Real-time Vehicle Matching for Multi-camera Tunnel Surveillance

Vedran Jelača, Jorge Oswaldo Niño Castañeda, Andrés Frías-Velázquez,
Aleksandra Pižurica and Wilfried Philips

ABSTRACT

Tracking multiple vehicles with multiple cameras is a challenging problem of great importance in tunnel surveillance. One of the main challenges is accurate vehicle matching across the cameras with non-overlapping fields of view. Since systems dedicated to this task can contain hundreds of cameras which observe dozens of vehicles each, for a real-time performance computational efficiency is essential. In this paper, we propose a low complexity, yet highly accurate method for vehicle matching using vehicle signatures composed of Radon transform like projection profiles of the vehicle image. The proposed signatures can be calculated by a simple scan-line algorithm, by the camera software itself and transmitted to the central server or to the other cameras in a smart camera environment. The amount of data is drastically reduced compared to the whole image, which relaxes the data link capacity requirements. Experiments on real vehicle images, extracted from video sequences recorded in a tunnel by two distant security cameras, validate our approach.

Keywords: Object recognition, feature extraction, tunnel surveillance, traffic monitoring

1. INTRODUCTION

Tunnels are environments prone to horrific traffic accidents. To enable timely actions that can save lives and minimize the damage, it is important to track vehicles throughout a tunnel. For this purpose, multiple surveillance cameras are typically mounted along tunnels, often with non-overlapping fields of view. As an assistance to human operators, computer vision algorithms can then be used for automatic detection and tracking of vehicles in tunnels. Such algorithms consist of three parts: vehicle detection, tracking of vehicles in a field of view of one camera and vehicle matching, which is used for a "handover" of vehicles between cameras. A real-time performance of these algorithms is crucial. In this context, this paper addresses the problem of real-time matching of vehicles as they are imaged by stationary surveillance cameras with non-overlapping views.

In our vehicle matching problem, to each detected vehicle from camera C_n we assign the corresponding vehicle from camera C_{n-1}. The vehicle to which the corresponding match is to be assigned is further on denoted as "template" and vehicles considered as possible matches are denoted as "candidates". For each template we define a set of candidates according to road constrains, inter camera distances and vehicle kinematics. A template-candidate assignment is then obtained based on a similarity measure between their appearances. In this paper we focus on the problem of vehicle appearance matching and propose a computationally efficient method for this purpose. Even though all cameras view the vehicles from the rear side, vehicle appearance matching is challenging due to the low resolution of surveillance cameras, poor lighting conditions in tunnels and various changes of the vehicle appearance. These changes are mainly due to illumination changes of the environment (e.g. a different lighting in different areas of the environment) and changes of the vehicle pose as it moves through the multi-camera environment. The motion blur and noise in the images impose an additional challenge for extraction of informative features that efficiently represent vehicle appearance. Fig. 1 shows images of five vehicles acquired by two cameras. The images illustrate a variety of matching challenges (scale difference, bounding box shifts, viewing angle and illumination changes).

Most of the previous work on object appearance matching[1–7] uses appearance representations based on color information (e.g. mean, histogram or correlogram of the color). Color alone is, however, not reliable feature in

V. Jelača, J. O. Niño Castañeda, A. Frías-Velázquez, A. Pižurica and Wilfried Philips are with Ghent University, Dept. of Telecommunications and Information Processing, TELIN-IPI-IBBT, Sint-Pietersnieuwstraat 41, Ghent, Belgium
{vedran.jelaca,jorge.nino,andres.friasvelazquez,aleksandra.pizurica,philips}@telin.ugent.be

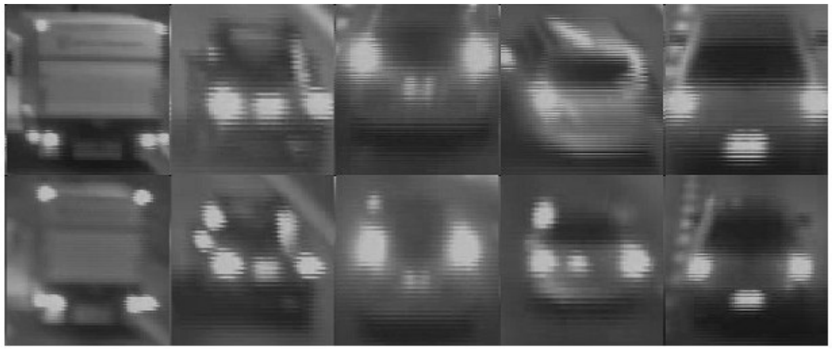

Figure 1. Examples of vehicle images from our database. Columns contain the same vehicles observed by two cameras along a tunnel pipe. Scale difference, shifts, viewing angle and illumination changes are present.

many traffic surveillance applications, especially in tunnels. Appearance representations that do not need color information are often based on eigenimages (mostly used for face recognition[8]) and local invariant features (e.g. SIFT[9] or SURF[10]). Methods based on *eigenimages* require offline training and their accuracy highly depends on variations of appearances present in the training set. Therefore, adaptation of these methods to appearance changes is limited. Accuracy of methods based on *local features* depends on the number of similar feature points found in compared images. Calculation of these features is computationally demanding and it is still difficult to achieve real-time performance when there are multiple objects that have to be compared simultaneously.

Our method is inspired by the work of *Betke et al.*,[11] which used vertical and horizontal projections of a vehicle edge map for accurate positioning of the bounding box in tracking. We go a step further, showing that it is possible to use projections for vehicle appearance representation and matching. Instead of using projections of the vehicle edge map, we represent vehicle appearances by projections of the vehicle images themselves. Such projections we call vehicle signatures. We use horizontal, vertical and two diagonal signatures of the grey value vehicle images, without any pre-processing step. In this way we save computational time because it is not necessary to calculate edge maps. Matching of the signatures is obtained by a simple combination of 1-D correlations. Finally, since each vehicle has one and only one correct match, we employ the Hungarian algorithm[12] for resolving ambiguities and matching optimization.

The remainder of the paper is organized as follows. Section 2 introduces the appearance model based on the vehicle signatures while the appearance matching and a template-candidate assignment are explained in Section 3. In Section 4 an overview of the whole matching algorithm is given, followed by the experimental results in Section 5. Finally, we conclude the paper in Section 6.

2. VEHICLE APPEARANCE MODEL

Let I be the vehicle image of size $M \times N$. We define the signatures of the image I as Radon transform like projection profiles along a certain direction. The vertical signature $\mathbf{v}_I$ consists of the arithmetic means of the pixel intensities in each image row,

$$v_I(m) = \frac{1}{N}\sum_{n=1}^{N} I(m,n), \quad m = 1, ..., M, \tag{1}$$

where $I(m,n)$ is an intensity value of the image pixel at the position (m,n). Analogously, the components of the horizontal signature $\mathbf{h}_I$ are the arithmetic means of the pixel intensities in each column of the image I,

$$h_I(n) = \frac{1}{M}\sum_{m=1}^{M} I(m,n), \quad n = 1, ..., N. \tag{2}$$

In Fig. 2 we see two images of the same vehicle observed by two grey scale cameras. Both images are represented by horizontal and vertical signatures. If we plot two horizontal signatures one over the other (the

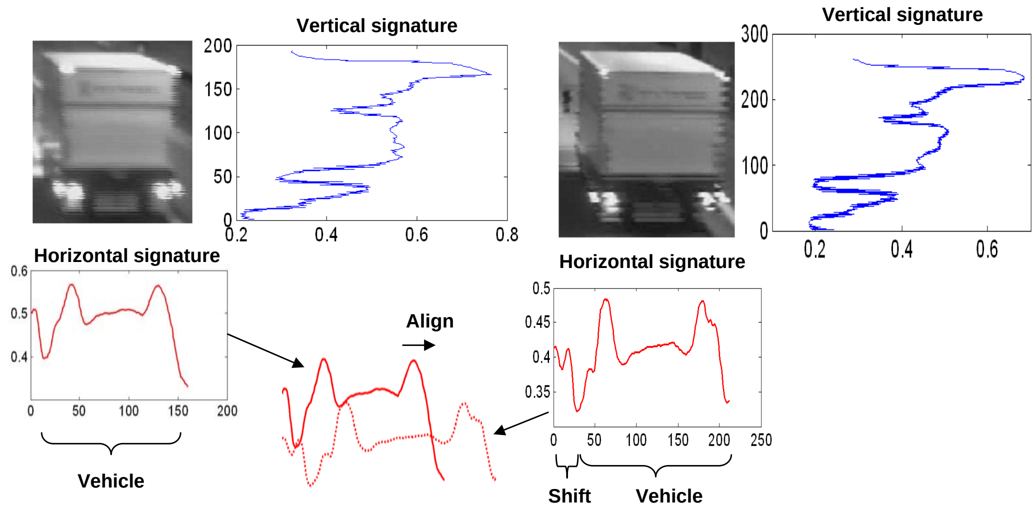

Figure 2. Vehicle images from two different cameras along a tunnel pipe with the corresponding horizontal (below the vehicle image) and vertical (on the right side of the vehicle image) signatures. There is a clear similarity in behaviour of the signature parts which correspond to the vehicle. The signatures can be shifted due to the bounding box misalignment.

signatures at the bottom of Fig. 2), we see they are quite similar apart from a shift. The characteristic parts, strong backlights in this case, are reflected in dominant peaks in both signatures. Since vehicles are rigid objects, the spatial relation between their different parts is preserved in all their observations.

Next to the vertical and horizontal signatures, we can use additional projections. We define the n-dimensional vehicle signature vector $\mathbf{s}_I$ calculated from the vehicle image I as an n-tuple of n projections (signatures) on different lines. In this paper we explicitly treat 2-D and 4-D signature vectors defined as following. The 2-D signature vector is a pair of the vertical and horizontal signature,

$$\mathbf{s}_I = (\mathbf{v}_I, \mathbf{h}_I), \tag{3}$$

while a 4-D signature vector contains also two diagonal signatures (see Fig. 4),

$$\mathbf{s}_I = (\mathbf{v}_I, \mathbf{h}_I, \mathbf{d}_I, \mathbf{a}_I), \tag{4}$$

where $\mathbf{d}_I$ and $\mathbf{a}_I$ are signatures on the main-diagonal and anti-diagonal, respectively. The signature vectors represent an image as multiple 1-D vectors, which significantly reduces the amount of the vehicle appearance representation data.

3. VEHICLE APPEARANCE MATCHING

3.1 Signature matching

The challenges for the signature matching come from scale difference, shift and viewing angle variations. Scale differences result from different distances between the vehicle and the camera, shifts result from different positions of the bounding boxes while viewing angle differences occur due to road bendings and lane changes between observations. Fig. 2 illustrates some of these effects and their influence on the vehicle signatures.

We remove the influence of scale by normalizing the signatures to the same length. In practice, we used for this the nearest neighbour interpolation for computation simplicity. After the normalization, to remove the influence of signature shifts, we align the signatures by correlating them in different shift positions. The position in which the correlation coefficient is maximal is their alignment position. Fig. 3 illustrates some possible cases of signature shifts. Since the cameras observe vehicles from behind, the rear part of the vehicle is the most informative and present in all observations. However, the rear part can be positioned in any part of the

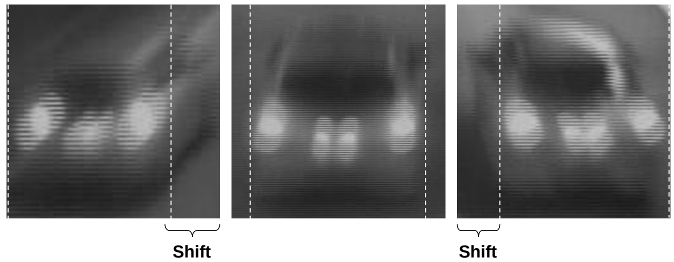

Figure 3. Three images of the same vehicle viewed from a different angle. Due to the viewing angle difference, there is a shift of the rear part of the vehicle.

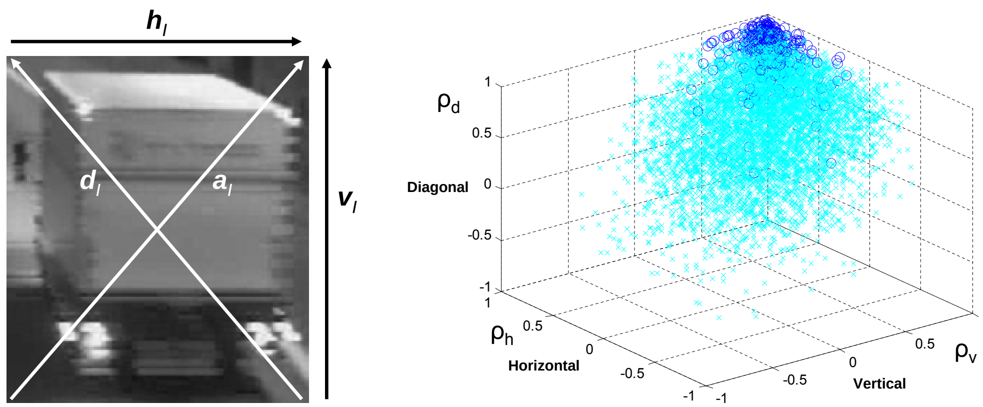

Figure 4. Left: Horizontal, vertical and two diagonal signatures are obtained as an arithmetic mean of image pixel intensities along the shown lines and their parallels, in the direction of arrows; Right: 3-D scatter plot of the correlation coefficients between the signatures from different vehicle pairs; circles represent the values for the same vehicles while crosses represent the values for different vehicles.

image (see Fig. 3), depending on the viewing angle and the bounding box shift. Therefore, when matching two signatures, we extract the first, the middle and the last part of template signature and shift it along candidate signature to find their alignment.

Let $\mathbf{x}$ and $\mathbf{y}$ be two signatures of length N. Further, let $\mathbf{x}_P$ be a part of $\mathbf{x}$, obtained by extracting P points from the signature $\mathbf{x}$.

The part $\mathbf{x}_P$ is shifted along the signature $\mathbf{y}$ and in each position $s \in [0, N-P]$ the correlation coefficient $\rho(s)$ is obtained:

$$\rho(s) = \frac{\sum_{i=1}^{P}(x_P(i) - \bar{\mathbf{x}}_P)(y(i+s) - \bar{\mathbf{y}}_s)}{\sqrt{\sum_{i=1}^{P}(x_P(i) - \bar{\mathbf{x}}_P)^2 (y(i+s) - \bar{\mathbf{y}}_s)^2}}, \tag{5}$$

Then, the matching measure between the signatures $\mathbf{x}$ and $\mathbf{y}$ is defined as the maximal correlation coefficient from all shift positions:

$$\rho_{\mathbf{xy}} = \max_{s} \rho(s). \tag{6}$$

This is a similarity measure of two signatures.

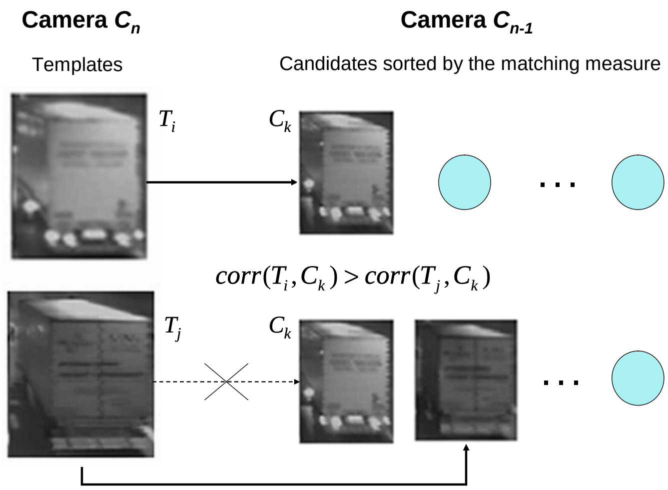

Figure 5. Match association. Example on two vehicles. Using the Hungarian algorithm an optimal assignment with minimal total cost (maximal sum of individual similarity measures) is made. Matching optimization does not allow multiple matches with one vehicle. In the given example, template T_i is matched to candidate C_k due to a higher matching measure between T_i and C_k than between template T_j and C_k. This leads to a correct matching of template T_j, too.

3.2 Matching of signature vectors

Matching of vehicle appearance models (the signature vectors) is done by combining the matching measures of their corresponding components. These are the correlation coefficients from Eq. 6, calculated separately for the signatures in different projection directions. Fig. 4 shows 3-D scatter plot of the correlation coefficients between the horizontal, vertical and main-diagonal signatures of 300 vehicles imaged by two distant security cameras. The vehicles were manually annotated for the evaluation purpose. Dark blue circles and light blue crosses in the scatter plot represent the correlation coefficients for the same and different vehicles, respectively. We can see that the results for the same vehicles are clearly clustered, with high correlation coefficient for all three signatures. Based on this observation, we define one matching measure between each two vehicle appearances as the Euclidian norm of an n-D vector, where each of n dimensions represents the correlation coefficient between two corresponding signatures (horizontal, vertical or two diagonal).

3.3 Template-candidate assignment

The set of candidates for each template is determined according to physical constrains of the environment and vehicle motion. Suppose that we observe vehicles by two successive cameras C_n and C_{n-1} and let $D_{n,n-1}$ be the distance between the two cameras. Suppose further that ν_j^n is a velocity of the template vehicle T_j, measured in the field of view of camera C_n at the time instance t. We measure the velocity according to the lane marks on the road. Let ν_{min} and ν_{max} be the minimal and maximal allowed vehicle velocities taking into account possible down and over speeding (these velocities can be determined relative to ν_j^n or in absolute values). Then, all vehicles that exited the field of view of camera C_{n-1} between time instances $t - \frac{D_{n,n-1}}{\nu_{min}}$ and $t - \frac{D_{n,n-1}}{\nu_{max}}$ are considered as matching candidates for the template T_j.

Each template is compared with all its candidates using the proposed appearance model and the matching procedure described in Section 3.2. From the matching measures the assignment cost matrix is obtained, where each matrix element ρ_{ij} is the matching measure between template T_i and candidate C_j. Since every template has one and only one correct match, we optimize the template-candidate assignment by using the Hungarian algorithm,[12] see Fig. 5. Note that only the template-candidate pairs that have positive correlation coefficients between all their signatures (vertical, horizontal and two diagonal) enter into the optimization process. The vehicle pairs that have at least one negative correlation coefficient between their signatures are immediately classified as "different vehicles".

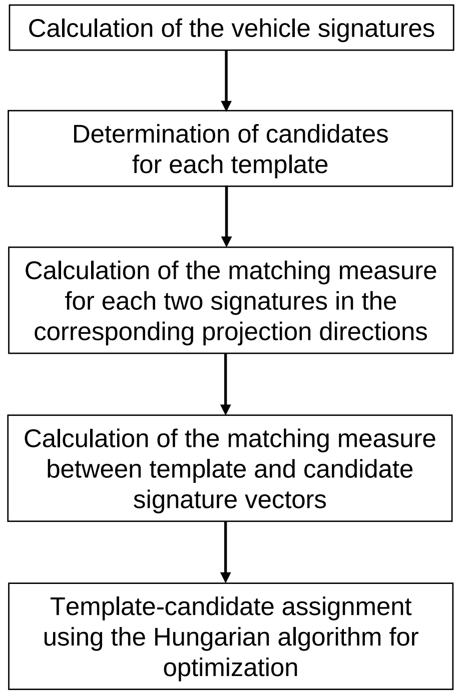

Figure 6. The matching algorithm, block diagram.

4. VEHICLE MATCHING ALGORITHM SUMMARY

The proposed vehicle matching algorithm is summarized in Fig. 6, with the following five modules:

1. Computation of vehicle signatures as explained in Section 2. This is done only once per each vehicle image.
2. Determination of candidates for each template as explained in Section 3.3.
3. Calculation of the similarity measure between each two template and candidate signatures in the corresponding projection directions, Eq. 6.
4. Calculation of the similarity measure between the template and candidate signature vectors (Section 3.2).
5. Optimal template-candidate assignment by Hungarian algorithm (Section 3.3).

5. EXPERIMENTAL RESULTS

We collected a database of 300 vehicle images from two distant security cameras mounted roughly in the center of a tunnel pipe ceiling and oriented in the direction of the traffic flow. The vehicles in the database were manually annotated for the evaluation purpose. Fig. 1 shows some examples of vehicle images in the database. All aforementioned matching challenges, explained in Section 3.1 are present: scale difference, vehicle shifts, viewing angle and illumination changes.

We have compared the matching score computed between the vehicles in our database using our matching algorithm, with three other appearance matching methods: 2-D image correlation, SIFT- and eigenimage-based method. 2-D image correlation was obtained using vehicle images normalized to the same size. In the SIFT-based method,[9] vehicle matching was done using the kd-tree and nearest neighbour search. For the eigenimage method,[8] the datasets were divided in two disjunct parts, the training and testing subset (in both databases 100 images are taken for training and tests were then performed on the other 200 images). The Hungarian algorithm was used to optimize the assignment in all methods.

Table 1. Experimental results

Method	Correct matching [%]
2-D correlation	84
Eigenimages	90
SIFT	89
Our method - two signatures	88
Our method - four signatures	91

In the experiments with our method, for the purpose of signature alignment (Section 3.1) we extracted the first, the middle and the last P points from each template signature ($P = 120$ of total $N = 150$ signature points), see Fig. 3. Then the template signature parts were correlated separately with the candidate signatures to obtain their best alignment and their matching measure (Eq. 5 and 6). The final matching measure between the signatures was calculated as the maximal of these three correlation coefficients.

The experimental results in Table 1 show that our signature based appearance matching performs similarly or slightly better than the reference methods while being much more efficient both in terms of data storage and computations. The complexity of our method is $O(N)$ compared to $O(N^2)$ for 2-D correlation, while SIFT and eigenimage based methods are even more complex than 2-D correlation. The inferior accuracy of 2-D correlation is due to its sensitivity to the vehicle misalignments whereas our method and the methods based on SIFT and eigenimages have certain robustness against such misalignments. In our method this is due to the correlation with shifting step.

The computational efficiency of our method is its essential advantage comparing to the other methods. In our C++ implementation, computation of the four signatures and the similarity measure for two vehicle images was achieved in 9.3 ms on a single-core 1.86 GHz CPU. Such efficiency allowed us to match in 27.2 seconds all 300 vehicles viewed by two cameras in a period of 8 minutes.

We also see that using the diagonal signatures in the appearance model increases matching accuracy. However, matching of only horizontal and vertical signatures is about three times computationally more efficient and requires three times less data than when all four signatures are used. Thus, using only horizontal and vertical signatures is a good solution in processing time and memory critical cases.

6. CONCLUSION

By using vehicle signatures composed of Radon transform like projection profiles, it is possible to significantly reduce the amount of data needed for vehicle matching. When the signatures are normalized to the same length and compared using 1-D correlation with shifting, scale and limited shift and viewing angle invariance are obtained. We showed that signature matching can reach or improve the accuracy of matching the whole images and it is yet computationally much more efficient. Hence, the proposed method can be used for the real-time vehicle matching on embedded systems (e.g. smart cameras) or by a central server without a need for sending the images between the cameras or to the server. This reduces requirements for communication links capacities. Additional improvement can be obtained by using multiple appearances per vehicle, acquired at appropriate time instants during the tracking, which is a subject of our ongoing research.

Acknowledgements

This research has been supported by the Flemish Interdisciplinary Institute for Broadband Technology (IBBT).

REFERENCES

[1] Huang, T. and Russell, S., "Object identification in a Bayesian context," *Proceedings of International Joint Conferences on Artificial Intelligence* (1997).

[2] Porikli, F., "Inter-camera color calibration using cross-correlation model function," *International Conference on Image Processing* (2003).

[3] Rahimi, A., Dunagan, B., and Darrell, T., "Simultaneous calibration and tracking with a network of non-overlapping sensors," *International Conference on Computer Vision and Pattern Recognition* (2004).

[4] Kettnaker, V. and Zabih, R., "Bayesian multi-camera surveillance," *International Conference on Computer Vision and Pattern Recognition* (1999).

[5] Javed, O., Rasheed, K., Shafique, M., and Shah, M., "Tracking across multiple cameras with disjoint views," *International Conference on Computer Vision* (2003).

[6] Javed, O., K., S., and Shah, M., "Appearance modelling for tracking in multiple non-overlapping cameras," *International Conference on Computer Vision and Pattern Recognition* (2005).

[7] Choe, T., M., L., and Hearing, N., "Traffic analysis with low frame rate camera networks," *First IEEE Workshop on Camera Networks* (2010).

[8] Turk, M. and Pentland, A., "Eigenfaces for recognition," *Journal of Cognitive Neuroscience* (1991).

[9] Lowe, D., "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision* **60, No. 2**, 91–110 (2004).

[10] Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L., "Surf: Speeded up robust features," *Computer Vision and Image Understanding* **110, No. 3**, 346–359 (2008).

[11] Betke, M., Haritaoglu, E., and Davis, L. S., "Real-time multiple vehicle detection and tracking from a moving vehicle," *Machine Vision and Applications, Springer-Verlag* **12**, 69–83 (2000).

[12] Munkres, J., "Algorithms for the assignment and transportation problems," *Society for Industrial and Applied Mathematics* (1957).