Surrounding Moving Obstacle Detection for Autonomous Driving Using Stereo Vision

Regular Paper

Hao Sun1,*, Huanxin Zou1, Shilin Zhou1, Cheng Wang2 and Naser El-Sheimy3

1 School of Electronic Science and Engineering, National University of Defense Technology, Changsha, China
2 Department of Computer Science, School of Information Science and Technology, Xiamen University, Xiamen, China
3 Department of Geomatics Engineering, University of Calgary, Calgary, Alberta, Canada

* Corresponding author E-mail: chaosun@gmail.com

Received 27 Jun 2012; Accepted 29 Apr 2013

DOI: 10.5772/56603

© 2013 Sun et al.; licensee InTech. This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract Detection and tracking surrounding moving obstacles such as vehicles and pedestrians are crucial for the safety of mobile robotics and autonomous vehicles. This is especially the case in urban driving scenarios. This paper presents a novel framework for surrounding moving obstacles detection using binocular stereo vision. The contributions of our work are threefold. Firstly, a multiview feature matching scheme is presented for simultaneous stereo correspondence and motion correspondence searching. Secondly, the multiview geometry constraint derived from the relative camera positions in pairs of consecutive stereo views is exploited for surrounding moving obstacles detection. Thirdly, an adaptive particle filter is proposed for tracking of multiple moving obstacles in surrounding areas. Experimental results from real-world driving sequences demonstrate the effectiveness and robustness of the proposed framework.

Keywords Moving Obstacle Detection, Autonomous Driving, Multiple View Geometry, Particle Filter

1. Introduction

Intelligent and autonomous vehicles are a very popular research topic in computer vision, robotics and machine learning [1-2]. It covers a very large set of subjects such as real world localization, mapping, scene understanding, object detection and tracking, and obstacle avoidance. Each of those topics has been actively studied in the last decades and impressive results have been achieved. Mobile robots and autonomous vehicles generally need to carry out trajectory planning in response to stationary or moving obstacles. Moving obstacles carry potentially higher risks of collision than stationary obstacles in traffic. When a robot is placed in a dynamic environment, it cannot remain still, otherwise it might be hit by a moving obstacle. In addition, the future motion of the moving obstacles is usually not known in advance, and will have to be predicted. The complexity of the problem is augmented by a variety of factors including motion blur, varying lighting, large numbers of independently moving objects (sometimes covering almost the entire image), static and dynamic partial occlusions, and
suboptimal camera placement dictated by the constraints of a moving platform.

As was proved by the recent success of the DARPA Urban Challenge, autonomous navigation for vehicles is now possible even in difficult conditions. Unfortunately, those vehicles operating autonomously need combinations of multiple sensors such as laser scanners, radars and ultrasonic sensors to perform navigation. In recent years, the availability of high quality and inexpensive stereo cameras and the proliferation of high-powered computers have generated a great deal of interest in stereo vision based approaches.

Moving obstacle detection is conceptually simple when the camera is stationary, and a variety of solutions exist under the general heading of background subtraction. However, if the camera is moving, it becomes a difficult task, since the image motion is generated by the combined effects of camera motion, structure, and the motion of independently moving objects. Camera motion leads to a number of multiview constraints, which can be applied to motion detection. Commonly used constraints include the planar-parallax constraint [3], the epipolar constraint [4] and the trilinear constraint [3]. This paper aims to develop an automatic algorithm for detection and tracking of surrounding moving obstacles at short range from an autonomous driving vehicle using binocular stereo cameras. Most moving obstacle detection approaches based on stereo vision [5-7] use traditional stereo vision in dual-camera systems, in which two cameras have the same focal length and the optical axes are parallel and perpendicular to the baseline. For ideal dual stereo vision, the system satisfies non-merged geometry [8] and is convenient for stereo rectification and matching. However, this assumption does not hold well for many realistic systems. Here, we adopt a more general stereo model, in which uncalibrated cameras with non-parallel optical axes are used.

In this paper, we propose a novel framework for surrounding moving obstacle detection. Firstly, local invariant features are detected and a multiview feature matching scheme is then proposed for simultaneous stereo matching and motion tracking of the detected features. Secondly, a multiview epipolar constraint based on the relative camera positions in pairs of consecutive stereo views is derived for surrounding moving obstacles detection. Finally, an adaptive particle filter is introduced, which combines an adaptive prediction strategy and the multiview epipolar constraint for efficient moving obstacles detection and tracking.

The outline of the paper is as follows. In Section 2, we briefly give a mathematical formulation of the matching problem and present a novel multiview matching method for simultaneous establishment of both stereo and motion correspondences. Section 3 reviews the epipolar constraint and introduces the multiview epipolar constraint. A new particle filter based on adaptive prediction is described in Section 4, and in Section 5 the results of the experiments are reported. We conclude the paper in Section 6.

2. Simultaneous stereo and motion correspondence searching

SIFT features [9] are used for both stereo matching and motion tracking because of their promising performance [10]. Since our goal is to detect and track independent motion, we want to capture as many features on independently moving obstacles as possible. The performance of motion detection and tracking improves with increasing density as well as increasing reliability of feature correspondences [11]. However, the existing pairwise SIFT matching methods in the literature [10] all suffer from the problem that the reliability of correspondences decreases with increasing density of correspondences. More constraints are needed to improve the matching results.

We propose a novel matching method that explores the closed loop constraint. The closed loop constraint is that matching chains among multiple images should be closed loops. In detail, given N sets of features f_i(t = 1, 2, ..., N) from N separated views, the two-view matching problem can be defined as a search for a mapping [12]

\[ H_i : f_i \cup \{q\} \leftrightarrow f_i \cup \{\xi\} \]

that maps every feature in f_i to either a feature in f_i or \{\xi\} representing no match, and vice versa. \{\xi\} is defined as a void matching and represents the no match element. The multiview matching problem can be defined as a search for a mapping \( H : f_1 \leftrightarrow f_2 \leftrightarrow ... \leftrightarrow f_{N-1} \leftrightarrow f_N \) where \( f_i = f_i \cup \{\xi\} \).

If each feature-to-feature correspondence in a mapping \( \psi \) is part of a closed correspondence loop, we denote \( \psi \) as a closed loop mapping (CLM).

In real-world multiview images, it is difficult to estimate the ideal CLM for feature matching, owing to the existence of different kinds of geometric and photometric transformation in multiple images. We develop an iteration algorithm to find an asymptotically optimal CLM. Given images from N separated views, SIFT features are first detected in each image independently. Then each SIFT feature is matched to its k nearest neighbours (kNN) in related views. For every feature \( f_i^p \) (\( p = 1, 2, ..., n_i \)) in \( f_i \), its kNN \( \{f_{ij}^p, f_{ij}^2, ..., f_{ij}^k\} \) in \( f_i \) (we use \( k = 2 \)) are found. \( k \) possible mappings \( h_{ij}^{p,q} (\psi) \) (\( q = 1, 2, ..., k \)) can be obtained. The weight of each possible mapping is calculated as follows:

\[ W_{ij} = 1 - \frac{\text{dis}(f_i^p, f_{ij}^q)}{\text{dis}(f_i^p, f_{ij}^q)} \]  

(1)

www.intechopen.com
where \( \text{dis}(.,.) \) is the Euclidean distance between two feature vectors. If a closed mapping chain \( h(f_1, f_2, \ldots, f_n) \) for multiple images exists, the confidence measurement of the closed mapping chain will be given as:

\[
W = \sum_i W_i + W_{nj} |1 \leq i \leq N - 1, j = i + 1|
\]  

(2)

Let \( L \) be a set of closed mapping chains. The iteration algorithm when \( N = 3 \) is detailed in Algorithm 1. We have presented Algorithm 1 specifically for the case of \( N = 3 \) as it is the best option for the stereo vision. On the one hand, a large number of feature correspondences can be obtained. On the other, it benefits the real-time online processing for surrounding moving obstacle detection.

```plaintext
for every feature \( f_p^1(p = 1, 2, ..., n) \) in \( f_1 \\
Find its kNN \( f_{[1]} \) \( f_{[2]} \), \( \ldots \), \( f_{[k]} \) in \( f_2 \\
for every feature \( f_q^1(q = 1, 2, ..., k) \) in \( f_2 \\
Find kNN of \( f_q^1 \), \( f_{[1]} \), \( f_{[2]} \), \( \ldots \), \( f_{[k]} \) in \( f_1 \\
c_{21} = \text{true}, \text{if} \ f_q^1 \in \{f_{[1]} f_{[2]} \ldots f_{[k]}\} \text{in} f_1 \\
Find kNN of \( f_q^2 \), \( f_{[1]} \), \( f_{[2]} \), \( \ldots \), \( f_{[k]} \) in \( f_2 \\
c_{32} = \text{true}, \text{if} \ f_q^2 \in \{f_{[1]} f_{[2]} \ldots f_{[k]}\} \text{in} f_2 \\
for every feature \( f_t^3(r = 1, 2, ..., k) \) in \( f_3 \\
Find kNN of \( f_t^3 \), \( f_{[1]} \), \( f_{[2]} \), \( \ldots \), \( f_{[k]} \) in \( f_1 \\
c_{31} = \text{true}, \text{if} \ f_t^3 \in \{f_{[1]} f_{[2]} \ldots f_{[k]}\} \text{in} f_1 \\
if (c_{21} & \& c_{32}) \| (c_{31}) \\
Calculate weights of \( h_{12}(f_p^1, f_{[k]}^1), h_{23}(f_q^2, f_{[k]}^2) \) and
\( h_{31}(f_t^3, f_{[k]}^3) \\
Save L = L + [h(f_p^1, f_{[k]}^1, f_q^2, f_{[k]}^2, f_t^3, f_{[k]}^3), W] \\
endfor
endfor
endfor

Algorithm 1. The iterative algorithm for CLM estimation.
```

After each iteration step, only the closed mapping chain with the largest confidence measurement is selected as a potential element of \( H \). Uniqueness of the CLM is enforced by removing the existing element in \( H \) out of the iteration process. The iteration stops when the largest confidence measurement is below a certain threshold. The multiview feature matching scheme is performed after SIFT feature detection and its matching performance is independent of various weather conditions. It should be noted that the number of detected features may be affected by weather conditions.

3. Multiview Epipolar Constraint

Given two images of a static scene taken from two camera systems \( C_1 \) and \( C_2 \), let \( p_1 \) be the projection point of a 3D point \( P \) in the first image. Its corresponding projection point \( p_2 \) in the second image is constrained to lie on the epipolar line derived from \( p_1 \). This epipolar line is the intersection of two planes: one plane is defined by three points: the two optical centres and \( p_1 \), and the other plane is the image plane of the second image. A symmetrical relation applies to \( p_2 \). This is known as the epipolar constraint. The epipolar constraint of a static scene describes the camera motion between two optical centres. It is a commonly used geometric constraint for motion detection in two views. In the uncalibrated cases, it is represented by the fundamental matrix \( F_{12} \) that relates the 2D images of the same 3D point as follows:

\[
p_{2}^\text{T} F_{12} P_{1} = 0
\]  

(3)

Let \( l_2 = F_{12} P_{1} \) denote an epipolar line in View 2 induced by \( P_{1} \). If the 3D point \( P \) is static, then \( p_2 \) should ideally lie on \( l_2 \). Similarly, \( p_1 \) should ideally lie on the epipolar line \( l_1 = F_{21} P_{2} \), as shown in Figure 1(a). \( P \) is detected as moving if the projection points deviate from the epipolar lines.

![Figure 1. Multiple view geometry. (a) Epipolar geometry. (b) Multiview epipolar geometry.](image-url)

Let \( C_{s,t} (s = r, l; n = 1, 2, \ldots, t) \) denote the binocular stereo camera system (\( r \) denotes right view camera and \( l \) denotes left view camera) at time instant \( n \). Four groups of epipolar constraints \( \{F_{r,t+(t+1)}, F_{r,t+t}, F_{r,t+(t+1)}, F_{l,t+(t+1)}\} \) exist in pairs of consecutive stereo views, as shown in Figure 1(b). Either \( F_{r,t+t} \) or \( F_{l,t+t} \) alone is not sufficient to detect motion in adjacent frames. Since a large amount of camera motion exists between \( C_{s,t} \) and \( C_{l,t+t} \), \( C_{s,t} \) and \( C_{r,t+t} \) due to the combined effects of stereo configuration and camera movement, robust estimation of either \( F_{r,t+t} \) or \( F_{l,t+t} \) is possible [13]. Furthermore, when used together, theoretically all kinds of 3D motion can be detected by \( F_{r,t+t} \) and \( F_{l,t+t} \). \( \{F_{r,t+t}, F_{l,t+t}\} \) is essentially two pairs of epipolar constraints and is defined as the multiview epipolar constraint. The multiview epipolar constraint can be used for motion detection in stereo sequences. Similar to the case of epipolar constraint, a point-to-line distance is
defined to measure how much the projection points deviate from the corresponding epipolar lines:

\[ d = \alpha(l_{r,t} \cdot p_{t+1} + l_{l,t} \cdot p_{t+1}) + \beta(l_{r,t} \cdot p_{t+1} + l_{l,t} \cdot p_{t+1}) \]

(4)

where \( l_{r,t} \cdot p_{t+1} \) is the perpendicular distance from projection point \( p_{t+1} \) to the corresponding epipolar line \( l_{r,t} \), and \( \alpha, \beta \) are the normalized weighting factors. Various methods have been proposed to compute the fundamental matrix from correspondences. In our experiments, MAPSAC [14] (Maximum A Posterior Sample Consensus) based on correspondences from CLM is adopted for robust estimation of the multiview epipolar constraint.

4. Adaptive Particle Filter

Motion tracking from a mobile platform also faces some significant challenges [15]. First, the number of independently moving objects in the scene varies, as multiple moving objects might enter or leave the field of view. As a result, an external track initialization and termination mechanism is constantly needed. Second, the dynamics of independently moving objects are often nonlinear and non-Gaussian, so that no closed-form analytic solution can be obtained. Third, the tracking algorithm needs to filter erroneous motion detection in the presence of static or dynamic occlusions in the scene.

In recent years, particle filters [16] have become a popular approach to tracking in image sequences due to their ability to efficiently handle the uncertainties associated with the visual data and the target’s dynamics. Here, we give the basic concept of particle filters. Let \( x_{t-1} \) denote the state of the system at time instant \( t-1 \); let \( y_{t-1} \) be an observation at \( t-1 \); and let \( Y_{t-1} \) denote a set of all observations up to \( t-1 \). From a Bayesian point of view, all of the interesting information about the system state is encompassed by its posterior \( p(x_{t-1} | Y_{t-1}) \). This posterior is recursively estimated as the new observations \( y_t \) are made, a process that is realized in two steps: prediction (5) and update (6).

\[
p(x_{t-1} | y_{t-1}) = \int p(x_{t-1} | x_{t-1})p(x_{t-1} | y_{t-1})dx_{t-1} \tag{5}
\]

\[
p(x_{t} | y_{t}) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(\frac{-q^2}{2\sigma^2}\right) \tag{6}
\]

The recursion for the posterior requires the specification of a dynamical model describing the state transition \( p(x_t | x_{t-1}) \) and a model that evaluates the likelihood of any state given the observation \( p(y_t | x_t) \).

In our case, the system state \( x_t = [u_t, v_t, o_t]^T \) consists of the position \( (u_t, v_t) \) and the tracked label \( o_t \) of the point resulting from the projection of independent moving objects, where \( (u_t, v_t) \) are pixel coordinates measured in the right view image coordinate system. Since stereo correspondences are readily available from CLM matching, the position of the point in the left view coordinate system can be easily determined. For objects that move independently from the mobile platform, there is no way of predicting their motion in an explicit form. However, optical flow vectors computed from motion correspondences provide a close approximation of the transition probability of the projection point. The state-transition model is thus defined as:

\[
x_t = \begin{cases} 
0 & \text{if } o_t < 0 \\
x_{t-1} + s_{t-1} + w_t & \text{if } o_t > 0 \\
\Phi_b + w_t & \text{if } o_t = 0
\end{cases}
\]

(7)

where \( s_{t-1} = [du, dv, 0]^T \) denotes the motion vector of the projection point in the right view image; \( \Phi_b \) denotes the points that have no backward motion correspondences (projection points of moving objects entering the scene) and \( w_t \) is a discrete-time white noise sequence defined by a zero mean normal distribution. Note that the system state has time-varying dimensions as independently moving obstacles enter \( o_t > 0 \) or leave \( o_t < 0 \) the field of view. The likelihood function models the observation process.

\[
p(y_t | x_t) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(\frac{-q^2}{2\sigma^2}\right)
\]

(8)

where \( d \) is the normalized point-to-line distance in equation (4). The measurement is assumed to be corrupted by zero-mean normally distributed noise of variance \( \sigma^2 \). In order to get from a particle set \( E_t \) representation to moving obstacles, a two-dimensional clustering algorithm involving both probability clustering and spatial coherence clustering has been applied.

5. Experimental Results

Some examples of images from the VISAT™ data [17] for evaluation of the CLM matching method are shown in Figure 2. The performance of the proposed CLM matching method is compared to three state-of-the-art SIFT matching methods: the threshold-based nearest neighbour (TNN) matching method, the threshold-based mutual nearest neighbour (MTNN) matching method and the nearest neighbour distance ratio (NNDR) matching method. Multiview matches obtained by these matching methods are based on the transitivity of pairwise matches. We use an evaluation criterion similar to the one proposed in [10]. The evaluation criterion is based on the number of correct and false matches obtained for multiview images.
The results are presented with Recall versus 1 – Precision. Recall is the number of correctly matched features with respect to the number of corresponding features in multiple images of the same scene:

\[ \text{Recall} = \frac{\text{correct matches}}{\text{correspondences}} \quad (9) \]

The number of false matches relative to the total number of matches is represented by 1 – Precision:

\[ 1 – \text{Precision} = \frac{\text{false matches}}{\text{correct matches} + \text{false matches}} \quad (10) \]

Table 1 lists the experimental results on the VISAT™ multiview images. For a fixed number of 160 multiview matches (400 corresponding features), the results from different matching methods are presented with respect to recall, false positive rate, and the number of correct matches. CLM performs better than the other three matching methods. In our experiments, feature correspondences from CLM matching are used for robust estimation of the multiview epipolar constraint and for independent motion detection and tracking.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Recall</th>
<th>1-Precision</th>
<th>Correct matches</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLM</td>
<td>0.3275</td>
<td>0.1813</td>
<td>131</td>
</tr>
<tr>
<td>NNDR</td>
<td>0.2925</td>
<td>0.2688</td>
<td>117</td>
</tr>
<tr>
<td>MTNN</td>
<td>0.2150</td>
<td>0.4625</td>
<td>86</td>
</tr>
<tr>
<td>TNN</td>
<td>0.1450</td>
<td>0.6375</td>
<td>58</td>
</tr>
</tbody>
</table>

Table 1. Matching performance comparison

To verify the effectiveness of the proposed SIFT-based CLM methods in the various weather conditions, we have collected 500 groups of stereo image pairs, including 120 groups in sunny weather, 180 groups in cloudy weather, 140 groups in light rain weather and 60 groups in conditions of heavy rain. The statistics of matching results are shown in Table 2, where \( N_{ad} \) represents the average detected SIFT feature number per image, \( N_{am} \) represents the average matched SIFT feature number per image.

<table>
<thead>
<tr>
<th>Weather Condition</th>
<th>Stereo Pairs</th>
<th>( N_{ad} )</th>
<th>( N_{am} )</th>
<th>( N_{am} / N_{ad} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>sunny</td>
<td>120</td>
<td>569</td>
<td>355</td>
<td>0.62</td>
</tr>
<tr>
<td>cloudy</td>
<td>180</td>
<td>552</td>
<td>326</td>
<td>0.59</td>
</tr>
<tr>
<td>light rain</td>
<td>140</td>
<td>319</td>
<td>183</td>
<td>0.57</td>
</tr>
<tr>
<td>heavy rain</td>
<td>60</td>
<td>166</td>
<td>79</td>
<td>0.48</td>
</tr>
</tbody>
</table>

Table 2. Matching performance in various weather conditions.

It can be seen from Table 2 that the number of average detected features varies greatly in different weather conditions. The number drops dramatically in heavy rain, which implies that additional feature detectors are needed under these challenging circumstances. However, the matching ratio remains the same level and proves the robustness of our CLM method in different weather conditions.
Figure 3. Detection of moving obstacles by the multiview epipolar constraint. (a) (b) (c) (d) Stereo images from adjacent stereo views at time steps $t - 1$ and $t$. (e) Inliers of the multiview epipolar constraint. (f) Outliers of the multiview epipolar constraint. (g) (h) Feature points with large independent motion probability and corresponding epipolar lines depicted in the stereo images at time step $t - 1$. (i) Potential projection points of independently moving objects depicted in the right view image. (j) Detected independently moving objects marked with their bounding boxes.

Figure 3 shows the results of moving obstacles detection on images from pairs of consecutive stereo views using the multiview epipolar constraint. Feature correspondences used for fundamental matrix estimation are from CLM matching. Inliers of the multiview epipolar constraint with small point-to-line distances are depicted in Figure 3(e), while outliers with large point-to-line distances are depicted in Figure 3(f). The reference coordinate system is the right view coordinate system at time step $t$. Feature points that violate the multiview epipolar constraint are depicted in the stereo images at time step $t - 1$ with their corresponding epipolar lines (Figure 3(g) and Figure 3(h)). Only feature points with large independent motion probabilities (Figure 3(i)) are spatially clustered to obtain the final results. Both of the two independently moving vehicles have been detected (Figure 3(j)). Moving obstacles are indicated by their bounding boxes in our experiments.

In order to illustrate the performance of our method in various scenarios, two different kinds of stereo sequences, namely the on-road stereo sequence and the intersection stereo sequence, have been used in our experiments. Figure 4 shows some results from an on-road sequence, and Figure 5 shows the results from an intersection sequence.

Figure 4 displays eight selected frames (right view) from a real-world on-road stereo sequence. Moving obstacles detection results are depicted in the images with rectangles in different colours. We observe that multiple independently moving obstacles are simultaneously detected and tracked in most frames.

Figure 4. Real-world on-road stereo sequence: detected and tracked moving obstacles are marked with a rectangle.

Figure 5. Real-world road intersection stereo sequence: detected and tracked independently moving objects are marked with rectangles. Images in the second and third row are magnified sections of the original results in the first row.

Figure 5 displays experimental results from an intersection stereo sequence. The results are depicted in the right view image sequence. The results show magnified sections of the original data.

As is shown in the results, independently moving vehicles far from the stereo cameras are not detected in frame 6, frame 17, or frame 272, as few SIFT feature points can be extracted on these objects. Our algorithm
fails when the targets are far from the camera (about 200–500 m). In frame 571, frame 573 and frame 575, the car (the blue one) driving away in the same direction as the moving platform and the car (the black one) making the right turn are not detected because of dynamic occlusion. The tracking of the car depicted in the red rectangle in frames 571 and 573 is terminated in frame 575 because of occlusion by the white car (the one depicted in the yellow rectangle) and is initialized as a new track (depicted in the green rectangle) in frame 577. It can be seen that our algorithm is inefficient for handling dynamic occlusions.

To measure the performance of moving obstacles detection, we use the following parameters:

\[
\text{Rate} = \frac{N_c}{N_c + N_M + N_f},
\]
\[
\text{Accuracy} = \frac{N_c}{N_c + N_M + N_f}
\]

where \(N_c\), \(N_M\), and \(N_f\) represent the numbers of correctly detected, missed and false positive moving obstacles, respectively. We have tested the proposed algorithm on six on-road stereo sequences and ten road intersection sequences. For effective processing, we sample two random frames from every 30 frames within each second. Generally, there are no moving obstacles entering or leaving the field of view in such a short period. On the one hand, the random sampling strategy does not deteriorate the overall detection performance. On the other, it benefits the online processing for autonomous driving. The statistics of the experimental results are shown in Table 3.

<table>
<thead>
<tr>
<th>Sequence Category</th>
<th>Total Frames</th>
<th>(\frac{N_c}{N_M + N_f})</th>
<th>Rate</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>On-road sequence</td>
<td>350</td>
<td>829/67/48</td>
<td>0.9252</td>
<td>0.9453</td>
</tr>
<tr>
<td>Intersection sequence</td>
<td>120</td>
<td>406/55/42</td>
<td>0.8807</td>
<td>0.9063</td>
</tr>
</tbody>
</table>

Table 3. Summary results for surrounding moving obstacles detection.

6 Conclusion

In this paper, we present a novel algorithm for detection and tracking of surrounding moving obstacles for autonomous driving using stereo vision. The major contributions of this paper are: 1) A novel multiview feature-matching scheme is proposed for simultaneous stereo correspondence and motion correspondence establishment. 2) The multiview epipolar constraint, derived from the relative camera positions in pairs of consecutive stereo views, is applied to surrounding moving obstacles detection. 3) We propose an adaptive particle filter that combines the adaptive prediction strategy and the multiview epipolar constraint for the joint detection and tracking of multiple moving obstacles.

Multiple moving obstacles entering or leaving the field of view are handled effectively within the framework without relying on an external track initialization and termination algorithm. Experimental results on real-world driving sequences demonstrate the effectiveness and robustness of our method.

7. Acknowledgments

The authors thank the editor and our anonymous reviewers for valuable comments, which helped improve the clarity of the presentation of this paper. The work was supported by the National Natural Science Foundation of China (Project No. 40971245).

8. References