Robotics and Autonomous Systems 58 (2010) 457–464

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

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

Biologically and psychophysically inspired adaptive support weights algorithm for stereo correspondence

Lazaros Nalpantidis *, Antonios Gasteratos

Robotics and Automation Lab., Production and Management Engineering Department, Democritus University of Thrace, University Campus, Kimmeria, GR-671 00 Xanthi, Greece

ARTICLE INFO

Article history:
Received 22 October 2008
Received in revised form
5 February 2010
Accepted 11 February 2010
Available online 17 February 2010

Keywords:
Stereo vision
Disparity map
Gestalt laws
Biologically inspired

ABSTRACT

In this paper a novel stereo correspondence algorithm is presented. It incorporates many biologically and psychologically inspired features to an adaptive weighted sum of absolute differences (SAD) framework in order to determine the correct depth of a scene. In addition to ideas already exploited, such as the color information utilization, gestalt laws of proximity and similarity, new ones have been adopted. The presented algorithm introduces the use of circular support regions, the gestalt law of continuity as well as the psychophysically-based logarithmic response law. All the aforementioned perceptual tools act complementarily inside a straightforward computational algorithm applicable to robotic applications. The results of the algorithm have been evaluated and compared to those of similar algorithms.

1. Introduction

Stereoscopic vision is a reliable tool in order to obtain image and depth data for a scene at the same time [1]. The accuracy of the results depend on the choice of the stereo camera system and the stereo correspondence algorithm utilized. Stereo correspondence has been a focus of the machine vision community for a few decades [2]. It is implied by the biological finding that two, slightly moved from each other, images of the same scene are enough to perceive the depth of the objects depicted. Thus, the importance of stereo correspondence is obvious in the fields of machine vision, virtual reality, autonomous robot navigation, Simultaneous Localization and Mapping (SLAM), depth measurements and 3D environment reconstruction, as well as in many other aspects of production, security, defense, exploration and entertainment [3–8].

The main function of a stereo correspondence algorithm is to match each (in the case of dense stereo) or some (in the case of sparse stereo) pixels of the first image to their corresponding ones in the second image [9]. The outcome of this process is a depth image, i.e. a disparity map [10]. If accurately rectified stereo image pairs are provided, any two matching points reside on the same epipolar line. As a consequence, the matching can be simplified to an 1D search. The difference of the two pixels' horizontal coordinates of these points is the disparity. The disparity map is the spatial distribution of all the calculated disparities. The depth of the scene is thus perceived, since the more distant an object is, the larger its disparity value.

The advances in the field of stereo vision during recent years are, to a large extent, dominated and guided by the test-bench of Scharstein and Szeliski [11]. This test-bench pushes researchers towards better quality, dense disparity maps. Very accurate results have been reported on the corresponding web site [12]. However, the high quality of these results comes mostly at the expense of computation power and, thus, processing time. The stereo correspondence methods utilized in order to obtain such accuracy and coverage typically involve some kind of iterative global disparity consideration. Consequently, they are referred to as global algorithms [13–16]. However their iterative, time consuming nature prevents them from being used in real-time robotics.

On the other hand, there are the local methods i.e. methods where each disparity computation depends on the data from a relatively small surrounding region [17,18]. They are the preferred solution for robotic applications as they are presumed to be less computationally intensive. However their accuracy is low. Indeed, methods that use fixed support regions, or even adaptively variable in size and/or shape support regions, for aggregation of the computed absolute differences (AD) have been proven to produce results of inferior quality. Not until recently have the results of local algorithms become comparable to those of the global ones. The adaptive support weights (ASW) based methods [19,20] achieved this, by using fix-sized support windows, whose pixel contributions in the aggregation stage vary depending on their degree of correlation to the windows' central pixel. Despite the

* Corresponding author. Tel.: +30 2541079359.
E-mail addresses: lanalpa@pme.duth.gr (L. Nalpantidis), agaster@pme.duth.gr (A. Gasteratos).

0921-8890/$ – see front matter
doi:10.1016/j.robot.2010.02.002

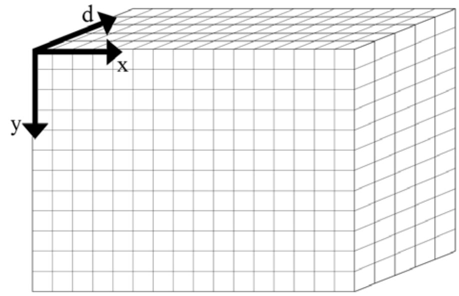

Fig. 1. Disparity space image volume.

acceptance that these methods have enjoyed, the determination of a correlation function is currently an active topic.

This paper proposes the incorporation of biological and psychological findings in the expression of a proper correlation function. The real world is the ultimate resource for finding the correct solutions in many fields of robotics, computer science and electronics [21–23]. The natural selection process is a strict judge that favors the more effective solutions for each problem. Of course, our understanding for the solutions that emerged from natural selection comes mainly from the sciences of biology and psychology. Applying ideas borrowed from these sciences in technological problems can lead to very effective results. Consequently, further blending of biological and psychological findings with the ASW method indicates a promising direction towards a simple, non-repetitive and accurate stereo correspondence algorithm.

2. Biologically inspired aggregation

The success of the human visual system (HVS) in obtaining depth information from two 2D images still remains a goal to be accomplished by machine vision. Incorporating procedures and features from the HVS into artificial stereo-equipped systems could improve their performance. The key concept behind this transfer of know-how from nature to science is identifying, understanding and expressing the basic principles of natural stereoscopic vision, aiming to improve the state-of-the-art in machine vision. These principles are mainly involved in the aggregation step that most existing algorithms employ. The AD computed for every pixel and for every candidate disparity value comprise a 3D volume called the disparity space image (DSI), as depicted in Fig. 1. Each value inside the DSI is updated by aggregating the values of constant disparity inside a support region centered at the value to be updated.

The HVS has been studied by many branches of the scientific community. Physics has expressed color information through color spaces, while biology has investigated the response of the eyes and the physiology of the eye. Psychophysics has studied the relationship between individual stimuli changes and the perceived intensity, which is applicable to vision as well as all the other modalities. On the other hand, the gestalt school of psychology suggested grouping as the key for interpreting human vision. The proposed algorithm embraces the findings of the scientific fields mentioned above and proposes a context within which all of them can coexist complementarily. Understanding each of the main features that the proposed algorithm incorporates is of great importance in order to clarify its structure.

2.1. Color usage

According to [24], using color information instead of gray values during stereo matching significantly improves the accuracy. Recently, the use of the CIELab color space has been proved to yield impressive results [19]. However, vision sensors, due to their structure, produce color images in the RGB color space. This fact is generally in accordance with the way the HVS perceives colors, as is shown in Fig. 2.

The conversion from the RGB to CIELab or similar color spaces demands non linear transformations and, as a result, it is computationally demanding. The use of the RGB color space's chromatic components is the simpler solution. Thus, the absolute differences for each channel of the RGB color space are taken into consideration. However, there are at least two possible methodologies for combining the three color channels. The International Telecommunications Union (ITU) in Recommendation BT.601-6 suggests that luminance (or intensity) information, represented as Y in color spaces such as YCbCr, YUV and XYZ, can be calculated as a weighted linear combination of the available RGB components. The weights for the red, green and blue chromatic channels are 0.299, 0.587 and 0.114 respectively. These values reflect photometric considerations and were derived from measurements of the response of the HVS to color stimuli. This equation is used in grayscale conversion by NTSC and JPEG. According to this, the aforementioned linear combination of the absolute differences (AD) calculated for each RGB channel becomes:

$$\mathrm{AD} = 0.299\mathrm{AD}_R + 0.587\mathrm{AD}_G + 0.114\mathrm{AD}_B \tag{1}$$

where AD denotes the total absolute luminance (or intensity) difference of two pixels in two images and AD_R, AD_G, AD_B denote, respectively, the absolute differences calculated for the red, green, blue chromatic channel only. In spite of being representative of the way the HVS accounts for each chromatic channel, this methodology is not the most credible one. The methodology most preferred in the literature [11,25] indicates that the same weight should be assigned to each one of the three chromatic channels, since each one contains the same amount of information. Thus, the total AD is a simple summation of the AD for each specific channel:

$$\mathrm{AD} = \mathrm{AD}_R + \mathrm{AD}_G + \mathrm{AD}_B. \tag{2}$$

This simpler treatment presents a better performance than the more sophisticated one, as indicated by the preliminary tests conducted.

2.2. Circular windows

The search for pixel correspondences between the two images of a stereo image pair is usually treated by comparing the surrounding regions of the examined pixels, rather than the examined pixels alone. The choice of those pixel regions, commonly referred to as support windows, plays an important role in the accuracy of the results. The support windows may vary in shape or dimensions and could be either of a fixed size or of an adaptive one. However, square or rectangular regions are the most common choices concerning the shape. The choice of the support windows' dimensions aims at establishing a compromise between noise compensation, which is favored by large dimensions, and detail preservation, which is favored by small dimensions. Support windows of fixed shape and size are the simplest solutions.

ASW as presented in [19] make use of fixed size, square windows with comparatively large size. However, the biological model of stereo vision seems to be better approximated by using circular shaped windows [26]. Aggregation inside circular windows is also preferable since the contribution of the neighboring pixels becomes perfectly isotropic, i.e. there are the same number of pixels contributing in any direction on the image plane. This fact makes the aggregation results produced by circular windows more reliable than any other window shape.