Intelligent Traffic Video Sensor: Architecture and Applications

Michael Bramberger[1], Roman Pflugfelder[2], Bernhard Rinner[1], Helmut Schwabach[3], Bernhard Strobl[3]

[1] Institute for Technical Informatics
Graz University of Technology
AUSTRIA

[2] Pattern Recognition and Image Processing Group
Vienna University of Technology
AUSTRIA

[3] Video & Safety Technology
Austrian Research Centers Seibersdorf
AUSTRIA

Abstract – This paper presents an intelligent video sensor (IVS) and possible applications in traffic surveillance. The IVS combines video sensing with video processing and data communication. It will be realized as embedded system and captures a video stream, computes traffic information and transfers the compressed video stream and the traffic information to the base station.

Keywords: visual traffic surveillance; video processing; embedded systems; long-term background change

I. INTRODUCTION

New information and communication technologies enable a fundamental innovation in transportation and traffic systems. There are many benefits from introducing these technologies into the existing transportation systems. The most important ones are: (i) The existing infrastructure can be more efficiently used, (ii) the consumption of resources as well as the travel time can be reduced, and (iii) the user can chose the “best” means of transportation for his purpose.

Especially the recent advances in sensor and processor technology offer dramatic benefits for the operation of transportation systems and help to increase the security of the users. In this paper, we present an *intelligent video sensor* (IVS) and its planed application in traffic systems. The IVS combines video sensing with image processing and data communication. It will be realized as embedded system and capture a video stream, compute high-level traffic parameters and transfer the video stream and the traffic parameters to a base station. Traffic parameters include average vehicle velocity, vehicle flow rate as well as detection of obstacles and standstill.

There are various application scenarios for such IVS. They can replace the existing cameras and improve the supervision of tunnels and construction sites. They can deliver valuable data for optimizing the traffic light control in urban areas.

The remainder of this paper is organized as follows: Section 2 introduces the IVS and its architecture as well as its onboard image processing. Section 3 sketches a few application scenarios and Section 4 concludes this paper with a short discussion.

II. INTELLIGENT VIDEO SENSOR

A. Overview

The entire IVS system is packed into a single cabinet, which is typically mounted in tunnels and aside highways. Thus, it is exposed to harsh environmental influences such as rapid changes in temperature and humidity as well as wind and rain. Many problems arise from these influences: The IVS has to operate from *-20°C* up to *+40°C*, so the cabinet has to be heated and cooled, whereby active cooling (e.g. with fans) is a technically poor solution. The power supply is typically 230VAC, or alternatively fed by a solar panel to ensure cheap and flexible operation without the need for expensive cabling to the IVS’s location.

The processing unit is designed to determine various traffic-parameters and to compress a live video-stream simultaneously using a state-of-the-art compression technique like MPEG-4.

The communication interface supports both, wired and wireless connections. A better error-tolerance is achieved, if a wireless connection (e.g. GSM/GPRS) is used concurrently with a wired (e.g. Ethernet) connection. Any of these connections can be used to transmit statistical data, alarm messages and, if enough bandwidth is available, a live video-stream from the scene.

The following two sections describe how the image processing tasks work, present some performance data and introduce the architecture of the IVS.

B. Video Processing

Images are rich of visual traffic information. Video streams are even richer, because they also include the temporal context. Therefore, video processing can deliver diverse traffic information. This information can be divided into quantitative and qualitative information [5]. Quantitative information, e.g., velocity, occupancy, number of vehicles, etc., is directly generated from the image data or by other sensor technologies, e.g., inductive loops. Qualitative information further needs a comparison of the traffic situation with a pre-modeled situation. For example, traffic offences or traffic jams could both be described by a spatiotemporal behavior model [6].

To show some of the necessary video processing capabilities of IVS, we present a solution for the problem of detecting stationary vehicles in tunnels from a video stream. The qualitative decision is based on long-term intensity changes of background pixels. Intensity values are grey values between 0 and 255. Pixels $(x\ y)^T$ of an image are semantically background pixels, if the difference $I_{t-1}(x,y)-I_t(x,y)$ of two consecutive images $I_{t-1}(x,y)$ and $I_t(x,y)$ is smaller than a threshold δ (stationary case) or the intensity value of a pixel is supported by a distribution over background intensity values (statistical case). The assumption is a static IVS and that intensity values of background pixels can be described by a Gaussian distribution. Intensity changes are only caused by the

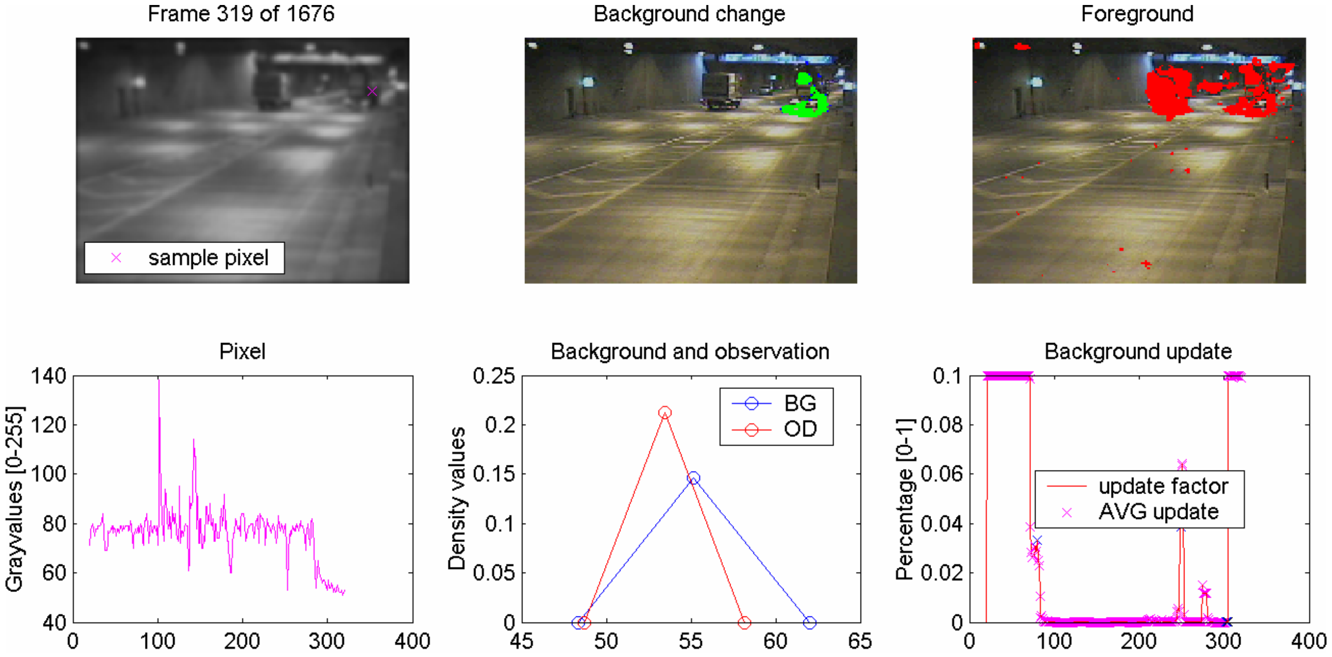

Figure 1: The results of the video processing: The first row shows a specific frame (left) with a stationary truck and a sample pixel, the detected long-term background change (middle) and the foreground regions (right). The second row shows analysis results for the sample pixel: The intensity profile (left), the significant part of the background and observation distribution (middle) and the adoption factor over the frames. "AVG update" indicates adoption case (1).

motion of vehicles.

Each pixel's background model is initialized with a Gaussian with mean equal to 128, 25%-quantile equal to zero and 75%-quantile equal to 255. Then, an observation distribution is updated for each pixel with every new available image of the video stream. The mean and the variance of the observation distribution are estimated by the sample mean and sample variance over the last k images.

In each step, the observation and background distribution are compared. If the significant parts, i.e. between 25%-quantile and 75%-quantile, of both distributions do not intersect each other, then a statistical long-term intensity change in this particular pixel is detected. To make the algorithm robust, a further morphological voting step in the 8×8 vicinity of this pixel is done. If the majority of neighbored pixels do not show the same separation of background and observation distribution, then the gap between both distributions is closed by an updated broader background model.

Regions of pixels with robust changes are found by a connected component algorithm. Stationary vehicles are detected, if two events happen: (i) The area of a region lies between a minimal and maximal threshold and (ii) each intensity profile $I_{t-k}(x,y),\ldots,I_t(x,y)$ over the last k images of all pixels of a region shows a difference $I_{i-1}(x,y)-I_i(x,y)$, $i\in[t-k+1,t]$ greater than δ. Besides, (ii) takes into account that the statistical background change was triggered by an abrupt intensity change, i.e. through a vehicle.

The adoption of the background model is realized with respect to the current observation distribution. Beside the case of the separation of both distributions, further three cases can be distinguished for adoption:

(1) The observation distribution is inside the background model

(2) The background model is inside the observation distribution

(3) Background model and observation distribution intersect each other

In (1), the background model is too broad. Therefore, the parameters of the model are adapted towards the parameters of the observation distribution. This is realized by exponential averaging [7]. The mean of the background model μ_{BG} is updated with the mean of the observation distribution μ_{OD} by $(1-\alpha)\cdot\mu_{BG}+\alpha\cdot\mu_{OD}$. α defines how quickly the background model is updated towards the current observation distribution. In (2) the background model is only updated by exponential averaging, if the difference of the inter-quantile ranges $iqr_{OD}-iqr_{BG}$ of both distributions is small. This is defined by a new adoption factor

$$\alpha \Big/ e^{a(iqr_{OD}-iqr_{BG})} .$$

a defines the sensitivity with respect to the inter-quantile range difference. In (3) the adoption rule is similar to the adoption in (2). The background model intersects the observation distribution from the left or from the right side. Accordingly, the 75%-quantile or 25%-quantile of the background model is adapted towards the 75%-quantile or 25%-quantile respectively of the observation distribution. The mean of the background distribution remains the same. Thus, the background model broadens and supports more of the current observation intensities.

In case of global intensity changes, a significant number of pixel intensities will not be supported by the current background models. Then, all pixel background models are reset to the initialization models.

Figure 1 shows the results of our solution where a stationary truck is detected. This specific example shows only a small portion of video processing algorithms for the IVS. Arithmetic, logical, statistical, threshold, comparison, filtering and morphological operations were used, which can be classified as video/image processing algorithms. But there is also a need of high-level computer vision

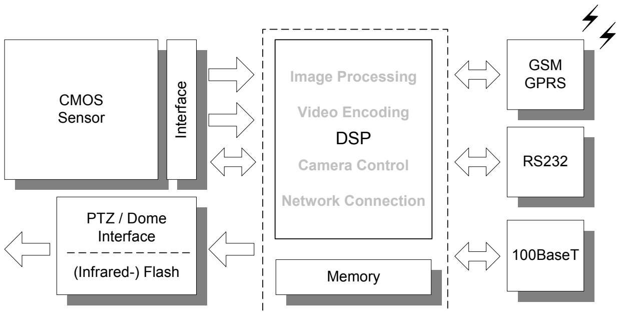

Figure 2: Schematic overview

algorithms like motion analysis, object tracking or calibration techniques to build powerful video processing applications.

C. Architecture

The *Intelligent Video Sensor (IVS)* is divided into three parts: The (i) video sensor captures the optical information, (ii) the processing unit extracts relevant information from the supplied video frames, and (iii) the generic network interface connects the IVS to a chosen network.

CMOS sensors that are produced in the same way than the most available semiconductor products are beginning to supersede the widely-used CCD sensors. These CMOS sensors can be equipped with analog-to-digital converts, which open the door to the digital world. Additionally, CMOS sensors may support non-linear sensitivity, glue less connection to processors as well as minimizing the disadvantages from the CCD devices. Especially logarithmic sensitivity is very important in traffic applications, where bright areas (e.g. high beam) may appear with dark areas (car-silhouette) in the same frame. It eases the development of lens apertures and introduces not yet known image quality.

As mentioned above, on-site processing is inevitable for optimal bandwidth usage as well as for processing the huge amount of data. The main tasks of the onboard processing unit are various image processing steps, video encoding, camera control, and finally the management of the used network connection. The image processing step should (i) extract statistical data from the video data like average speed, number of cars per time interval as well as (ii) detect suspicious driving like cars stopping on highways, cars driving against the traffic, dropped or lost objects on highways and of course, detect accidents. The collected data should then be transmitted via the connected network to a central station where it may be used for traffic control and to initiate the appropriate actions in case of danger or accidents.

Image processing is only the half way to a full surveillance system. A live video stream has to be transmitted, too. Uncompressed video at full resolution would result in a fairly high bit-rate, which is not acceptable these days. Therefore, the video data has to be compressed using a state-of-the-art compression technique like MPEG-4. This compression technique reduces the stream down to a 1/20th of the original size with no visible loss of quality. The combination of the image processing and video compression module enables the transmission of low quality – and therefore low bit-rate video during "normal" mode, and to switch to high-quality video if an alarm situation encounters. This way many (e.g. when using 100Mbit Ethernet up to 70) IVSs can be connected using the same network connection.

The camera control does not require a high performance CPU, since the tasks are not as demanding as image processing or video encoding. Typically, PTZ (Pan-Tilt-Zoom)-installations or domes have to be controlled by the system. The control commands are typically received via a network connection from a central-station, which is operated by user. Optionally, a photoflash or an infrared photoflash should be triggered by the image processing system due to dangerous or conspicuous situations at night or during low-light conditions.

The network interface from the processing unit is generic to allow the connection of different network interfaces without much effort. Possible interfaces are actually Ethernet, GSM (GPRS), ISDN and Wireless LAN. To achieve better error resistance, more network interfaces can work in parallel, whereby the type of connection may differ between the connections. Since the bandwidth of a GPRS connection is much lower than the bandwidth of an Ethernet connection, video streams will not be transmitted via GPRS. However, status signals and still images may be transmitted via GPRS. Therefore the IVS can be used without any cabling required. Although most information will be uploaded by the IVS, a download path to the IVS is present too. Using this download path various information like firmware, parameter sets and camera control-commands can be transmitted. Firmware can be downloaded to fix bugs or more likely to change the image processing tasks to achieve highly flexible systems. Changeable parameter-sets include the bitrate and quality of the video stream as well as optimizing parameters for the image processing subsystem.

Low-power system design is the key to a self-sufficient solution. The power dissipation is kept low, so the IVS can be supplied with a local solar-panel.

III. APPLICATION SCENARIOS

The IVS could potentially and advantageously replace every known camera, frame grabber and computer solution in visual traffic surveillance [2]. The logarithmic CMOS sensor with high dynamic range has several benefits in contrast to the CCD traffic cameras. Problems like large intensity contrasts due to weather conditions or road lights, further blooming, which is an inherent problem of CCD with vehicle lights or low illumination at night would be tackled by the IVS. Furthermore, noise in the video data is reduced by the capability of video computation close to the CMOS sensor. Thus, the IVS delivers a new video quality and better video analysis results, if it is compared to existing solutions.

Beside these qualitative arguments and from a system architecture point of view, the IVS is an important concept in future digital and heterogeneous third generation visual

Traffic area	SS	PD	AS	AC	SF	TI
Motorway surveillance	✓	✓	✓		✓	✓
Surveillance of dangerous traffic sections		✓	✓	✓	✓	✓
Traffic signal control		✓		✓	✓	✓

Table 1: Three different application scenarios for the IVS are shown, which are defined by the traffic area and the application properties, self-sufficiency (SS), plug and detect (PD), active surveillance (AS), actuator control (AC), sensor fusion (SF), traffic information (TI). The key properties in each scenario are marked (✓).

surveillance systems [3].

Flexible video transmission and computing in scalable networks with thousands of cameras will only be possible, if cameras become “intelligent”. Table 1 summarizes three possible application scenarios of the IVS. Every row represents a traffic area, where IVS could be used. Every column shows the potential properties of the IVS. These properties can be explained as follows:

Self-sufficiency: Due to its self-sustained energy and wireless communication capabilities, the sensor could be used in traffic areas, where either no energy or communication infrastructure is available or the installation, i.e. interfacing, cabling, is too expensive.

“Plug and detect”: The wish of the transportation departments to easily install cameras by their road workers without any special knowledge implies sophisticated recognition and self-calibration procedures.

Active surveillance: Active PTZ (Pan-Tilt-Zoom) cameras become increasingly important in contrast to static cameras [4]. The greater flexibility in observing several spots with large inter-distances is a significant cost factor. Consequently, video computing with dependent camera control will become an important property.

Actuator control: Direct influence of actuators, e.g. traffic signals, is possible by the CPU and the communication interface.

Sensor fusion: Several sensors deliver different traffic data with different quality. To increase accuracy and reliability of traffic information, data fusion is a proper way. Nevertheless, in large and scalable networks this requires computation on the sensor. For example, detection of vehicles at all weather conditions could be improved by evaluating the fused video, radar or laser sensor data.

Traffic information: Individual parameterization, depending on quality needs and bandwidth capacities, of MPEG-4 video streams is possible for each IVS. Video computing delivers qualitative information like stationary vehicles, traffic flow behavior, traffic jams, traffic offences, etc. This information could in future be used to index video streams, as it is suggested in MPEG-7. Certainly, gaining quantitative information like velocity, occupancy, vehicle count, etc. is also directly executable on an IVS.

Necessary key properties are marked in table 1 for each specific scenario. “Plug and detect” installation, sensor fusion and the collection of traffic information are possible demands in all scenarios. The first row describes a scenario, where IVS is used to collect traffic information on rural motorways. Rural sites are characterized by the lack of infrastructure. Certainly, self-sufficiency is a key property. Furthermore, to increase the field of view, active surveillance could be an option.

The second scenario describes the close observation of dangerous traffic sections, e.g. winding roads, tunnels or traffic junctions. In some cases, infrastructure is available. There, self-sufficiency is not necessary. On the other hand, actuators, e.g. traffic signals or displays, should be controlled by the IVS within these traffic areas.

A special case - traffic signal control - describes the last row of table 1, where a static IVS could be integrated within an existing infrastructure. This scenario shows the IVS as sensor to collect traffic information like presence and as control unit to perform the light change at urban intersections.

IV. CONCLUSION

In this paper we have presented an intelligent video sensor (IVS) for traffic surveillance. In this ongoing project the IVS will be realized as an embedded system and combines video sensing with image processing and data communication. Due to on-board image processing qualitative and quantitative traffic parameters can be computed by the sensor. Video compression dramatically reduces the bandwidth requirements. The IVS will, therefore, become an important component of future video traffic surveillance systems.

V. REFERENCES

[1] Gian Lica Foresti, Petri Mähönen and Carlo S. Regazzoni (eds), *Multimedia Video-based Surveillance Systems*, Kluwer 2000

[2] Kevin Borras, *In life you must have vision*, Traffic Technology International, pages 36-39, UK & International Press, Oct/Nov, 2002

[3] Carlo Regazzoni, Visvanathan Ramesh, Gian Luca Foresti, *Introduction of the special Issue*, Proceedings of the IEEE, Vol. 89, No. 10, Oct, 2001

[4] Eric Toffin, Gregory Pernot, Salah Bouzar, *Active video surveillance*, Traffic Technology International, pages 28-30, UK & International Press, Oct/Nov, 2002

[5] Roman P. Pflugfelder, *Visual Surveillance Using Real-time Tracking*, master’s thesis, University of Technology, Vienna, 2002

[6] Roman P. Pflugfelder, Horst Bischof, *Learning Spatiotemporal traffic behaviour and traffic patterns for unusual event detection*, 26th Workshop of the Austrian Association for Pattern Recognition, pages 125-133, Graz, 2002

[7] Richard Lyons, *Understanding Digital Signal Processing*, Addison-Wesley, 1997

[8] Paolo Remagnino, Graeme A. Jones, Nikos Paragios and Carlo S. Regazzoni (eds), *Video-based Surveillance Systems*, Kluwer, 2002