Detectability of Moving Objects Using Correspondences over Two and Three Frames

Jens Klappstein, Fridtjof Stein, and Uwe Franke

DaimlerChrysler AG
71059 Sindelfingen, Germany

Abstract. The detection of moving objects is crucial for robot navigation and driver assistance systems. In this paper the detectability of moving objects is studied. To this end, image correspondences over two and three frames are considered whereas the images are acquired by a moving monocular camera. The detection is based on the constraints linked to static 3D points. These constraints (epipolar, positive depth, positive height, and trifocal constraint) are discussed briefly, and an algorithm incorporating all of them is proposed. The individual constraints differ in their action depending on the motion of the object. Thus, the detectability of a moving object is influenced by its motion. Three types of motions are investigated: parallel, lateral, and circular motion. The study of the detection limits is applied to real imagery.

1 Introduction

Robots and autonomous vehicles require the knowledge about objects moving in the scene in order to avoid collisions with them. Beside radar and lidar sensors also cameras can be utilized to observe the 3D scene in front of the vehicle. In this paper up to three images taken by a moving monocular camera are evaluated. Since we do not know a priori where moving objects are in the scene we cannot check for them directly. However, given the optical flow (image correspondences) and the ego-motion we are able to triangulate the viewing rays yielding reconstructed 3D points. If the 3D point is actually a static point the reconstruction will be fine, but if the actual 3D point is moving the reconstruction will fail (in general). What does this mean?

A reconstructed 3D point has to fulfill certain constraints in order to be a valid static 3D point. If it violates any of them the 3D point is not static, hence it must move. Thus, the detection of moving objects is based on the constraints a static point fulfills.

Although many constraints exist, there are some kinds of motion which (nearly) fulfill all constraints and thus are not detectable. This paper investigates these detection limits, and is organized as follows: At first (section 2) the available constraints are discussed. In section 3 an error metric is developed combining all constraints. Based on this metric the detection limits are investigated in section 4. Experimental results are given in section 5.

F.A. Hamprecht, C. Schnörr, and B. Jähne (Eds.): DAGM 2007, LNCS 4713, pp. 112–121, 2007.

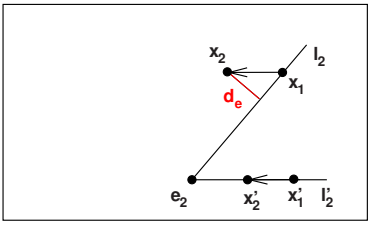


Fig. 1. Epipolar constraint. The image of the second view is shown. The camera moves along its optical axis. An object moves lateral w.r.t. the camera inducing an horizontal optical flow shown by the correspondences $x_1 \leftrightarrow x_2$ and $x'_1 \leftrightarrow x'_2$. The subscripts 1 and 2 denote entities in the first and the second view, respectively. x_2 does not lie on the epipolar line l_2 inducing the epipolar error d_e. x'_1 moves along its epipolar line l'_2 and thus fulfills the epipolar constraint. e_2 is the epipole.

Please note that the ego-motion, i.e. the motion of the camera from frame to frame, must be known in order to perform the detection. Furthermore, the location of the camera with respect to the road (ground plane) is required. The information is considered as given here. Specifically, it is assumed that the fundamental matrix, the road homography between the first two views, and the trifocal tensor are given.

The reader is referred to [1,3,7] which address the estimation of the ego-motion. Beside these two-view methods one can estimate the ego-motion over all three views [9].

2 Constraints for Static 3D Points

In this section we discuss briefly the constraints a static 3D point fulfills. On the basis of traffic scenarios we will see how each constraint acts on different kinds of motion. Thereby we differentiate between parallel motion (preceding and overtaking objects), and lateral motion (crossing objects).

The first three constraints, discussed in detail in [5], apply for correspondences over two frames. The fourth constraint is applicable if correspondences over three views are available. Each individual constraint raises the quality of detection.

- **Epipolar Constraint**
  The epipolar constraint expresses that the viewing rays of a static 3D point (the lines joining the projection centers and the 3D point) must meet. A moving 3D point in general induces skew viewing rays violating the constraint. Figure 1 illustrates it.
- **Positive Depth Constraint**
  The fact that all points seen by the camera must lie in front of it is known as the positive depth constraint. It is also called cheirality constraint. If viewing rays intersect behind the camera, as in figure 2a, the actual 3D point must be moving.

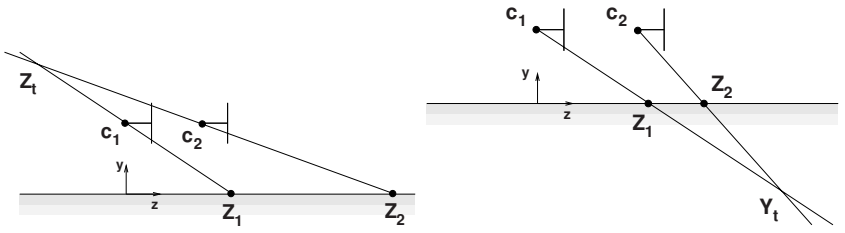


Fig. 2. Side view: Positive depth (a) and positive height (b) constraint. The camera is moving from c_1 to c_2. A 3D point on the road is moving from Z_1 to Z_2. In (a) the traveled distance of the point is greater than the distance of the camera (overtaking object). The triangulated 3D point Z_t lies behind the camera, violating the positive depth constraint. In (b) the traveled distance of the point is smaller (preceding object). The triangulated 3D point Y_t lies underneath the road, violating the positive height constraint.

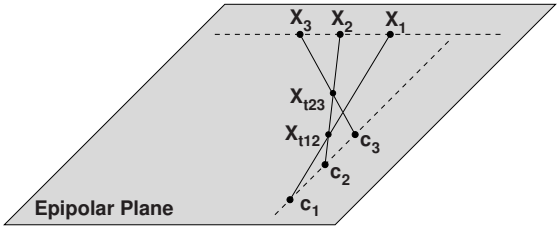


Fig. 3. Trifocal Constraint. The camera observes a lateral moving 3D point (X_1 to X_3) while moving itself from c_1 to c_3. The triangulated point of the first two views is X_{t12}. The triangulation of the last two views yields X_{t23} which does not coincide with X_{t12} violating the trifocal constraint.

- **Positive Height Constraint**
  All 3D points must lie above the road. If viewing rays intersect underneath the road, as in figure 2b, the actual 3D point must be moving.
- **Trifocal Constraint**
  A triangulated 3D point utilizing the first two views must triangulate to the same 3D point when the third view comes into consideration. This constraint is also called trilinear constraint. In figure 3 it is violated.

3 Error Metric Combining All Constraints

With the constraints described above, the objective is to measure quantitatively to which extent these constraints are violated. The resulting measurement function, called error metric, shall be correlated to the likelihood that the point is moving, i.e. higher values indicate a higher probability.

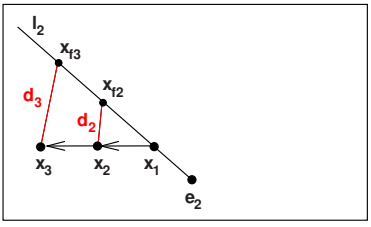


Fig. 4. Combined error metric. The image of the second view is shown. The camera moves along its optical axis observing a lateral moving point $x_1 \leftrightarrow x_2 \leftrightarrow x_3$. The closest point to x_2 fulfilling the two-view constraints is x_{f2}. The error arising from two-views is the distance d_2. Transferring the points x_1 and x_{f2} into the third view yields x_{f3}. If the observed 3D point was actually static its image x_3 would coincide with x_{f3}. However, the 3D point is moving which causes the trifocal error d_3. The overall error is $d = d_2 + d_3$. Note, that in general x_1 and x_{f3} do not lie on the epipolar line l_2.

The error metric is developed in two steps. First, the two-view constraints are evaluated taking view one and two into account. Afterwards, the trifocal constraint is evaluated using the third view, too.

3.1 Two-View Constraints

An error metric combining the two-view constraints has been introduced in [5]. It measures the distance of a given image point in the first view to the closest point fulfilling all constraints (epipolar, positive depth, and positive height constraint). For the ease of computational complexity image points in the second view are considered noise free. We use this metric here but swap the roles of the views, i.e. we compute the error (distance) in the second view. This is illustrated in figure 4.

We first consider the correspondence $x_1 \leftrightarrow x_2$ in the views one and two. The closest point to x_2 fulfilling the two-view constraints is x_{f2}. It lies on the epipolar line $l_2 = Fx_1$ with F the fundamental matrix. Note that the vector from x_{f2} to x_2 is not necessarily perpendicular to l_2. The distance d_2 between x_{f2} and x_2 is the error arising from the first two views. For the computation of d_2 see [5].

3.2 Three-View Constraint

We now add the third view and consider the correspondence $x_1 \leftrightarrow x_2 \leftrightarrow x_3$. As the point x_{f2} is defined such that it fulfills the two-view constraints the reconstructed 3D point arising from the triangulation of the points x_1 and x_{f2} constitute a valid 3D point. This 3D point is projected into the third view yielding x_{f3}. The measured image point x_3 will coincide with x_{f3} if the observed 3D point is actually static. Otherwise there is a distance d_3 (figure 4) between them which we call trifocal error. x_{f3} is computed via the point-point-point transfer using the trifocal tensor [2]. This approach avoids the explicit triangulation of the 3D point.

The overall error combining the two-view constraints and the three-view constraint is $d = d_2 + d_3$. It measures the minimal required displacement in pixels necessary to change a given correspondence into a correspondence belonging to a valid static 3D point.

4 Detection Limit

In this section we deal with the key question: Utilizing the different constraints, which kinds of motion are detectable and to which extent? In order to detect a moving object reliably the error metric developed in section 3 must be greater than a certain threshold T, whereas the threshold should reflect the noise in the correspondences (optical flow). A reasonable choice is $T = 3\sigma$ with σ the standard deviation of the correspondences.

In the following we consider the three most frequent kinds of motion in traffic: parallel, lateral and circular motion. We model the motion of the camera and the object as shown in figure 5. It is not necessary to investigate camera rotations about its projection center, since they do not influence the detection limit. One can always compensate these rotations by a virtual inverse rotation.

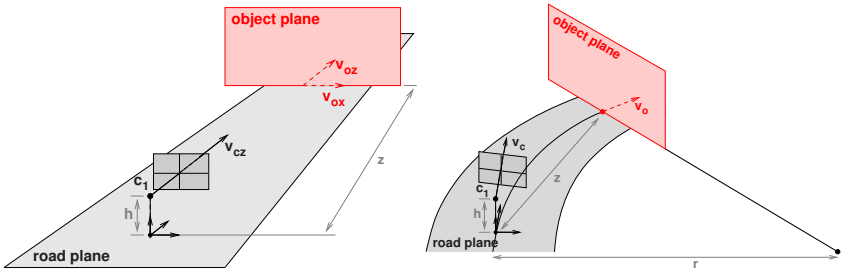


Fig. 5. Motion model utilized for the investigation of the detection limit. The cameras projection center in the first view is c_1. The moving object is modeled as a plane. (a) Linear motion: The (object)plane moves parallel (w.r.t. the camera) with speed v_{oz} and lateral with speed v_{ox}. The distance of the camera to the object is z, to the road it is h. The camera moves along its optical axis with speed v_{cz}. (b) Circular motion: Both, camera and object, move along a circle with radius r. The tangential speed of the camera is v_c, that of the object is v_o.

4.1 Linear Motion

The detection limits for the linear motions (parallel and lateral motion) are illustrated by means of three examples:

1. Overtaking object: The object moves parallel to the camera but faster. $v_{cz} = 30\text{km/h}$, $v_{oz} = 40\text{km/h}$, $v_{ox} = 0\text{km/h}$

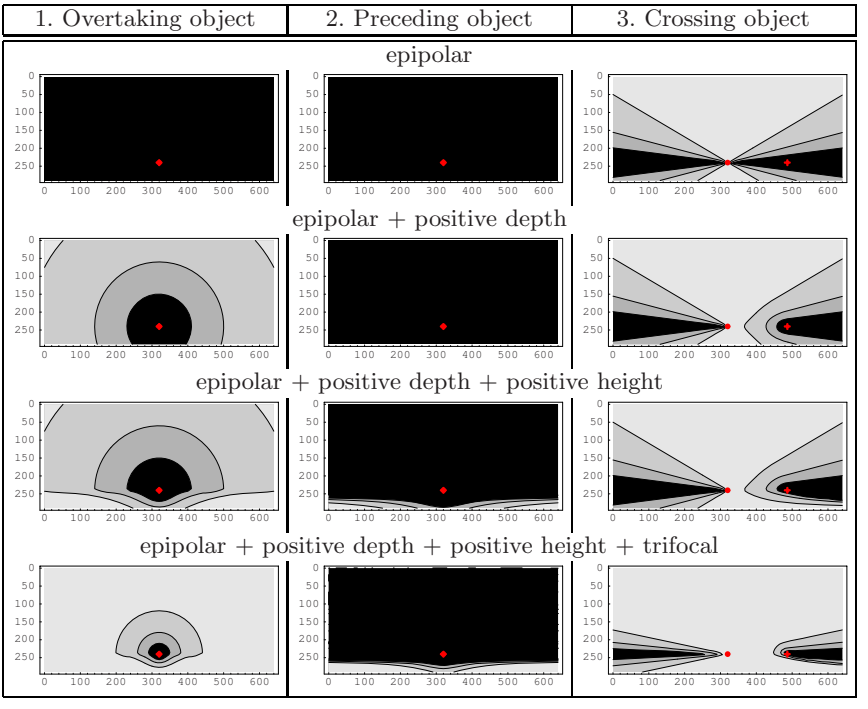


Fig. 6. Detection limits for different kinds of linear motion and constraints. The images show the first view (compare to fig. 5). They are truncated at row 290, since below there is no object but the road. Inside the black regions the motion is not detected. The contour lines $2T$ and $4T$ are also shown. The red point marks the epipole, the red cross is the point of collision. Further explanation is given in the text.

2. Preceding object: The object moves parallel to the camera but slower.
   $v_{\mathrm{cz}} = 30\mathrm{km/h}$, $v_{\mathrm{oz}} = 20\mathrm{km/h}$, $v_{\mathrm{ox}} = 0\mathrm{km/h}$
3. Crossing object: The object moves lateral to the camera.
   $v_{\mathrm{cz}} = 30\mathrm{km/h}$, $v_{\mathrm{oz}} = 0\mathrm{km/h}$, $v_{\mathrm{ox}} = -5\mathrm{km/h}$

The subscripts stand for: c = camera, o = object, z = longitudinal direction, x = lateral direction. Anti-parallel motion ($v_{\mathrm{cz}} > 0\mathrm{km/h}$, $v_{\mathrm{oz}} < 0\mathrm{km/h}$, $v_{\mathrm{ox}} = 0\mathrm{km/h}$) is not of interest here, since it is completely not detectable [4]. In the examples other important parameters are: focal length $f = 1000\mathrm{px}$, principal point $(x_0, y_0) = (320, 240)$, height of camera above the road $h = 1\mathrm{m}$, distance to object $z = 20\mathrm{m}$, time between consecutive frames $\Delta t = 40\mathrm{ms}$.

The detection limits of the linear motions are shown in figure 6. Each image shows the first view. Inside the black regions the error metric is lower than $T = 0.5\mathrm{px}$ (assuming a std. dev. in the correspondences of $\sigma = 0.167\mathrm{px}$). Parts of the object seen in these regions are not detected as moving. There is one important point in the image: the point of collision. This is the point where the

camera will collide with the object, provided that the object is slower than the camera. We will see that this dangerous point is not detectable in many cases.

The first row of figure 6 considers the epipolar constraint only. As can be seen parallel motion is not detected at all. Lateral motion is detected to a high extent. The black region is shaped like a bow tie.

In the second row the positive depth constraint is added. Overtaking objects are now detected. The error metric in this case is identical to the motion parallax induced by the plane at infinity. The optical flow of points at infinity is zero (camera does not rotate). Thus, the motion parallax is equal to the length of the measured optical flow. The contour lines (lines where the error metric takes on a constant value) are circular around the epipole. Preceding objects are still not detected. In the case of lateral motion the bow tie is cracked. The motion is also detected between the epipole and the point of collision due to the violation of the positive depth constraint.

The use of the positive height constraint (third row) gains the power of detection for the image part below the horizon. In the cases of parallel motion (overtaking and preceding objects) the error metric below the horizon is identical to the motion parallax induced by the road plane. It is possible to detect preceding objects but it is a challenging task. Lateral motion benefits from the positive height constraint only on the right-hand side of the epipole.

Adding the trifocal constraint yields the best achievable results. The parallel motion profits mainly from the larger driven distance of the camera, since the camera moves from c_1 to c_3 (not just to c_2). This just increases the signal to noise ratio. Similar results would be obtained if only the first and the third view would be evaluated. This does not hold for the lateral motion. The trifocal constraint allows a detection also to the left of the epipole.

The reason for that is given in figure 3. There the camera moves from c_1 to c_3 observing a point moving from X_1 to X_3. A situation is chosen such that the trajectories of the camera and the point are co-planar. They move within the epipolar plane. Considering the first two views the two-view constraints are fulfilled. The viewing rays meet perfectly in the point X_{t12}. This point lies in front of the cameras and above the road. Consequently, this kind of motion is not detected over two views alone. Considering the third view reveals the motion, since the triangulated point X_{t23} of the second and third view is different from X_{t12}.

We have seen that in case of the linear motion the strength of the trifocal constraint is not very high. The trifocal constraint shows its strength if the cameras translational direction changes over time as it is the case in the circular motion.

4.2 Circular Motion

The circular motion is modeled as shown in figure 5b. To demonstrate the detection limit for this case we consider an example similar to the "preceding object" example: $v_\mathrm{c} = 30\mathrm{km/h}$, $v_\mathrm{o} = 20\mathrm{km/h}$, $z = 20\mathrm{m}$, and $r = 100\mathrm{m}$.

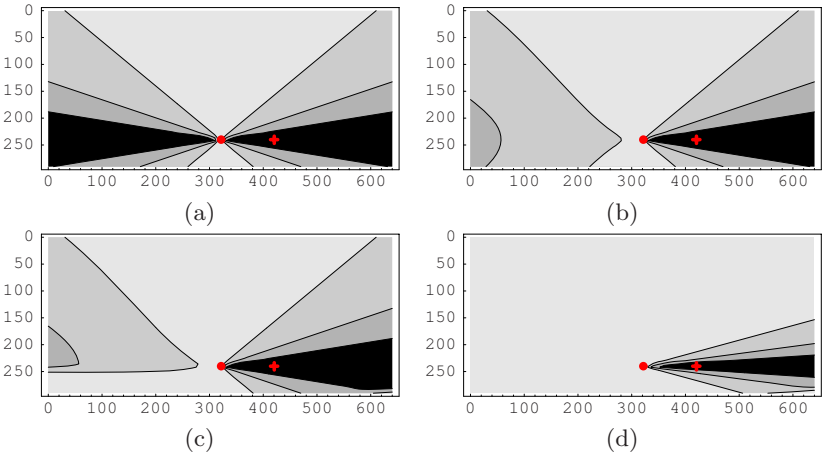


Fig. 7. Detection limit in the case of circular motion. The images show the first view (compare to fig. 5b). They are truncated at row 274, since below there is no object but the road. Inside the black regions the motion is not detected. The contour lines $2T$ and $4T$ are also shown. The red point marks the epipole, the red cross is the point of collision. (a) Epipolar constraint. (b) + positive depth constraint. (c) + positive height constraint. (d) + trifocal constraint.

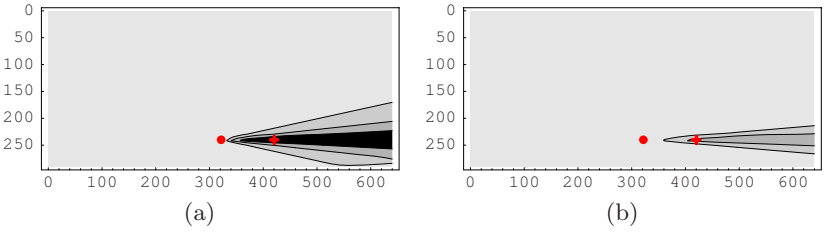


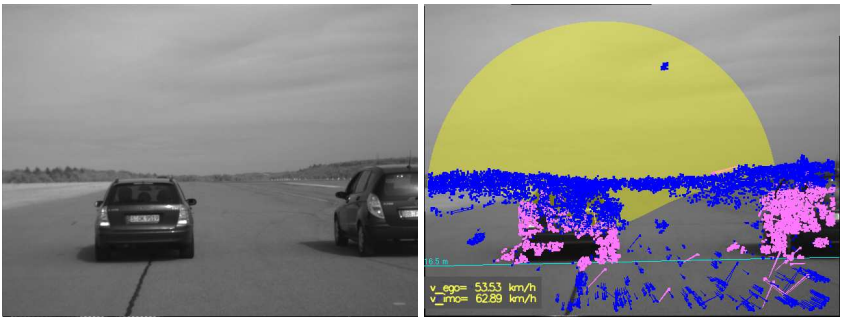
Fig. 8. Detection limit in the case of circular motion with tripled time period Δt compared to figure 7. (a) Epipolar + positive depth + positive height constraint. (b) + trifocal constraint.

Figure 7 shows the detection limit. Although the object is slower than the camera, which was a problem for the parallel motion case, the circular motion is detected to a high extent (fig. 7a). With the positive depth constraint taken into account the entire region to the left of the epipole is detected. It seems that the trifocal constraint (fig. 7d) just shrinks the black region, meaning that it only improves the signal to noise ratio. This is, however, not true. If we triple the time period $\Delta t = 120$ms the black region vanishes (figure 8b). Consequently, the entire object is detected as moving and so is the point of collision. The power of the two-view constraints is insufficient to detect that point.

Taking more than three views into account just increases the signal to noise ratio and hence shrinks the black regions but does not change the shapes of the contour lines (unless camera and object accelerate differently).

5 Experimental Results

In this section we apply the study on the detection limit to real imagery. Further, we detect the moving objects based on the measured optical flow and the proposed error metric d_2. The detection result is compared to the theoretical detection limit.



(a) (b)

Fig. 9. Experimental result. (a) Original image with two moving vehicles in front. (b) The semi-transparent yellow region shows the image region where the motion is not detectable. The measured optical flow vectors are classified as static (blue / dark) and moving (magenta / bright).

Figure 9a shows two vehicles driving in front of the camera (ego-vehicle). They are faster than the camera and move parallel to it. First, the detection limit is computed. To this end, the distance to the objects and the speed of them are required. The on-board radar sensor provides this information: $z = 16.5\text{m}$ and $v_{\text{oz}} = 62.9\text{km/h}$. The speed of the camera, retrieved by odometry, is $v_{\text{cz}} = 53.5\text{km/h}$. With this information together with the camera calibration the non-detectable region computes to that shown in figure 9b. Thereby the two-view constraints are considered.

The actual detection of the vehicles is carried out by the evaluation of the two-view error metric d_2 utilizing the measured optical flow. Radar data are ignored. The required ego-motion as well as the road homography are estimated using [6]. Flow vectors with $d_2 > T = 1.7\text{px}$ are classified as moving. The result is shown in figure 9b. One can see that the theoretical detection limit matches well to the practical one.

The vehicle on the right side is completely detected whereas only the lower part of the vehicle in the middle of the image is detected.

6 Conclusion

We have presented the detection limits of independently moving objects utilizing all available constraints existing for static 3D points. We have seen that:

- Objects which are faster than the camera are detected to a higher extent than those which are slower. That is a pity because slower objects are the dangerous ones. We will not collide with a faster object.
- In the event of linear motion the dangerous point of collision is not detected at all, what an irony of fate!
- The trifocal constraint emphasizes its potential if the motion of the camera is circular (non-linear). Then the point of collision is detectable (in principle).

References

1. Armangué, X., Araújo, H., Salvi, J.: Differential Epipolar Constraint in Mobile Robot Egomotion Estimation. In: IEEE International Conference on Pattern Recognition (ICPR), Québec, Canada, pp. 599–602 (2002)
2. Hartley, R., Zisserman, A.: Multiple View Geometry in computer vision, 2nd edn. Cambridge Press (2003)
3. Ke, Q., Kanade, T.: Transforming Camera Geometry to A Virtual Downward-Looking Camera: Robust Ego-Motion Estimation and Ground-Layer Detection. In: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), Madison, USA, pp. I-390- I-397 (2003)
4. Klappstein, J., Stein, F., Franke, U.: Flussbasierte Eigenbewegungsschätzung und Detektion von fremdbewegten Objekten. In: Workshop Fahrerassistenzsysteme (FAS), Löwenstein, Germany, pp. 78–88 (2006)
5. Klappstein, J., Stein, F., Franke, U.: Monocular Motion Detection Using Spatial Constraints in a Unified Manner. In: IEEE Intelligent Vehicles Symposium (IV), Tokyo, Japan, pp. 261–266 (2006)
6. Klappstein, J., Stein, F., Franke, U.: Applying Kalman Filtering to Road Homography Estimation. In: Workshop on Planning, Perception and Navigation for Intelligent Vehicles in conjunction with IEEE International Conference on Robotics and Automation (ICRA), Rome, Italy (2007)
7. Nistér, D.: An efficient solution to the five-point relative pose problem. In: IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), June 2004, pp. 756–770 (2004)
8. Torr, P.H.S., Zisserman, A., Murray, D.W.: Motion Clustering using the Trilinear Constraint over Three Views. In: Mohr, R., Wu, C. (eds.) Europe-China Workshop on Geometrical Modelling and Invariants for Computer Vision, pp. 118–125. Xidan University Press/Springer–Verlag (1995)
9. Trautwein, S., Mühlich, M., Feiden, D., Mester, R.: Estimating Consistent Motion From Three Views: An Alternative To Trifocal Analysis. In: International Conference on the Analysis of Images and Patterns (CAIP), Ljubljana, Slovenia, pp. 311–320 (1999)