IMPROVEMENT OF A PROCEDURE FOR VEHICLE DETECTION AND TRACKING BY BASE FRAME UPDATING AND KALMAN FILTER

G.Artese[a]

[a]Land Planning Dept., University of Calabria, 87036 Rende, Italy - g.artese@unical.it

Commission III - WG III/5

KEY WORDS: Digital Photogrammetry, Vehicle Tracking, Kalman Filter, Change Detection

ABSTRACT:

For the vehicle detection and tracking in the roundabouts, several difficulties have to be faced, e.g. the use of non nadiral perspectives, vehicle images overlapping and radiometric differences between top and side of the cars; the silhouette of the cars is, besides, not regular and the size is variable. In the paper, the upgrade of a procedure for detecting and tracking vehicles in a roundabout is presented. The image subtraction technique is used, along with geometric and radiometric filters. The problems due to the variation of the background radiometric characteristics are discussed. Kalman filter has been used to foreseen the position of vehicles for a better tracking, and to improve the determination of the trajectories. The results of a test are presented.

1. INTRODUCTION

Object detection and tracking can be very useful in a range of industries and applications, like surveillance, vehicle and pedestrian tracking. Trajectory and velocity estimation of vehicles is very important for the management of traffic, above all in the urban zones and for road intersections.

The most used methods for traffic monitoring allow to obtain the number of vehicles crossing a given section of a road; automatic sensors (pressure hose sensors, magnetic buried loops) have been used for decades, to automatically count the axes or the vehicles, but no information about trajectories are generally obtained.

Several procedures, based on digital photogrammetry techniques and computer vision have been proposed in the last years for the estimation of traffic flow. New available technologies have been also used, and multisensor systems have been tested (Grejner-Brzezinska et al., 2007, Toth and Grejner-Brzezinska, 2007, Yao et al., 2009, Goyat et al., 2009). The main advantage of the computer vision techniques is the possibility to obtain information regarding not only the number of entering and outgoing vehicles, but also their trajectories. Some authors proposed algorithms set up by using a single camera (Reulke et al., 2002, Broggi and Dickmanns, 2000) or multi-camera systems (Pedersini et al., 2001). In more complex systems, images obtained with infrared cameras are also used (Dalaff et al., 2003, Hinz and Stilla, 2006, Kirchhof and Stilla, 2006). The detection of vehicles is generally performed by using procedures based on the segmentation of the groups of pixels having similar chromatic values, and on the edges extraction. Morphologic operations (regions filling and erosion, edges dilatation) can be performed to optimize the results. Depending on the kind of sensor data, a data base with the characteristics of the vehicles can be used for facilitating the recognition, while geometric conditions, such as the parallelism between street edge and vehicle trajectories can be imposed only for straight roads (Puntavungkour and Shibasaki, 2003).

If the differences of images are used, the frames obtained by a camera are compared to detect the presence and the position of the vehicles. One can follow essentially two ways: to compare every frame with the previous one, or to compare every frame with a *base frame* (background) without vehicles.

Due to the short time interval between two consecutive frames, the background is practically identical in the first case, and it is possible to detect the vehicles in motion, but not those immobile ones; if this way is followed, the block-based motion estimation is usually chosen, by using high performance computers (Min Tan and Siegel, 1999). Correlation techniques are used, by subdividing two consecutive frames in small blocks, and by obtaining for every block the movement vector: a neighbourhood window of pixels in a given image, centred on a specific pixel, is searched over a larger neighbourhood window of pixels in the previous image, centred on the same pixel. It is assumed that from frame-to-frame in a video sequence pixel intensity values do not change and that the video source is stationary. The position where the minimum absolute or squared difference is obtained, gives the motion vector for the centre pixel. To save run time, the block-based motion estimation is run only for a subset of pixels in the image and the results are interpolated over the entire image.

If a base frame is used, it is possible to detect both mobile or immobile vehicles, but some problems must be solved, essentially due to:

- variations of the lighting conditions, that make practically impossible to use a single *base frame*;
- presence of noises in the images;
- shadows projected from the vehicles, trees and surrounding buildings;
- movements of the video camera (oscillations and spins, even if modest, make the elimination of the background difficult and create false moving objects).

Other difficulties to overcome are the superimposition of the images of close vehicles and the necessity of a real time elaboration.

A key element for many target tracking algorithms is an accurate background subtraction (Seki and Wada, 2003, Isenegger et al., 2005). An overview of several techniques for background detection and updating has been made by Cheung and Kamath (2004). In this paper, simple Frame Difference (FD), Median Filter (MF) (Cucchiara at al., 2003), Approximated Median Filter (AMF) (McFarlane and Schofield, 1995) , Kalman Filter (KF) and Mixture of Gaussian (MoG) (Friedman and Russell, 1997, Stauffer and Grimson, 1999) are compared. The best results are obtained by MoG and MF, followed by AMF. In spite of its simple implementation, low storage requirements and computing rapidity, Approximated Median Filter shows good performances. Also the Running Gaussian Average (RGA) (Wren et al., 1997) can be used with good results. More sophisticated approaches, like Bayesian Background Estimation (Rahimizadeh et al., 2009) or multimodal information integration (Kato and Wada, 2004) has been proposed. Kalman Filter for background detection has been implemented in commercial software packages (MVTec , 2009).

In the following, a technique for detecting and tracking vehicles in a roundabout is described. Image subtraction, geometric and

radiometric filters are used. The problems due to the variation of background are shown, along with the chosen solution.
In order to effectively perform the detection of vehicles and their trajectory, a prediction of their positions is obtained by using data previously collected and a Kalman filter.

2. METHODOLOGY

2.1 Candidate vehicles detection and labelling

A base frame is chosen, with no vehicles in the roundabout (Figure 1). For every frame (Figure 2), the difference of the base image is performed (Figure 3). By applying a non-maxima suppression, the pixels with null or negligible radiometric absolute differences are black, and those for which variations have been detected have non zero values. The following operations are then carried out: pixels are grouped through a segmentation and some regions are detected, whose edges are extracted with classic techniques (Canny, 1986); open lines are connected through an expansion of the edges, and a clustering is performed by filling the regions surrounded by closed lines; an erosion of the edges is then performed, in order to restore the original dimensions of the filled regions. A filter is used to eliminate the regions having a width less than a selected threshold; every obtained area, corresponding to a candidate vehicle, is finally labelled (Figure 4). In this way, small noises are eliminated, along with the false foregrounds (known as *ghost objects*). A counterpart of this procedure is the elimination of some real foregrounds: in figure 4 the region corresponding to the red car on the left side of figure 2 is

Figure 1. Base Frame

Figure 2. Compared Image (Frame 78880)

Figure 3. Frames difference before non-maxima suppression

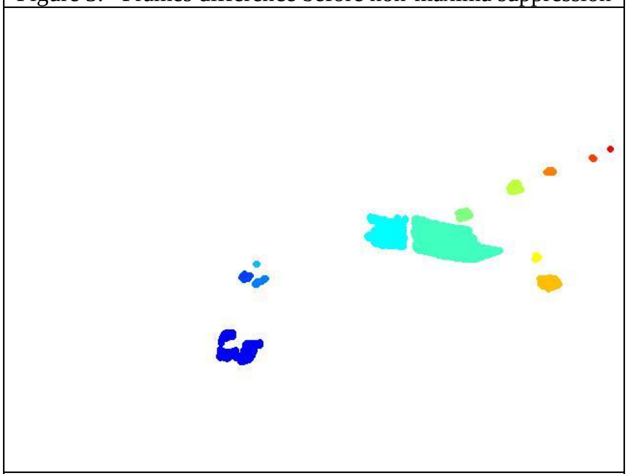

Figure 4. Detected regions labelling for the frame 78880

present in figure 3, but it was eliminated by non-maxima suppression performed using a high threshold. Kalman filter will be useful in this case, as described later.
Generally, to have the view of a whole roundabout, the camera should be positioned in a high place; light poles can be used, but in this case oscillations are expectable, the confronted frames are translated and rotated, and several “false” areas will be detected (Artese, 2008).
In order to solve this problem, a registration should be performed; automatically detectable targets (Fraser, 1997), or known points external to the road area, can be effectively used.

2.2 Base frame updating

Background subtraction is used primarily to identify image regions that contain foreground information. The ideal background model is dynamic and should handle slight variations in the background conditions. For this reason, an updating must be performed. The changes in the background are often not predictable, above all in a partially cloudy day. Also in sunny days the changes due to the motion of the shadows of poles and buildings are fast.
Figure 5 shows a base frame obtained only 30 minutes before the frame of figure 1. It is possible to observe some big differences: the shadow of the building on the left side, the shadow of the central pole and the colour of the road. In fact, the surface of the road was rapidly drying, and, consequently, the radiometric characteristics were changing. If we perform the difference between figure 2 and figure 5, we obtain a result (Figure 6) very different from Figure 3. In Figure 7 the difference between the base frames is shown. It is evident that few minutes are sufficient to change the background.

Figure 5. First Base Frame

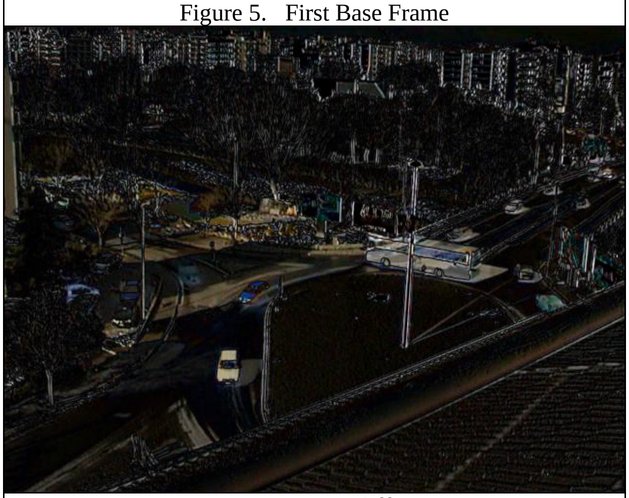

Figure 6. Frames difference

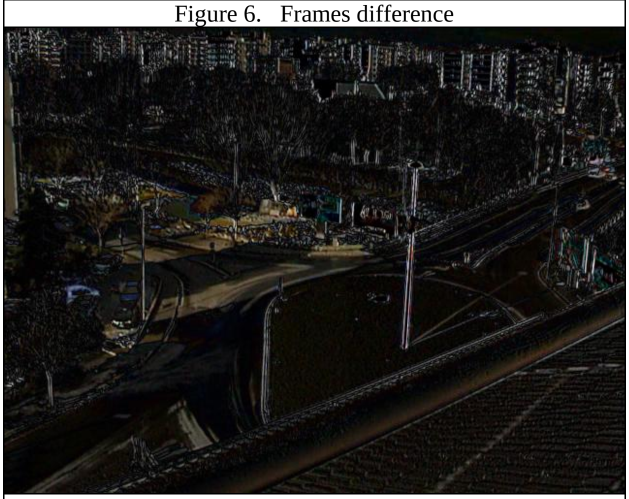

Figure 7. Difference between base frames

To evaluate the performances of the techniques for background detection, some information retrieval measurements are generally used, based on the number of pixels correctly detected by the algorithm. For this aims, "ground-truth" frames are obtained, by manually highlighting the foreground.

It can be observed that, when the described methods are directly used for background updating and foreground detection, some drawbacks are present:

- all background algorithms are sensitive to environmental noises;
- algorithms that adapt more slowly (MF, KF, RGA, MoG) have better performance than those that adapt quickly, but, in case of sudden changes, produce "ghost objects";
- For RGA and MoG, a rapid variation in global illumination, after a long stationary period, can turn the entire frame into foreground, due to the very small variances of the background components.

In our case, two operations are carried out for the current frame processing: after the Frame Difference, both morphological operation and non maxima suppression are performed. To update the base frame the following criterion has been adopted: the pixels corresponding to the black ones after the current frame processing (i.e. the pixels not belonging to the candidate vehicles), are used to update the base frame. With reference to the described example, the pixels of figure 2 corresponding to the white ones of figure 4 are used to update the corresponding ones of figure 1. In fact, figure 1 is not a *real* frame, but it is the frame of figure 5 after the updating.

2.3 Vehicle recognition and tracking

The blobs obtained and labelled are used to recognize the entering vehicles and those already present in the previous frames, to determine their trajectories. The flow chart of the algorithm used to build the compatibility table, along with some examples, is shown in (Artese, 2008). The adopted strategy is the following:

- in the first frame the coordinates of the moving vehicles are known; these vehicles have been opportunely labelled;
- for every region detected in the second frame, a vector is built, which elements are the coordinates of the barycentre, both orthogonal (pixel) and polar with reference to the centre of the roundabout, along with other characteristics (area, average RGB and HSI values, bounding box);
- every vehicle of the first frame is confronted with every region of the second frame and there are obtained the distance, the march direction (clockwise or counter-clockwise) and the ratio between the areas;
- a table is obtained in which, for every vehicle, the regions of the second frame compatible for trajectory and dimension are reported. The radiometric characteristics are then compared and a figure of merit is obtained for every couple vehicle-region.

– by using the table, the number of the corresponding vehicle should be assigned to every region; in case of incoming vehicles, the number of the last vehicle increased by one will be assigned.

In several cases, the correspondence between vehicles and blobs is not trivial. If we exclude the entering or outgoing vehicles, and the case of compatibility with only a region, several possibilities should be investigated. By comparing figures 3 and 4, we can observe that the coach has been divided into two regions due to the light pole, while the blue car in the middle of the frame has been divided into three blobs.

In this case, the sum of the areas should be compared with the vehicle having compatible barycentre.

The obtained coordinates of the vehicles allow to obtain both trajectory and velocity, once known the time interval between consecutive frames.

2.4 The use of Kalman filter

The prediction of the barycentres of the vehicles can be very useful to improve the results of the comparisons above described. For this aims, the use of Kalman filter (Kalman, R.E., 1960) has been foreseen.

The Kalman filter addresses the general problem of trying to estimate the state of a discrete-time controlled process that is governed by the linear stochastic difference equation:

$$x_k = Ax_{k-1} + Bu_{k-1} + w_{k-1} \qquad (1)$$

with a measurement

$$z_k = Hx_k + v_k \qquad (2)$$

In these equations the random variables $\mathbf{w_k}$ and $\mathbf{v_k}$ represent, respectively, the process and measurement noise at the time k.
In the case of a roundabout, the state equation (1) is non linear and the extended Kalman filter must be applied.
The dynamic state variables are the vectorial position and velocity of the vehicles.
Position and velocity are characterized by the components in x and y (or, in a polar system, radial and angular)
The roundabout studied in this paper, has been monitored (Guido et al., 2009) by using virtual detectors (Figure 8); mean trajectories and mean velocity profiles has been obtained (Figure 9).
By using the obtained trajectories, it is possible to write the simplified state equation:

$$X_{k+1} = \Phi_k X_k + w_k \qquad (3)$$

where X_{k+1} is the position and velocity vector at the *k+1* time (frame), and Φ_k is the system transition matrix. Given the time interval $\mathbf{\Delta t}$, the expressions of X_{k+1} and Φ_k are:

$$X_{k+1} = \begin{bmatrix} x_{k+1} \\ y_{k+1} \\ v_{x,k+1} \\ v_{y,k+1} \end{bmatrix} \qquad \Phi_k = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The measurement equation is:

$$z_{k+1} = HX_{k+1} + v_{k+1}$$

where the measurement matrix is:

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

The process noise $\mathbf{w_k}$ is obtained by the trajectories and velocities variances given by Guido et al.. The measurement noise $\mathbf{v_k}$ is obtained by the variance of the barycentres of the vehicles detected in the frame.
For every recognized vehicle, the analysis starts after the first two frames in which the vehicle is present. Position and velocity are obtained; these parameters are utilized for estimating new position and velocity by using equation (3). We utilize these predicted positions for recognize the blob (or the blobs) in the next frame, corresponding to the vehicle: a simple test on the distance between blob and vehicle predicted barycentres is applied. The individuated blob is used for obtaining the new position and velocity values at k+1 step. In this way we can correct errors due to failed vehicle recognition.

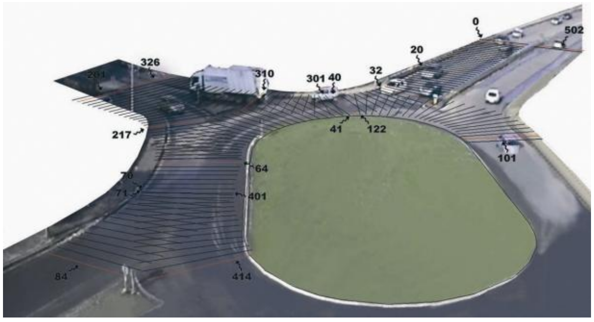

Figure 8. The roundabout with the virtual detectors

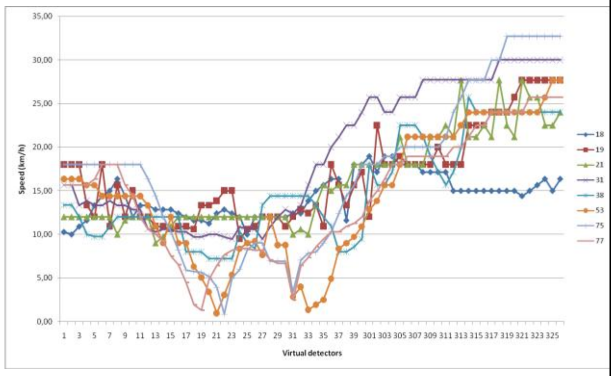

Figure 9. Velocity profiles for 8 vehicles

The Kalman filter is also applied to estimate the *a posteriori* state $\hat{X}_{k+1}$ and the *a posteriori* error covariance P_{k+1}^{+} . The procedure is iterated, and the obtained values can be useful to detect overlapping vehicles, when their trajectories diverge.

3. THE TEST

A test has been executed on a portion of roundabout in the city of Cosenza, Italy. A digital camera "Sony Handycam TRV14E" has been used, with 640 x 480 pixels frames. The acquisition rate is 25 frames per second. For the elaboration of the images and the comparisons the Matlab^{TM} software has been used. The comparison have been performed every 10 frames, with a time step of 0.4 seconds. The data obtained by Guido et al. have been used, along with Kalman filter, to obtain the foreseen positions of the vehicles.
To evaluate the effectiveness of the Kalman filter, we can consider the detection and tracking of the red car on the left side of figure 2.
Figure 10 shows the frame 78830, with the bounding boxes of the selected regions; figure 11 shows the relevant labelling: it is possible to observe that the red car, entering the shadowed zone, is recognized. Figure 2 (frame 78880) has been obtained two seconds after figure 10; as shown in the figure 4, the red car wasn't detected and no regions were identified, compatible with its previous position. Due to the prediction of Kalman filter (obtained taking into account also frames from 78840 to 78870), a car should be present in the shadowed zone of figure 2; for this reason, the threshold used for non maxima suppression (0.2) has been reduced to 0.05 for a neighborhood of the foreseen position of the red car. The result of this operation is shown in figure 12, where the outlines of the detected regions are superimposed to the frame 78880. The positions foreseen by Kalman filter have been used for all frames, and an enhancement of the vehicles detection has been obtained.

Figure 13 shows the trajectories reconstructed for four vehicles present in figure 2. The image is obtained by using the regions detected in 8 consecutive frames. The trajectories are the lines connecting the centroids of the regions.
The data manually obtained by Guido et al. for the next hour, used as reference data, have been compared with the results of the described procedure. Traffic ad meteo conditions was variable and influenced the performances. The best results were obtained for low traffic intensity and cloudy weather; in this case 95% of vehicles have been correctly recognized and tracked. Errors are mainly due to cars hidden by trucks or busses. The performances decrease in case of high traffic density, and the percentage of correctly recognized vehicles decreases to 80%. During the test 2475 vehicles crossed the roundabout; by using the described procedure without Kalman filter, only 1847 were correctly recognized and tracked; with Kalman filter 2153 vehicles (87%) were correctly recognized and tracked.


Figure 10. Frame 78830 with bounding boxes

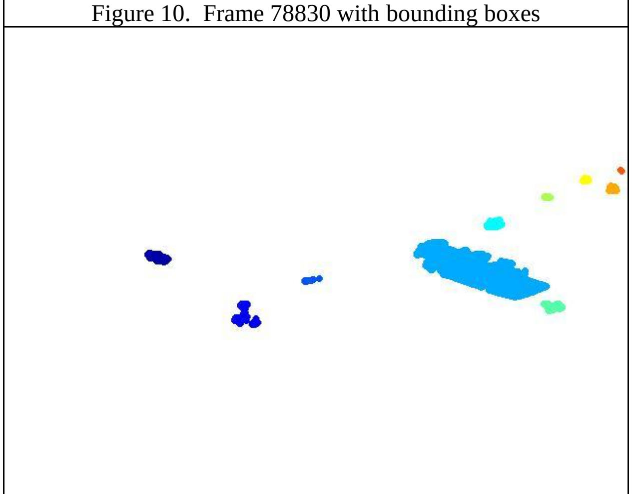
Figure 11. Detected regions labelling for the frame 78830

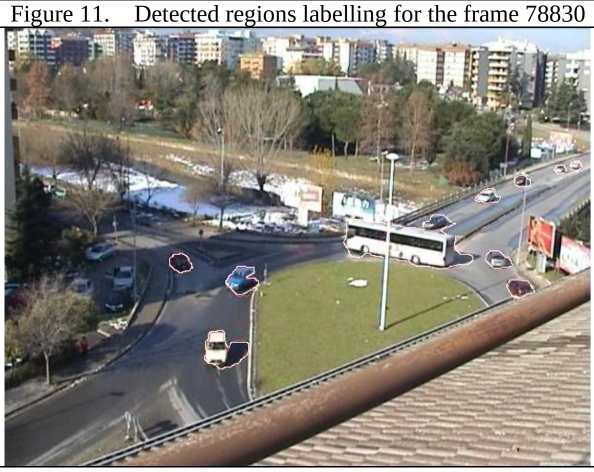
Figure 12. Frame 78880 with the outlines of the detected regions

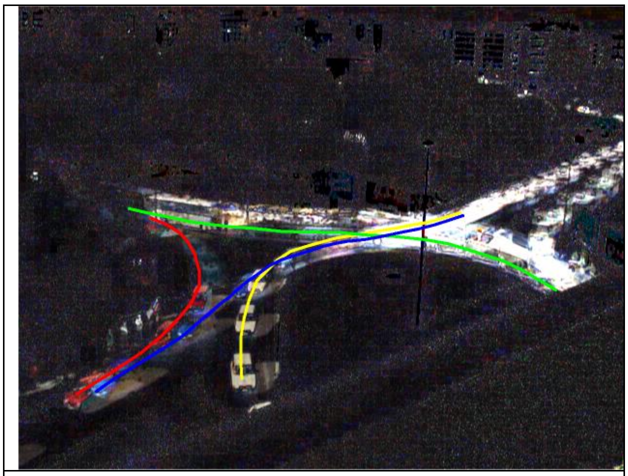
Figure 13. Trajectories of four vehicles

4. CONCLUSIONS

A methodology has been described, for the vehicles detection in a roundabout and for the determination of their trajectories.
The differences of images are used, and the frames obtained by a camera are compared with a *base frame* (background) to detect the presence and the position of the vehicles. The regions, corresponding to the possible vehicles present in the roundabout, are isolated by a segmentation operation, using edge extraction, region filling and dimensional filtering. Recognition and tracking are performed by comparing the extracted regions of a frame, and the vehicles present in the former frame. A compatibility table is built, which allows the vehicle recognition. The trajectory and velocity are obtained.
The importance of an updated base frame has been underlined, and the technique used for the updating has been illustrated. The use of Kalman filter to foreseen the position of the vehicles has been described and the results of a test lead on a real case have been shown.
The present work regards the optimization of Kalman filter and of the procedure. Next studies will be addressed to the resolution of the problems to face, with particular regard to the elimination of the shadows from the regions associated to the vehicles.

REFERENCES

Artese, G., 2008. Detecting and Tracking Vehicles in a Roundabout. In: *International Archives of Photogrammetry and Remote Sensing and Spatial Geoinformation Sciences*, XXXVI, PART 5/C55, pp. 6 – 13

Broggi, A., Dickmanns, E. D., 2000. Applications of computer vision to intelligent vehicles. *Image and Vision Computing*, 18(5): pp. 365–366.

Canny, J., 1986. A computational approach to edge detection. In: *IEEE Transactions on Pattern Analysis and Machine*

Intelligence, 8 (6), pp. 679–698.

Caselles, V., Kimmel, R., Sapiro, G., 1997. Geodesic active contours. *International Journal of Computer Vision*, 22(1):61–79, 1997. 3

Cheung, S.C., Kamath, C., 2004. Robust techniques for background subtraction in urban traffic video. In: *Proceedings of SPIE, Visual Communications and Image Processing 2004*, 5308, pp. 881–892.

Cucchiara, M.P.R., Grana, C., Prati, A., 2003. Detecting moving objects, ghosts and shadows in video streams. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25 (10), pp.1337–1342.

Dalaff, C., Reulke, R., Kroen, A., Kahl, T., Ruhe, M., Schischmanow, A., Schlotzhauer, G., Tuchscheerer, W., 2003. A Traffic Object Detection System for Road Traffic Measurement and Management. http://www.sprg.massey.ac.nz/ivcnz/Proceedings/IVCNZ_15.pdf , (accessed 5 April 2009)

Di Stefano, L., Viarani, E., 1999. Vehicle Detection and Tracking Using the Block Matching Algorithm. *Proc. of "3rd IMACS/IEEE Int'l Multiconference on Circuits, Systems, Communications and Computer"* Athens, Greece, July 4-8 1999, Vol. 1, pp. 4491-4496.

Fraser, C.S., 1997. Innovations in Automation for Vision Metrology Systems. *The Photogrammetric Record*, 15(90), pp. 901-911.

Friedman, N., Russell, S., 1997. Image segmentation in video sequences: A probabilistic approach. *Proceedings of the Thirteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI{97)*, pp. 175 - 181.

Goyat, Y., Chateau, T. and Trassoudaine, L., 2009. Tracking of vehicle trajectory by combining a camera and a laser rangefinder. *Machine Vision and Applications*, 21 (3), pp. 275 - 286.

Grejner-Brzezinska, D.A., Toth, C.K. and Paska, E., 2007. Airborne remote sensing supporting traffic flow estimation. In: *Advances in Mobile Mapping Technology*, Tao V. C. and Li J. editors, Taylor & Francis, London, pp. 51-60.

Guido G. , Saccomanno F. , Astarita V. , Vitale A. , 2009. Measuring safety performance at roundabout using videotaped vehicle tracking data. In: *Proceedings of 88th Annual Meeting Transportation Research Board*, Washington D.C., January 11-15 2009, CD-ROM.

Hinz, S., Stilla, U., 2006. Car detection in aerial thermal images by local and global evidence accumulation. *Pattern Recognition Letters* 27 (4), pp. 308-315.

Isenegger, L., Salgado, L., Garcia N., 2005. Moving objects segmentation based on automatic foreground/background identification of static elements. *Lecture notes in computer science* , vol. 3708, pp. 491-498.

Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. In: *ASME Transactions, Series D: Journal of basic engineering*, n. 82, pp. 35 – 45.

Kato, T., Wada, T., 2004. Integration between background subtraction and colour detection based on nearest neighbour classifier: Instance based multimodal information integration. In: *Information Processing Society of Japan Transactions on Computer Vision and Image Media*, Vol. 45, n.13, pp.110–117.

Kirchhof, M., Stilla, U., 2006. Detection of moving objects in airborne thermal videos. *ISPRS Journal of Photogrammetry and Remote Sensing*, 61(3-4), pp. 187-196

McFarlane, N., Schofield, C., 1995. Segmentation and tracking of piglets in images. *Machine Vision and Applications*, 8(3), pp. 187-193.

Min Tan, J.M.S., Siegel, H.J., 1999. Parallel implementati ons of block-based motion vector estimation for video compression on four parallel processing systems. *International Journal of Parallel Programming*, 27 (3), pp.195–225.

Pedersini, F., Sarti, A. and Tubaro, S., 2001. Multi-camera parameter tracking. In: *IEE Proceedings -Vision,Image and Signal Processing*, 148(1): pp. 70– 77.

Puntavungkour, S., Shibasaki, R., 2003. Novel Algorithm of Vehicle Detection by Using New Ultra Resolution Aerial Image, Three Line Scanner. In: *Proceeding of the IEEE International Conference on Systems, Man and Cybernetics, 5 – 8 Oct. 2003*, Vol. 1, pp 234- 239.

Rahimizadeh, H., Marhaban, M.H., Kamil, R.M., Ismail. N.B., 2009. Color Image Segmentation Based on Bayesian Theorem and Kernel Density Estimation, *European Journal of Scientific Research*, 26 (3), pp.430-436.

Reulke, R., Boerner, A., Hetzheim, H., Schischmanow A. and Venus, H., 2002, A sensor web for road-traffic observation. In *Image and Vision Computing - New Zealand*, pp. 293–298.

Seki, M., Wada, T., Fujiwara, H., Sumi, K., 2003. Background detection based on the cooccurrence of image variations. *Proceedings of CVPR*, Vol. 2, pp. 65–72.

Stauffer, C., Grimson, W.E.L., 1999. Adaptive Background Mixture Models for Real-Time Tracking. *Proceedings of CVPR* Vol. 2, pp. 246-252.

Toth, C.K., Grejner-Brzezinska, D.A., 2007. Vehicle classification from LIDAR data to support traffic flow estimates. In: *Advances in Mobile Mapping Technology*, Tao V. C. and Li J. editors, Taylor & Francis, London, pp. 119-130.

Wren, C., Azarbayejani, A., Darrell, T., Pentland, A., 1997. Pfinder:Real-time Tracking of the Human Body. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7), pp. 780-785.

Yao, W., Hinz, S., Stilla, U., 2009. Vehicle activity indication from airborne LiDAR data of urban areas by binary shape classification of point sets. In: *International Archives of Photogrammetry, Remote Sensing and Spatial Geoinformation Sciences*, XXXVI, (3/W4), pp. 187-192.

MVTec Software GmbH, 2009. Halcon 9.0 documentation. http://www.halcon.de/download/documentation/documentation-9.0.html (accessed 5 April 2010).