Image and Vision Computing 32 (2014) 391–404

Contents lists available at ScienceDirect

Image and Vision Computing

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

Editor's Choice Article

Unsupervised flow-based motion analysis for an autonomous moving system ☆

Andry Maykol Pinto *, Miguel V. Correia, A. Paulo Moreira, Paulo G. Costa

INESC Technology and Science, Faculty of Engineering of University of Porto, Rua Dr. Roberto Frias, s/n 4200-465 Porto, Portugal

ARTICLE INFO

Article history:
Received 28 June 2013
Received in revised form 14 February 2014
Accepted 2 April 2014
Available online 13 April 2014

Keywords:
Motion segmentation
Optical flow
Moving observer
Active surveillance
Mobile robot

ABSTRACT

This article discusses the motion analysis based on dense optical flow fields and for a new generation of robotic moving systems with real-time constraints. It focuses on a surveillance scenario where an especially designed autonomous mobile robot uses a monocular camera for perceiving motion in the environment.
The computational resources and the processing-time are two of the most critical aspects in robotics and therefore, two non-parametric techniques are proposed, namely, the Hybrid Hierarchical Optical Flow Segmentation and the Hybrid Density-Based Optical Flow Segmentation. Both methods are able to extract the moving objects by performing two consecutive operations: refining and collecting. During the refining phase, the flow field is decomposed in a set of clusters and based on descriptive motion properties. These properties are used in the collecting stage by a hierarchical or density-based scheme to merge the set of clusters that represent different motion models. In addition, a model selection method is introduced. This novel method analyzes the flow field and estimates the number of distinct moving objects using a Bayesian formulation.
The research evaluates the performance achieved by the methods in a realistic surveillance situation. The experiments conducted proved that the proposed methods extract reliable motion information in real-time and without using specialized computers. Moreover, the resulting segmentation is less computationally demanding compared to other recent methods and therefore, they are suitable for most of the robotic or surveillance applications.

1. Introduction

For the generalization of robotic applications it is crucial to overcome certain problems related to the perception and interpretation of the dynamic scene, for instance, the extraction of high level information in order to increase the robots' ability to perform suitable motion detection [4,11], tracking [17], object recognition [30,24] and navigation [6,7]. It is also imperative to increase their ability to interact with the environment and, thus, the robot must be able to detect and analyze their surrounding scene.

In the scientific community, motion perception is one of the most relevant areas under discussion, and there are several models to perform motion analysis in a variety of environments. However, most of the methods cannot achieve the real-time constraints imposed by mobile robots without specialized computers. In some cases, these computer devices cannot be used due to the small size of the vehicles or they cause a higher consumption of energy which reduces the autonomy of such robots. Nowadays, there are pixel-wise techniques that have good results [5]; however, the segmentation of motion commonly takes more than a pair of seconds. Techniques for robotic applications are computationally more efficient although, this improvement is usually done at the expense of using lower resolution of images and feature-based approaches. The computational resources and the processing-time are some of the most critical aspects for vision-based techniques applied to robotics. Usually, these applications tolerate some loss of accuracy in the algorithms to ensure a fast response [15].

The work presented by this research studies the real-time motion analysis using dense optical flow fields and for a practical use in a mobile robot. Motion segmentation is the process of dividing an image into different regions in a way that each region presents homogeneous motion characteristics. The goal is to segment different objects according to their motion coherence. In particular, the current research describes a real application that is installed in a corridor with large homogeneous regions that does not have a significant amount of structural clues for feature-based techniques. Optical flow techniques provide relevant motion information about the environment around the robot [15]. Moreover, dense optical flow fields provide a good representation of the visual and apparent motion for robotic applications [27, 25]; however, the analysis of these fields is a complex and challenging process that requires for sophisticated techniques.

☆ Editor's Choice Articles are invited and handled by a select rotating 12 member Editorial Board committee. This paper has been recommended for acceptance by Thomas Pock.

* Corresponding author at: Departamento de Electrotecnia da Faculdade de Engenharia da Universidade do Porto, Rua Dr. Roberto Frias, s/n 4200-465 Porto, Portugal. Tel.: +351 22 508 14 00; fax: +351 22 508 14 40.
E-mail address: andry.pinto@fe.up.pt (A.M. Pinto).

http://dx.doi.org/10.1016/j.imavis.2014.04.003
0262-8856/

In this article, the estimation of dense flow fields is conducted by the optical flow technique of [29] that was especially designed for small robotic applications equipped with generic computers. This technique identifies motion properties which are considered as high level information about the sequence and originates flow fields in a short period of time. The research proposes two unsupervised clustering techniques for segmenting dense flow fields: the Hybrid Hierarchical Optical Flow Segmentation (HHOFS) and the Hybrid Density-Based Optical Flow Segmentation (HDBOFS). These two techniques were designed for robotic applications with a vision system and limited computer resources. Two major and distinct phases form both methods, namely, *refining* and *collecting*. The *refining* stage decomposes the flow field in a set of distinctive clusters that represent image regions with different motion models and the *collecting* stage merges the set of clusters that were obtained in the previous phase (using a hierarchical scheme or a density-based scheme). This architecture reduces the computational requirements of the proposed methods (HHOFS and HDBOFS). An extensive and interesting comparison between parametric (K-means and EM) and the proposed non-parametric techniques (no assumptions about the distribution of the data) is presented.

In addition, this article proposes a model selection method, called Fusing Distributed Bayesian Hypothesis (FDBH), which combines a histogram-based approach with cost functions (that balance fitness and model complexity). The estimation of the number of clusters is incorporated in all methods.

Experimental considerations prove that modeling a clustering technique in a structure formed by two consecutive stages is computationally rewarding. The computational demands of the HHOFS and the HDBOFS are substantially lower than that of the EM and K-means that are conducted at flow level. The behavior of the proposed techniques can be adjusted to specific characteristics of the application. For instance, motion segmentation in surveillance operations can be appropriately performed without processing at pixel level. Therefore, the proposed techniques are completely capable of perceiving and understanding external motions in real-time and using low computational resources. However, the results have a blocky aspect which is usually tolerable by robotic moving systems.

Therefore, the contributions of this paper include:

1. Novel motion analysis methods with a reduced computational complexity: the HHOFS and the HDBOFS. The proposed architecture guides the motion analysis in both methods, enabling a reliable segmentation process while preserving the computational time requirements;
2. An efficient method to decompose the optical flow field into exclusive regions based on similarity properties of motion;
3. A model selection method to enhance the performance of the clustering techniques, the FDBH method. The Bayesian formulation combines a histogram-based analysis of the flow field in the polar space with the decay-ratio of a model selection criterion;
4. A comparative study of several unsupervised motion segmentation techniques (pixel-wise and block-wise) is provided;
5. An extensive qualitative and quantitative evaluation under realistic working conditions (with moving observers);

The article is organized as follows. Section 2 presents a brief review of motion segmentation methods that are commonly used in robotic systems to perceive motion. Section 3 shows the concept of the robotic application, named EEyeRobot. Section 4 presents the two unsupervised clustering techniques that are proposed and used in this research. Both non-parametric techniques are described with detail in Sections 4.1 and 4.2. Afterwards, a model selection method is proposed in Section 4.3. Experimental results are presented in Section 5. These results include the comparisons of the proposed techniques with the EM and the K-means. The experiments were conducted using the EEyeRobot in a real surveillance scenario. The results demonstrate that the HHOFS and the HDBOFS methods perform satisfactorily better, and can be used as a tool for motion analysis in applications with limited resources. Finally, Section 6 presents the most important conclusions of this research.

2. Related work

In the literature, it is possible to identify three motion perception methods for conventional fixed systems [19,33]: background subtraction, temporal differencing and optical flow. The most conventional techniques for motion perception consider that the visual changes are only caused by the movement of the external objects since they assume the stationary position of the observer. Therefore, they fail almost completely when the dynamic scene is captured by a non-static observer due to their inability to distinguish both motion components. For this reason, motion perception and analysis for moving observers is becoming an active research field and preliminary techniques typically use one of the following approaches: organizing the background into mosaics [6], modifying background subtraction methods [19,4] and optical flow or geometrical models [12,23].

Quian Yu and Gerard Medioni (2008) [36] focus on motion detection for a moving observer. They propose a mosaic approach that computes the homography between consecutive frames in order to compensate the egomotion. This assumes that the depth of the scene is much smaller than the distance between the object and the camera. To prevent registration errors from spreading, the authors adopt a sliding window and only a number of frames are considered. The movement of the sliding window demands for a high computational effort because all the registration processes must be executed. However, the algorithm was implemented in GPU and the time required to compute a 320×240 image is less than 100 ms.

Fernández-Caballero et al. (2010) [11] present a human detection method based on a thermal infrared camera mounted on an autonomous mobile robot. The detection is accomplished using a combination of optical flow and temporal differencing. The non-pyramidal Lucas–Kanade is used when the robot moves and the temporal differencing is used to detect human candidates based on thermal signatures when the robot stops. The authors focus on detecting motion interactions; however, the thermal cameras facilitate the detection of humans. A tracking application that resorts to pyramidal Lucas–Kanade optical flow to compute the flow field is presented in Jay et al. (2011) [8]. The main goal is to detect and extract regions where the flow does not represent the UAV's egomotion, for instance, for tracking a target that moves at different velocity compared to the background. The authors compute the pure movement of the target by searching along sparse diagonal lines which limits the Lucas–Kanade incoherence assumptions and reduces the computational complexity. An approach for motion clustering and classification based on consecutive images and a free-moving camera is presented in Jiman et al. (2010) [20]. The approach uses an optical flow technique and the random sample consensus (RANSAC) which removes outliers (scattered points) in the flow field. The flow field in Cartesian coordinates is transformed into polar axis (magnitude and orientation), and then, divided into blocks. The initial number and the respective cluster center are obtained by counting the selected block and by computing the density of the moving points. The clusters are redefined using the RANSAC, where each point is assigned to the initial cluster. Iteratively, the Euclidean distances to the clusters are computed and each point is updated with the cluster that has the minimum distance. The foreground and background are classified using the eigenvalue analysis based on the scatter of the cluster distributions, because they assume that the background is more scattered than the moving objects (due to their highest number of pixels).

Marco Tagliasacchi (2007) [34] presents a genetic-based optical flow estimation algorithm. The current frame is segmented using a watershed algorithm and by grouping the pixels with the same spatial position and similar color. This approach performs better at the border