Telecommunications in 2002 and Ph.D. degree in Telecommunications in 2006 both from the AGH University of Science and Technology. During 2007 he worked on a post-doc position in CNRS-LAAS (Laboratory for Analysis and Architecture of Systems of CNRS) in France where prepared both malicious traffic analysis and anomaly detection algorithms. In 2010/2011 he spent half a year on a post-doc position in University of Geneva working on QoE for health applications. His main interests are statistics and probabilistic modelling of subjects and subjective rates used in QoE evaluation. He has participated in several commercial and scientific projects. He is an author of several research papers.



**Piotr Kozłowski** is a graduate of Automatics and Robotics at the AGH University of Science and Technology (2011). He continues his education at Telecommunications at the same university. During his studies he received LLP Erasmus scholarship in the University of Ljubljana and participated in internship in GE Hitachi Nuclear Energy in Wilmington, NC, USA. His professional interests include advanced computer vision algorithms, machine learning and artificial intelligence. Currently he develops ANPR system in the INSIGMA research project.



**Remigiusz Baran** received the M. Sc. degree in Electrical Engineering from the Faculty of Electrical and Control Engineering, Kielce University of Technology in 1993 and the Ph.D. degree in Telecommunications from the Faculty of Electrical, Control, Electronic and Computer Engineering, AGH University of Science and Technology in Krakow in 2004. The main areas of his scientific interest are embedded systems and digital image processing including especially feature extraction and object recognition techniques.



**Piotr Romaniak** received PhD degree in Telecommunications (2011). His area of interest includes knowledge about multimedia systems, content-based indexing, perceptual no-reference video and image quality assessment (objective and subjective techniques), video streaming and video delivery technologies (IPTV, VoD), security and watermarking issues. Piotr Romaniak was/is involved in international projects funded by Culture 2000, eContentPlus, FP6 and FP7, as well as nation projects funded by Polish Ministry of Science and Higher Education. He was also involved in few research activities: for polish top telecommunication services provider (assessment of the 3rd generation fax image quality, subjective and objective evaluation methods) and in cooperation with other educational entities (panoramic images concatenating, subjective tests - MOS). Currently he is working on 3DTV Quality of Experience assessment and 2D content quality optimization and assessment in surveillance systems (FP7 IP INDECT Project).



**Andrzej Glowacz** received Ph.D. in Telecommunications from the AGH University of Science and Technology, in 2007. Currently he is Assistant Professor in AGH University. Andrzej Glowacz has been working in numerous industrial works, and the EU research projects including: INSIGMA, INTOM (in both as Deputy Project Coordinator), INDECT, Future Internet Engineering, GAMA, OASIS Archive, CARMEN, DAIDALOS, DAIDALOS 2, EuroNGI and EuroFGI. His main professional areas are: intelligent information systems, multimedia systems, pattern recognition, wireless QoS, modern transport protocols and advanced systems programming. He is author of over seventy scientific papers and technical reports, he also serves as a reviewer of international journals and conferences.



**Tomasz Rusc** received the M. Sc. degree in Telecommunications from the Faculty of Electrical and Control Engineering, Kielce University of Technology in 1998 and the Ph.D. degree in Institute for Circuit and System Theory from the Faculty of Engineering, Christian-Albrechts-University in Kiel in 2006. The main areas of his scientific interest are digital image processing including filter bank design, and semantic modelling and searching of multimedia content.