Neurocomputing 100 (2013) 74–85

Contents lists available at SciVerse ScienceDirect

Neurocomputing

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

Pylon grid: A fast method for human head detection in range images

Rusi Antonov Filipov [a,*], Flávio Luis Cardeal Pádua [b], Marco Aurélio Buono Carone [b]

[a] Faculty of Computer Science, Hochschule Karlsruhe, Germany
[b] Department of Computing, Federal Center of Technological Education of Minas Gerais, Brazil

ARTICLE INFO

Available online 7 May 2012

Keywords:
Stereoscopy
Range image
Head detection
Local minimum search
Hill climbing
Perspective distortion
Orthographic projection
Structural convexization
Computer vision
People counting

ABSTRACT

We present a fast and accurate method for human head detection in range images captured by a stereo camera that is positioned vertically, pointing from the roof to the ground. We show how a static grid of measure points (pylons) can outperform hill climbing techniques and how a fast median filter can be used for effective preprocessing of the range data. The Pylon Grid algorithm detects all local minima in the range image and has a linear time complexity with respect to the number of pylons. One important prerequisite for applying the Pylon Grid algorithm to human head detection is a one-to-one relationship between human heads in the scene and local minima in the range image. This is achieved in a preprocessing phase, where an orthographic projection, convexization and noise filtering is applied to the range data. The preprocessing steps also run in linear time and can be parametrized for a further trade-off between computational cost and accuracy. The method was tested with crowded scenes, where multiple dense groups of up to six people move in random directions and have physical contact.

1. Introduction

There is a class of applications that make use of automatic tracking and counting of people in a scene. One example is consumer path tracking in shopping malls in order to collect statistical data about the consumers behavior. Another application people counting and pedestrian flow analysis in public areas.

An essential component of any tracking system is the *detector*, which determines the positions of the people in each frame of the image sequence. In the context of stereo vision, the problem of precise detection can be attacked effectively with a vertical camera orientation (Fig. 1).

While the vertical camera orientation limits the size of the observed area, it provides a clear view, with minimal occlusions among the people. This enables a reliable segmentation of the people from the ground, with accurate head detection, even in crowded scenes with hight people density.

Another advantage of the vertical camera orientation is that human heads appear as local minima in the range image (Fig. 2). Considering this, and assuming that human bodies have convex silhouettes in the range image, the problem of head detection can be reduced to finding all local minima in the range image.

2. Related work

Vision-based algorithms for the people counting and tracking problem may be divided into two main classes: the *direct* and the *indirect* methods [1]. The direct methods [2–5] are based on the detection of each person in the scene, by using some form of segmentation and object detection algorithm. The counting, in turn, is performed as a second step. The indirect methods [1,6–8] instead perform the people counting by using the measurement of some features that do not require the separate detection of each person in the environment.

This work presents a new algorithm, the Pylon Grid, which is used to detect efficiently human heads in range images. This algorithm lies at the heart of a novel direct vision-based approach to estimate people flow in both indoor and outdoor environments.

In [1], the authors focus on large groups of people which partially occlude each other. They propose a clustering algorithm that estimates the point density in each cluster and the distance to the camera, which faces the scene from an angle of about 45°. Khan and Shah [2] propose a multi-view approach to solving the problem of occlusion and lack of visibility in crowded and cluttered scenes. They gather evidence from every camera and use purely image-based 2D constructs. The authors of [3] attack the problem of occlusion with a Bayesian framework. In [4] the authors employ a learning approach that builds models of the people while the tracking. The authors of [5] focus on estimating the number of people in a big crowded scene.

In our method, we focus on small scenes instead and attack the occlusion problem with a vertical camera orientation.

* Corresponding author.
E-mail address: rfilipov@googlemail.com (R.A. Filipov).

0925-2312/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2011.12.040

Fig. 1. People flow at a shopping mall, a subway and an airport.

Fig. 2. A dense scene (center) captured from above and two corresponding range images. The student's heads appear at the darkest spots in the blobs. The right image was extracted using *sub-pixel interpolation* and has higher range resolution.

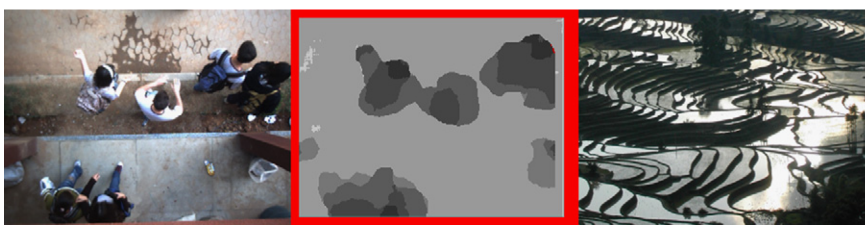

Fig. 3. The range image contains plateaus that resemble a rice field.

Our detector is designed to precisely detect each individual, because accurate detection can be a solid base for robust tracking and counting.

3. Similar schemes

3.1. Hill climbing

There is a class of *hill-climbing* algorithms that are able to find one local minimum in sublinear time. The approach is based on putting a climber somewhere in the range image and then check the surrounding neighbor pixels at a close distance. The climber descends by moving to the neighbor pixel with the smallest value. He continues the search from there, until he reaches a position where no neighbor has a smaller value, which is a local minimum.

This method is well suited for strictly monotonous functions, with local minima that spread over just a few positions. Unfortunately, one inherent property of range images is that they are *discrete functions*, where each pixel can only have a certain discrete value. This effect is visible in Fig. 3 (center), where the student's heads are covered by multiple pixels with the same range value. We can also observe that some areas of the upper body (shoulders and backpacks) have the same range value. Once a climber reaches such a *plateau*, it is hard to decide in which direction to continue the search.

The discrete values in the range image originate from the *disparity image* captured by the stereo camera, which is also stores only discrete values which represent the correlation of a pixel in the left image to the same pixel in the right image. Moreover, the range resolution of the range image decreases when the camera is placed higher above the ground (Fig. 4).

Fig. 5 shows some difficult situations in a hill climbing approach. In (a) it is difficult to decide where to start the search with the climber and in (d) it is hard to chose in which direction to continue, once a plateau is reached. In (c) the climber must decide where to continue his path, if two possible moves are equally adequate. Furthermore, a strict requirement of the problem presented in this text is to find the positions of *all* local minima in the range image, not just any local minimum. This is difficult to achieve with a single climber. And without knowing the number of human heads beforehand, it is hard to decide how many climbers to use (Fig. 5b).

Yet another goal is to find the *center* of the head plateau. In (e) the climber has reached the minimum, but now the problem is to find its center. Situation (f) shows two climbers that found the same minimum which requires special care in order not to count the minimum twice.

3.2. Watershed

The Watershed algorithm gradually "fills" layers of the range image with imaginary water, until all regions are filled enough to touch each other, so that borders between the filled regions emerge. As a result, the whole scene is divided into multiple adjacent cells. Hoang and Won use a marker-free Watershed algorithm for segmentation of protein spots [9]. Rambabu and Chakrabarti propose a combined Watershed and Hillclimbing technique targeting a hardware implementation [10]. While the Watershed method places a cell around every local minimum, the geometry of the cell does not accurately reflect the shape of the human silhouette, and the position of the local minimum within the cell is not obvious. This happens because the cell's shape depends on the relative positions of the adjacent local minima.