Robotics and Autonomous Systems 57 (2009) 268–278

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

journal homepage: www.elsevier.com/locate/robot

Cheap Joint Probabilistic Data Association filters in an Interacting Multiple Model design

Christian Hoffmann *, Thao Dang

Institute of Measurement and Control Engineering, Universität Karlsruhe (TH), 76131 Karlsruhe, Germany

ARTICLE INFO

Article history:
Available online 5 November 2008

Keywords:
Driver assistance
Vision sensor
Shadow
Symmetry
Sensor fusion

ABSTRACT

This paper presents an approach to fuse multiple sensors in an Interacting Multiple Model design. Visual features like shadow and symmetry, treated as independent stand-alone virtual sensors, are employed for detection and tracking of vehicles for driver assistance tasks. Cheap Joint Probabilistic Data Association is utilised to account for the large amount of clutter in the measurements provided by these sensors. Special attention is devoted to the different noise characteristics of the measurements. The individual sensors are considered in a sequential manner, leading to a versatile fusion architecture that allows easy integration of further sensor modules.

1. Introduction

As the computational power of modern microcomputers becomes available in automotive environments, high-level tasks like keeping the lane and safety margin to the leading vehicle are introduced in driver assistance systems. Even security-related topics such as emergency breaking and collision avoidance seem to be in reach. Crucial for all of these systems is a reliable perception of the surrounding environment, including, but not limited to, detection and tracking of leading and oncoming road users.

Currently, mainly radar sensors are used for this task. Although these sensors have many advantages, they suffer from a limited lateral field of view and lack the ability to classify detected objects. Thus, video sensors might usefully supplement or even replace radar sensors [1]. Apart from stereo and motion stereo approaches, evaluation of monocular features like vehicle shadows or symmetry has been used in autonomous applications [2–8]. A single 2D feature on its own, however, is too weak for reliable object detection. Thus, we combine the strengths of several features by fusing them to obtain robust and reliable detection results.

As object tracking has to cope with different traffic situations, an approach with multiple system models is implemented. The ability to switch between several motion models leads to reliable tracking even on lane change and breaking manoeuvres. An *Interacting Multiple Model (IMM)* algorithm ensures automatic soft model switching.

The different system models are represented by *Cheap Joint Probabilistic Data Association (cJPDA)* filters. This improves robustness in the presence of cluttered data, which is a common problem encountered when using the aforementioned visual features. We take into consideration that these measurements have differing covariances, which is discarded in many applications using *cJPDA* filters. However, we argue that this is important due to the fact that the measurements pass a nonlinear transformation from the image sensor into world coordinates. Thus, we provide a consistent derivation of an *Interacting Multiple Model Joint Probabilistic Data Association* filter for multiple sensors.

Combinations of *IMM* and probabilistic data association techniques showed good results for several problems [9,10], which encouraged the development of this vision based vehicle tracking system for driver assistance tasks. A comparative study of examples demonstrates the benefits for our application.

A brief overview of the overall system is given in Section 2, followed by an introduction of the visual features in Section 3. Section 4 describes the fusion architecture: Section 4.1 outlines validation of measurements prior to their association to specific tracks. The employed system models and the *Cheap Joint Probabilistic Data Association* filter are described in detail in Sections 4.2 and 4.3, respectively. Section 4.4 combines those in the *Interacting Multiple Model* approach, the multi-sensor extension is presented in Section 4.5. The assignment of the track control stage is explained in Section 4.6, followed by examples computed from real world image sequences in Section 5. Section 6 summarises our results and concludes the article.

2. Overview

Fig. 1 depicts the overall system. The image source, a progressive scan grey-level camera with 640×480 pixels and 8

* Corresponding author. Tel.: +49 6222 774686.
E-mail addresses: christian.hoffmann@heidelberg.com (C. Hoffmann), thao.dang@daimler.com (T. Dang).

0921-8890/$ – see front matter
doi:10.1016/j.robot.2008.10.009

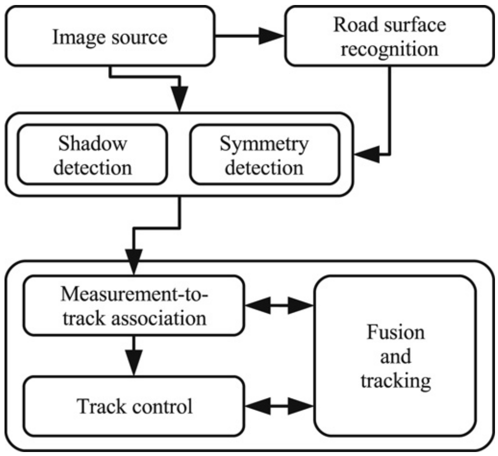

Fig. 1. System overview.

Fig. 2. Shadow and symmetry features.

bit depth, is mounted near the rear mirror behind the windshield of the test vehicle. Processing all 25 frames per second delivered by the camera will be called real-time in the sequel.

A road surface recognition module provides information about the position and orientation of the ground plane. This is necessary to determine the distance of an object detected by the feature extraction steps. As road surface recognition is not in the scope of this article, we refer to [11,12].

Every frame is scanned for shadow and symmetry cues independently [13]. In combination with road surface information, they may be considered as separate virtual sensors delivering 3D positions of object hypotheses.

These hypotheses are then fed into the fusion and tracking module. Besides pruning to ensure that only valid tracks survive, the track control takes care of spawning new tracks for measurements that cannot be assigned to an existing one.

For each track, a separate *Interacting Multiple Model* filter is instantiated. It combines two *Cheap Joint Probabilistic Data Association* filters with different system models. A model assuming constant velocity is sufficient for common car following. Situations demanding higher dynamics, like lane changes or breaking manoeuvres, are covered by a model assuming constant acceleration.

3. Visual features

Various features like shape [14], shadow [2,4], and symmetry [8,6,5,7] may be extracted from a single video image. Because of their good detection results for a wide range of situations at moderate computational cost, we chose to evaluate shadow and symmetry for vehicle detection. Fig. 2 depicts both features for a leading vehicle.

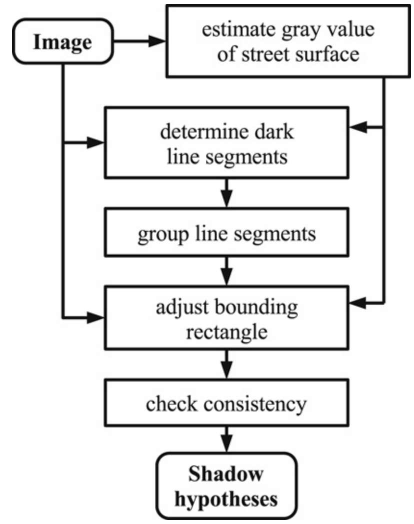

Fig. 3. Structure of the shadow detection process.

3.1. Shadow detection

Several processing steps are performed to obtain shadow hypotheses. Since shadow regions below vehicles exhibit sharp boundaries, we simultaneously evaluate region darkness and region gradient for shadow localisation. Fig. 3 shows the structure of the algorithm.

The primary characteristic of a shadow is its reduced brightness with respect to the road. For this reason, we propose to consider an estimate of the mean grey value g_S of the street surface within the shadow detection process.

For this purpose, a region from the bottom of the camera frame is selected that images the street surface straight in front of the observing vehicle. This assumption is validated by a test based on the variance of the grey values. If the variance is low, the hypothesis is accepted that there are no line markings or objects in the area in question. Otherwise, the region is shifted until the criterion is met. We describe the road intensity pattern by its mean g_S and variance σ_g^2.

The mean grey value g_S of the street surface is then used to determine initial shadow hypotheses, which are represented by rectangular boxes surrounding dark image regions. Besides their grey value, these hypotheses are required to match the expected width of a vehicle and are delimited by gradients at the bottom and at the left and right sides [13].

All boundaries of the initial shadow hypotheses are then refined individually. Refinement of the lower boundary is based on detection of the steepest gradient of the grey values in a vertical direction, i.e. the transition from shadow to street. Adjusting the left and right boundaries is analogously conducted.

Finding the upper boundary is the hardest part, because no previous knowledge about the region above the shadow exists, as no further assumptions about the observed vehicle are to be made. Thus, the algorithm tries to detect brighter spots that are often present due to reflexes on the rear push rod, or due to light shining through beneath the car. However, if no such bright spot can be detected, the hypothesis is left unchanged.

Finally, hypotheses have to pass a linear discriminant classificator rating the following attributes:

- markedness of the bottom gradient,
- markedness of the side gradients
- width/height ratio,
- mean grey value,
- grey value variance,