INTERNATIONAL SYMPOSIUM ON
MODERN TECHNOLOGIES, EDUCATION AND PROFESSIONAL PRACTICE IN GEODESY AND RELATED FIELDS
Sofia, 03 – 04 November, 2005

AUTOMATIC ESTIMATION OF VEHICLE SPEED FROM UNCALIBRATED VIDEO SEQUENCES

Lazaros Grammatikopoulos, George Karras, Elli Petsa (GR)

ABSTRACT

Video sequences of road and traffic scenes are currently used for various purposes, such as studies of the traffic character of freeways. The task here is to automatically estimate vehicle speed from video sequences, acquired with a downward tilted camera from a bridge. Assuming that the studied road segment is planar and straight, the vanishing point in the road direction is extracted automatically by exploiting lane demarcations. Thus, the projective distortion of the road surface can be removed allowing affine rectification. Consequently, given one known ground distance along the road axis, 1D measurement of vehicle position in the correctly scaled road direction is possible. Vehicles are automatically detected and tracked along frames. First, the background image (the empty road) is created from several frames by an iterative per channel exclusion of outlying colour values based on thresholding. Next, the subtraction of the background image from the current frame is binarized, and morphological filters are employed for vehicle clustering. At the lowest part of vehicle clusters a window is defined for normalised cross-correlation among frames to allow vehicle tracking. The reference data for vehicle speed came from rigorous 2D projective transformation based on control points (which had been previously evaluated against GPS measurements). Compared to these, our automatic approach gave a very satisfactory estimated accuracy in vehicle speed of about ±3 km/h.

1. INTRODUCTION

Video sequences of road scenes are increasingly used in several contexts with an emphasis on automation, notably for tracking moving objects in a static background (video processing techniques for traffic applications are surveyed in Kastrinaki et al., 2003). Among the mobile mapping and video-logging systems reported in literature (Tao & El-Sheimy, 2000) one finds multi-sensor systems with more than one camera, providing geo-referenced image sequences, as well as simpler options based on a single camera. In these single-image cases, measurement in 3D is generally impossible; hence, certain geometric constraints (such as the "flat-earth" model) must be adopted. Several approaches are founded on a priori knowledge of interior and exterior camera orientation (Dailey et al., 2000). However, the use of un-calibrated cameras with minimal external information allows flexibility and low cost. In such applications, the vanishing point in the road direction usually plays an important role. Simond & Rives (2003), for instance, compute the projectivity between two images using constraints from the dominant vanishing point, while Bose & Grimson (2003) use vanishing points estimated from constant vehicle velocity.

The potential of automatic single-image approaches is being currently studied for various purposes. Besides tasks such as the development of algorithms for automatic lane and obstacle detection, the studies of traffic flow parameters constitute a significant field of research. Video sequences from a stationary camera may be used in this respect, with the application of image processing and vision algorithms to traffic scenes for queue detection and vehicle classification or counting (Dailey et al., 2000). Ordinary video cameras present certain advantages over other means for monitoring vehicle speed with object tracking (Chun & Li, 2000). Indeed, the task was here to develop a simple method for automatically estimating vehicle speed, a crucial variable in studies of traffic flow and the traffic character of freeways. Vehicle speed estimations on congested highways based on an un-calibrated camera have already been reported (e.g. Dailey et al., 2000). In this case, a mean vehicle dimension was used for scaling purposes, thus leading to estimates for time-averaged mean vehicle velocity. Contrary to this, our task here was to measure the speed of individual cars for obtaining detailed information on speed distribution. The present implementation of our approach – which extends previous work of Grammatikopoulos et al. (2002) – is limited to the un-congested case.

Given the time intervals between frames, estimation of speed is essentially a question of measuring distances in 1D. The video cameras used here are un-calibrated (i.e. camera constant, principal point location and image affinity parameters are irrelevant; on the contrary, radial lens distortion is taken into account). As regards exterior orientation, following basic assumptions are made: planar ground in front of a fixed camera; negligible image rotations about the vertical and camera axes. Thus, for 1D measurements on a straight road segment the only requirements are knowledge of the vanishing point in the road direction and one known distance on the road plane. This information allows removing perspective distortion by an affine rectification of the image sequence. Object segmentation and tracking are applied to the uniformly scaled frames for a more precise estimation of vehicle velocity. Since in the existing literature numerical results from algorithms for individual vehicle speed estimation are sparse, aim of this contribution is to present the mathematical model and the vehicle tracking technique, but also to assess their practical performance with sufficient measurements.

2. IMAGE RECTIFICATION

The basic geometric model, described in more detail in Grammatikopoulos et al. (2002), makes use of the fact that, once the image horizon of a plane is defined, the affine properties of this image can be recovered. The vanishing line I_∞ of the ground plane (horizon line) may be identified through the vanishing points in two orthogonal directions. The projective transformation between the image and the plane coordinates is expressed in homogeneous representation as $\mathbf{Hx} = \mathbf{X}$ (Hartley & Zisserman, 2000), whereby $\mathbf{H}$ is a 3x3 matrix with 8 independent coefficients, $\mathbf{x}$ is an image point and $\mathbf{X}$ a point on the ground plane. Knowledge of the third row of $\mathbf{H}$ – which represents I_∞ and is found from the cross product of two vanishing points – allows the removal of pure projectivity (i.e. affine rectification of the image). Orthogonality of the two directions restores angular relations. Yet, due to the 1D character of the studied problem (measurement of distances in a single direction), no need arises for correcting aspect ratio. Thus, the scale only along the road axis has to be restored, using one known ground distance in this direction.

2.1 Rectification with one vanishing point

Nevertheless, available lines do not suffice here for establishing two vanishing points, since parallel lines are identified along the road axis (Y-axis) but lane demarcations cannot be actually trusted as regards the orthogonal direction (X-axis). Grammatikopoulos et al. (2002) suggest an approach for cases where both vanishing points are finite, but only one of them is identified on the image. Here, however, the camera is assumed to have only a downward tilt (about its x-axis). Hence, lines orthogonal to the road axis are imaged parallel to the x-image axis, i.e. the second vanishing point F_2 is at infinity. On the image plane, vanishing points F_1 in the direction of the road axis and F_2 are then expressed as $F_1 = [f_1, f_2, 1]^T$ and $F_2 = [1, 0, 0]^T$.

Points F_1 and F_2 define I_∞, a line parallel to the image x-axis. The point at infinity in the direction orthogonal to the road axis is, thus, forced to be transformed through **H** to the point at infinity of the image x-axis. Ignoring the coefficients which concern pure translation, affine rectification **Hx** = **X** between the image (**x**) and the ground plane (**X**) is finally expressed as follows (Grammatikopoulos et al., 2002):

$$\begin{bmatrix} 1 & -f_1/f_2 & 0 \\ 0 & 1 & 0 \\ 0 & -1/f_2 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}$$

As already mentioned, the aspect ratio does not have to be corrected. A suitably chosen scale factor, however, will improve the visual appearance of the resulting rectified image. A measured distance along the road axis relates the pixel size of the rectified image to the ground units in this particular direction. Thus, for a vehicle assumed to travel in the direction of the road axis, measurement in successive frames of its position along the image y-axis allows the estimation of its speed.

2.2 Detection of vanishing point

In Grammatikopoulos et al. (2002) the vanishing point of the road direction was extracted manually. Here, on the contrary, image edges were first extracted automatically using the Canny edge detector (Fig. 1, left). Next, relying on an extension of the approach presented by Rother (2000), edges were grouped in directions with a vanishing point voting process, and finally the vanishing point in the direction of the road axis was estimated automatically with a least squares adjustment (Fig. 1, right).

Figure 1. Left: extracted edges using the Canny detector. Right: automatic estimation of the vanishing point in the road direction.

3. OBJECT TRACKING

In our previous work, measurement of corresponding vehicle points was made manually (Grammatikopoulos et al., 2002). In the present contribution, vehicles were recognized and followed automatically with the following process.

3.1 Vehicle detection

Blob detection was based on a simple background subtraction approach, in which an “empty” background image is generated from several frames by the iterative exclusion of outlying colour values.

Thus, for each pixel the RGB values from several frames are recorded, and the mean value g_m along with the standard deviation σ for all three channels are estimated. The background model includes all pixels with values within the range $g_m \pm \sigma$. Image pixels with even one colour value falling outside this range are regarded as originating from moving objects and excluded. Outlying values are discarded one by one, and the statistical test is repeated among the remaining pixel values until it is satisfied. Fig. 2 shows a background image generated from four original images.

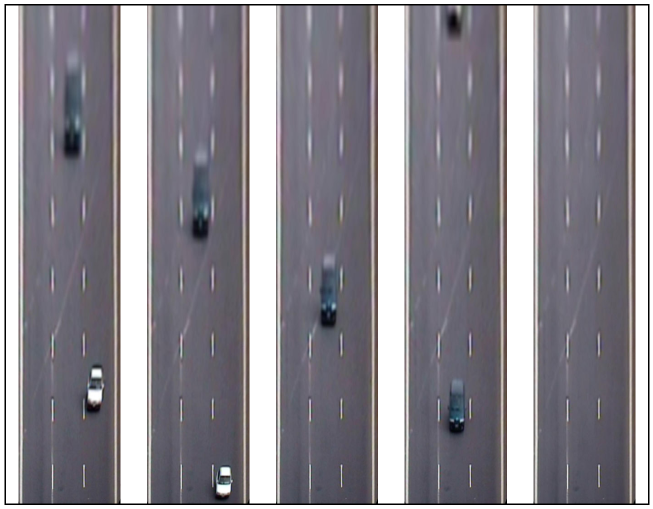

Figure 2. Creation of background image (far right)
from four successive rectified frames.

The generation of background image is repeated over time or can be updated frame-by-frame taking into account possible illumination changes due to shadows and weather conditions (more sophisticated tools may also be used to ensure higher accuracy, e.g. Stauffer & Grimson, 2000; Gutchess et al., 2001). Subtraction of the background image from a current frame results in an image containing only the foreground objects. This image is then binarized, and small holes within blobs are eliminated with the dilation and erosion operations (morphological closing). Finally, vehicles are identified using connected-component labelling. Blobs with areas below a threshold (noisy pixels) are omitted from the process. This vehicle detection process is performed for every successive frame. The steps of the described procedure are seen in Fig. 3

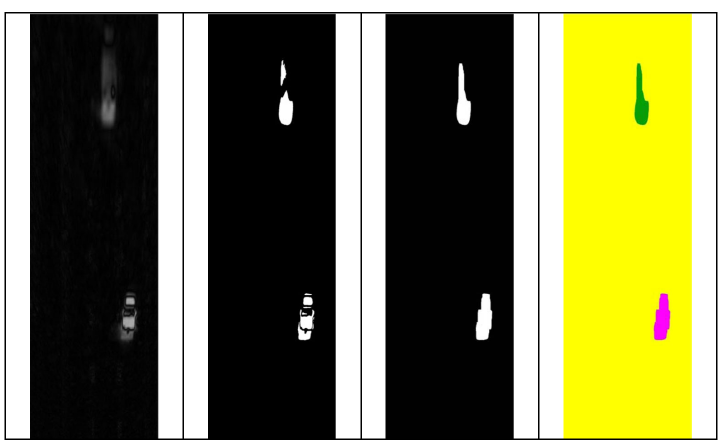

Figure 3. From left to right: foreground image
(subtraction of background image from rectified image);
binarization; morphological closing; detected blobs.

3.2 Vehicle tracking

Once blobs have been identified, tracking is performed for all detected objects within the sequence using normalized cross-correlation. Goal of this process is the accurate estimation of speed for each individual vehicle appearing in the sequence by computing its vertical displacement through all rectified frames. One known length in the road direction (e.g. the distance between two lane marks) suffices, as the only external information, for estimating Y-displacements of vehicles, as long as the scale in this direction has been restored. Most reported approaches for vehicle tracking compute the average velocity of the road lane by following the entire detected vehicle, or vehicle features, from frame to frame (e.g. Beymer et al., 1997; Dailey & Schoepflin, 2003). Such an estimation of speed, involving points of the vehicle which do not belong to the road plane, may be subject to a certain inaccuracy since, in a strict sense, only the image-to-ground relation has been restored.

Based on this consideration, our approach estimates displacement by tracking the lowest part of vehicles (which are closer to the ground plane) using normalized cross-correlation among frames. For each detected blob a rectangular image window is defined, containing sufficient information for the whole lower vehicle profile. The template width is adjusted according to the width of the profile, while its height depends on the resolution of the rectified images and an average vehicle length. For instance, in Fig. 4 (left) the height of the correlation window was set to 40 pixels. This area was tracked along the following frames until the vehicle exits. The number in blue gives the correlation coefficient; the number in yellow is the measured distance in the vertical direction.

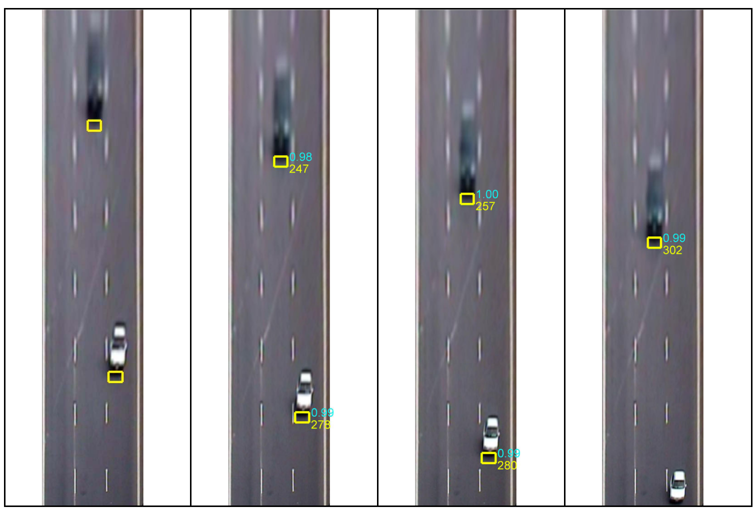

Figure 4. Tracking of detected profiles in three successive affine frames with cross-correlation.

It is remarked that, since image scale is uniform along the image y-axis, no need exists for window resizing, in contrast to the direct use of perspective imagery (Beymer et al., 1997). An additional advantage of our approach is the possibility of tracking cars, which either change lanes or travel large distances between frames. It is also worth mentioning that a lane mark may lead to mismatching if it is present in a correlation window. Therefore, cross-correlation is applied to subtracted foreground images rather than the initial rectified frames.

In its current implementation the presented algorithm cannot handle, partial or complete, vehicle occlusions. However, it could be extended on the basis of other more robust position-predicting methods (Melo et al., 2004), in order to ensure a more precise and generic estimation of vehicle movement over time.

4. Practical evaluation

Traffic flow was recorded from a bridge (10 m high) over a freeway with three lanes in each stream. The camera was looking centrally along the axis of the road with a downward tilt against the horizontal (one finite vanishing point with the other assumed at infinity). The performance of the automatic tracking and measuring algorithm was assessed against rigorous 2D projective transformation with sufficient ground control points based on manual image measurement. Frame dimensions were 768x576 pixels, and the frequency was known as 25 frames/sec. A total of 20 vehicles in all three lanes were tracked. Relying on one known distance, speed was estimated from 4–8 frames, yielding different estimates. Assuming constant speed for each vehicle, from these estimates a mean speed (v) and a standard deviation (σ) were calculated, as seen in Table 1.

Table 1. Speed estimation v ± σ (in km/h)
1D: automatic, affine images
2D: manual, metric images

1D		2D		1D		2D	
v	σ	v	σ	v	σ	v	σ
78	±1	79	±1	134	±2	133	±1
124	±1	124	±1	97	±1	98	±2
137	±2	136	±1	118	±2	118	±1
109	±1	107	±2	117	±2	117	±1
110	±2	110	±1	134	±1	134	±1
97	±1	97	±2	137	±2	136	±2
126	±3	127	±2	112	±1	111	±1
149	±3	148	±1	95	±2	95	±1
112	±1	111	±1	156	±3	155	±1
115	±1	115	±1	122	±2	122	±2

The above results clearly indicate that the automatic estimation of vehicle speed with the described tracking and measuring process, using frames subject to affine rectification, compares indeed well to the full 2D projective transformation based on manual image measurement. The overall algebraic mean difference between the two sets is 0.3 km/h and their RMS difference does not exceed 1 km/h. The uncertainty of individual measurements (σ) is below ±2 km/h. The validity of the 2D projective transformation had been checked against speed measurements with a GPS system on a moving car, and the differences of mean velocities were found below 1 km/h (Grammatikopoulos et al., 2002). Thus, the performed experiments indicate that the single image approach presented here may produce an estimated accuracy of about ±3 km/h.

5. Concluding remarks

The presented algorithm allows automatic 1D measurements and subsequent estimation of vehicle speed form single un-calibrated images. It requires minimal external data (one known distance), if basic assumptions are adopted. It has been also shown that this method, using one vanishing point, is capable of providing metric results of satisfactory accuracy. The accuracy of the approach will be further assessed with more data. Basic error propagation tests, however, have shown that – at least for this particular configuration – results do not appear as very sensitive to small errors. Of course, certain aspects need further investigation. The configuration with two finite vanishing points, for instance, is under implementation. Furthermore, the question of vehicle occlusion is obviously very important for a more generic approach of both un-congested and congested traffic.

Acknowledgements
This research has been financially supported by the Project of Basic Research 'Heraclitus', co-funded by the European Social Fund (75%) and National Resources (25%).

REFERENCES

Beymer, D., McLauchlan, P., Coifman, B., Malik, J., 1997. A real-time computer vision system for measuring traffic parameters. *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR '97)*, pp. 495-501.
Bose, B., Grimson, E., 2003. Ground plane rectification by tracking moving objects. *Proceedings of the Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*.
Chun, K.M., Li, C,K., 2000. Motions of multiple objects detection based on video frames. *International Symposium on Consumer Electronics (ICSE 2000)*, Hong Kong.
Dailey, D.J., Cathey, F.W., Pumrin, S., 2000. An algorithm to estimate mean traffic speed using un-calibrated cameras. *IEEE Trans. on Intelligent Transportation Systems*, 1:98-107.
Dailey, D.J., Schoepflin T.N., 2003. A correlation technique for estimating traffic speed from cameras. *Transportation Research Board Annual Meeting*, TRB no. 03-3414.
Grammatikopoulos, L., Karras, G., Petsa, E., 2002. Geometric information from single uncalibrated images of roads. *International Archives of Photogrammetry & Remote Sensing*, 34(5):21-26.
Hartley, R., Zisserman, A., 2000. *Multiple View Geometry in Computer Vision*. Cambridge University Press.
Gutchess D., Trajkovics M., Cohen-Solal E., Lyons D., Jain A.K, 2001. A background model initialization algorithm for video vurveillance. *Proc. IEEE International Conference on Computer Vision (ICCV '01)*, Part 1, pp. 744-740.
Kastrinaki, V., Zervakis, M., Kalaitzakis, K., 2003. A survey of video processing techniques for traffic applications. *Image and Vision Computing*, 21:359-381.
Melo, J., Naftel, A., Bernardino, A., Santos-Victor J., 2004. Viewpoint independent detection of vehicle trajectories and lane geometry from un-calibrated traffic surveillance cameras. *International Conference on Image Analysis and Recognition*, Springer, LNCS 3212, pp. 454-462.
Rother C., 2000. A new approach for vanishing point detection in architectural environments. *Proc. 11th British Machine Vision Conference (BMVC 2000)*, pp. 382–391.
Simond, N., Rives, P., 2003. Homography from a vanishing point in urban scenes. *Proc. International Conference on Intelligent Robots and Systems*.
Stauffer, C., Grimson, E., 2000. Learning patterns of activity using real-time tracking. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22:747-757.
Tao, C.V., El-Sheimy N., 2000. Highway mobile mapping. *GIM International*, 14(10):81-85.

Authors:
Lazaros Grammatikopoulos (PhD Student)
Dept. of Surveying, National Technical University of Athens (NTUA), GR-15780 Athens, Greece
Tel. (30) 210 7722683, Fax. (30) 210 7722677, E-mail: lazaros@central.ntua.gr
George Karras (Associate Professor)
Dept. of Surveying, National Technical University of Athens (NTUA), GR-15780 Athens, Greece
Tel. (30) 210 7722685, Fax. (30) 210 7722677, E-mail: gkarras@central.ntua.gr
Elli Petsa (Professor)
Dept. of Surveying, Technological Educational Institute of Athens (TEI-A), GR-12210 Athens, Greece
Tel. (30) 210 5385343, Fax. (30) 210 5385366, E-mail: petsa@teiath.gr