Image and Vision Computing 38 (2015) 13–23

Contents lists available at ScienceDirect

Image and Vision Computing

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

A stereo matching approach based on particle filters and scattered control landmarks☆

Stylianos Ploumpis, Angelos Amanatiadis *, Antonios Gasteratos

School of Engineering, Democritus University of Thrace, 12 Vas. Sofias Str, GR-67100 Xanthi, Greece

ARTICLE INFO

Article history:
Received 2 December 2013
Received in revised form 20 October 2014
Accepted 1 April 2015
Available online 18 April 2015

Keywords:
Stereo matching
Particle filters
Ground control points
Markov chains
Plane fitting

ABSTRACT

In robot localization, particle filtering can estimate the position of a robot in a known environment with the help of sensor data. In this paper, we present an approach based on particle filtering, for accurate stereo matching. The proposed method consists of three parts. First, we utilize multiple disparity maps in order to acquire a very distinctive set of features called landmarks, and then we use segmentation as a grouping technique. Secondly, we apply scan line particle filtering using the corresponding landmarks as a virtual sensor data to estimate the best disparity value. Lastly, we reduce the computational redundancy of particle filtering in our stereo correspondence with a Markov chain model, given the previous scan line values. More precisely, we assist particle filtering convergence by adding a proportional weight in the predicted disparity value estimated by Markov chains. In addition to this, we optimize our results by applying a plane fitting algorithm along with a histogram technique to refine any outliers. This work provides new insights into stereo matching methodologies by taking advantage of global geometrical and spatial information from distinctive landmarks. Experimental results show that our approach is capable of providing high-quality disparity maps comparable to other well-known contemporary techniques.

1. Introduction

While common digital images provide sufficient 2D information of a scene, several applications including robotics, entertainment and other require full 3D information, such as depth information, which can be accomplished by stereo imaging. Stereo matching is the field of study that aims to determine correspondences in two or more images shot from different viewpoints for obtaining a depth map of the actual scene. This major field of computer vision has been reviewed and categorized accordingly in [1,2] and [3]. Stereo matching techniques can be explicitly classified into two major groups, global and local approaches. The common approach in local algorithms is that the disparity value of a given pixel is calculated only by intensities populating a certain region around that pixel, which is called support window. Several strategies have been introduced to optimize the results of local techniques such as adaptive-weight cost aggregation strategies aiming to reduce implicit assumptions [4]. In global approaches, algorithms determine the disparity values simultaneously based on smoothness assumptions and tend to be iterative for refining their results with energy minimization techniques. The general framework is that local techniques are less computationally expensive than the global techniques and, therefore, they are mostly real time/hardware applicable. Despite this fact, global algorithms can achieve higher accuracy if they deploy cost aggregation techniques in their process. Several applications of belief propagation [5], graph cuts [6] and segmentation combined with plane fitting [7] have been integrated in global and local approaches to boost their performance.

In this paper, we present an approach which tackles the stereo matching problem from a novel perspective compared to known techniques. Our inspiration has been derived from the problem of robot localization in a known environment. This problem can be solved accurately by particle filters, where each particle is a possible state of the robot and each state has its own probability of the robot's actual positioning [8]. The probability is calculated by means of linked series of geometrical and spatial information derived from known landmarks in the environment. We address the stereo matching problem with the same technique, where at first we acquire a set of ground control points (GCPs) [9] by computing multiple disparity maps and subsequently we label them as landmarks. Those features possess an extremely precise disparity value and we utilize them for applying the particle filtering context in a scan line for accurate estimation of disparity values. For improving the effectiveness of particle filtering, we first classify each landmark using a segmentation based technique in each of the paired images. Additionally, we introduce a Markov chain model to increase the accuracy and decrease the computational expensiveness of particle filtering. Finally, we refine our results by applying the RANSAC algorithm [10] combined with a histogram technique

☆ This paper has been recommended for acceptance by Enrique Dunn.

* Corresponding author. Tel.: +30 2541079329.
E-mail address: aamanat@ee.duth.gr (A. Amanatiadis).

http://dx.doi.org/10.1016/j.imavis.2015.04.001
0262-8856/

which, by drawing out the outliers, allows us to obtain high quality disparity maps.

2. Related work

The proposed stereo matching approach is based on particle filtering. Although recent work has introduced particle filtering in 3D optical flow [11], our framework of stereo matching is completely different in every aspect. Moreover, it tends to relate to a small portion of global and local approaches as referred in the stereo related literature.

Global algorithms have lately increased their accuracy due to segmentation-based techniques as introduced by [12,13]. These approaches rely on the assumption of homogeneous color segments which approximate non-overlapping regions in the reference image. Over-segmentation is preferred, since it helps to meet these assumptions in practice, where in every segment the disparity values vary smoothly on a planar surface [7]. Moreover, global approaches have employed belief propagation in their frameworks. Belief propagation was initially established and described in [14] and it has been a popular method for state of the art global algorithms as an energy minimization technique, which is usually constructed by Markov random fields [7,5]. Furthermore, various strategies of graph cuts have been implemented in global approaches to address the same problem of energy minimization [6].

On the contrary, local algorithms examine each pixel independently with the assistance of cost aggregation strategies. Cost aggregation approaches compose a broad chapter in stereo literature which has been evaluated and analyzed extensively in [15]. Several methods of cost aggregation strategies have been introduced over the years based on shiftable windows [16,17], adaptive weights [4], multiple windows [18], and segment based windows [19]. Traditional local algorithms produce less accurate results compared to global ones, but lately this gap has been reduced. A popular method in this category is based on adaptive weights proposed by Yoon and Kweon [4], inspired by the Gestalt principles based on spatial proximity and color similarity. Additional improvements to this method have been proposed by deploying a segmentation based support window [19].

Lastly, there is a category of stereo algorithms which takes advantage of global smoothness assumptions combined with several local constrains. These algorithms fall under the category of semi-global approaches where both global and local techniques are applied for the same correspondence purpose. One of the first semi-global approaches was proposed by Hirschmuller [20], where a different approach of energy minimization was utilized. Semi-global techniques aim at minimizing the global 2D energy function by applying several 1D minimization methods. Generally, semi-global matching algorithms take advantage of scan line energy minimization combined with dynamic programming from several 1D directions. Due to the computational efficiency of semi-global approaches, a wide range of techniques have been introduced to address the stereo correspondence problem, such as discontinuity preserving interpolation in structured environments [21] and segmentation based techniques combined with plane fitting [22].

The overall framework of our approach has been also motivated by recent global approaches which utilize ground control points. These points are described as high confidence matches and their first appearance was in the work of Bobick and Intille [9]. There are two well known methods that can obtain such high confident matches. The first method requires strong feature correspondences, for obtaining high confidence starting points in order to initiate the GCPs calculation [23]. The second technique for acquiring the GCPs relies on the computation of multiple disparity maps based on local approaches and winner-take-all (WTA) strategies as it has been presented in [24]. A similar technique has been presented in [25] where the GCPs are obtained from local matching by oriented spatial filters.

3. Proposed approach

The proposed method has been motivated by the particle filter framework in robot localization. Early particle filter implementations in robot localization can be found in the literature, in which a robot's position has to be recovered from sensor data [8]. In all tracking problems it is essential to decompose the system into three basic models, the *inference* the *state* and the *dynamics* one. In the state model which describes the environment of a mobile robot, a state is usually measured by a two-dimensional Cartesian coordinate system and the orientation of heading. Additionally, in the inference model the sensor data provides the system with an estimation of its position in the environment combining spatial information from the surrounding objects. This spatial sensory information is often called a landmark. Finally, the dynamics model describes the evolution of the system states over time. When the system passes from a certain state $\mathbf{x}_t$ to the next $\mathbf{x}_{t+1}$ in discrete time t, the sensors gather new spatial information from the environment. Furthermore, the current state is related to the previous state $P(\mathbf{x}_{t+1}|\mathbf{x}_t)$ by a Markov chain model. Each state in the Markov chain model is not observable, instead the only observable variable is the set of measurements Z_t acquired from the sensors leading to a stochastic prominence of the true state $\mathbf{x}_t$.

More precisely, particle filters are approximate methods for the calculation of non-linear posterior probabilities in partially observable Markov chain models in discrete time, where an analytic solution to an integral equation is not feasible. In a typical non-linear Bayesian tracking model the list of total measurements up to t is denoted as Z_t and the feature measurements at time t is expressed as $\mathbf{z}_t$, while the set of states over time is expressed as X_t:

$$Z_t = \{\mathbf{z}_1, \ldots, \mathbf{z}_t\}, X_t = \{\mathbf{x}_1, \ldots, \mathbf{x}_t\}.$$

In order to calculate the posterior density probability $P(\mathbf{x}_t|Z_t)$, conditioned over all given observations until time t, we utilize Bayes' formula:

$$P(\mathbf{x}_t|Z_t) = \frac{p(\mathbf{z}_t|x_t, Z_{t-1})p(\mathbf{x}_t|Z_{t-1})}{p(\mathbf{z}_t|Z_{t-1})}. \quad (1)$$

Furthermore, using the Chapman–Kolmogorov equation for joint probability distributions, and assuming there is independence between observations, Eq. (1) can be rewritten as:

$$P(\mathrm{x}_t|Z_t) = \frac{p(\mathbf{z}_t|\mathbf{x}_t)\int_{x_t-1} p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|Z_{t-1})dx_{t-1}}{p(\mathbf{z}_t)}. \quad (2)$$

This solution of the non-linear Bayesian tracking problem is a conceptual solution and it cannot be determined analytically. Several approximations have been proposed over the years in certain sets of cases. In the simplified case of Kalman filters [26] where the dynamics of the system and the observations are not linear, the measurement noise derived from observations forms a distinctive Gaussian distribution. In the case of non-linearities a considerable amount of techniques have been introduced to determine an approximate solution including Monte Carlo approximations [27]. Monte Carlo methods are simulation-based techniques for computing posterior distributions. Particle filters are sequential Monte Carlo models where each particle is a possible state scattered in the known environment as a three dimensional variable containing the orientation and the Cartesian coordinates. In every particle i, there is a certain associated weight $w_t^{(i)}$, which is proportional to the prior probability of importance.