Expert Systems with Applications 41 (2014) 2043–2051

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Automatic expert system for 3D terrain reconstruction based on stereo vision and histogram matching

R. Correal [a,*], G. Pajares [a], J.J. Ruz [b]

[a] University Complutense of Madrid, Software Engineering and Artificial Intelligence Dept., C/ Profesor, José García Santesmases, s/n, 28040 Madrid, Spain
[b] University Complutense of Madrid, Computers Architecture and Automation Dept., C/ Profesor, José García Santesmases, s/n, 28040 Madrid, Spain

ARTICLE INFO

Keywords:
Expert system
Terrain reconstruction
Histogram matching
Stereo vision
Image processing

ABSTRACT

This paper proposes an automatic expert system for 3D terrain reconstruction and automatic intensity correction in stereo pairs of images based on histogram matching. Different applications in robotics, particularly those based on autonomous navigation in rough and natural environments, require a high-quality reconstruction of the surface. The stereo vision system is designed with a defined geometry and installed onboard a mobile robot, together with other sensors such as an Inertial Measurement Unit (IMU), necessary for sensor fusion. It is generally assumed the intensities of corresponding points in two images of a stereo pair are equal. However, this assumption is often false, even though they are acquired from a vision system composed of two identical cameras. We have also found this issue in our dataset. Because of the above undesired effects the stereo matching process is significantly affected, as many correspondence algorithms are very sensitive to these deviations in the brightness pattern, resulting in an inaccurate terrain reconstruction. The proposed expert system exploits the human knowledge which is mapped into three modules based on image processing techniques. The first one is intended for correcting intensities of the stereo pair coordinately, adjusting one as a function of the other. The second one is based in computing disparity, obtaining a set of correspondences. The last one computes a reconstruction of the terrain by reprojecting the computed points to 2D and applying a series of geometrical transformations. The performance of this method is verified favorably.

1. Introduction

1.1. Problem statement

Machine vision is an excellent sensor, widely used for a multitude of different applications. In the domain of robotic autonomous navigation in natural, rough terrain and planetary rovers for space exploration, one of the most important features is 3D perception and terrain reconstruction for path planning and navigation. Stereoscopic vision is a mechanism to obtain depth or range data based on images. It consists of two cameras separated by a given distance to obtain two differing views of a scene, similar to human binocular vision. By comparing both images, relative depth information is obtained, in the form of disparities, which are inversely proportional to distance to objects. A matching process computes the difference in position of a set of features or pixels from one image relative to the other. Provided that the position of centers of projection, the focal length and the orientation of the optical axes are known, the depth can be established by triangulation of the disparities obtained from the matching process. It is then reprojected to 3D space using perspective transformations, obtaining a set of world coordinates.

Different methods and strategies for 3D environment reconstruction using stereo vision have been applied in different works (Bakambu, Allard, & Dupuis, 2006; Goldberg, Maimone, & Matthies, 2002; Lin & Zhou, 2009; Morisset, 2009; Song & et al., 2012; Xing-zhe, 2010). While most existing strategies focus in the problem of computation of disparities and the matching process, there is little work devoted to the correction and validation of the input images, beyond vertical alignment and rectification (Papadimitriou & Dennis, 1996; Kang & Ho, 2012). The constant image brightness (CIB) approach assumes that the intensities of corresponding points in two images of a stereo pair are equal. This assumption is central to much of computer vision works. However, surprisingly little work has been performed to support this assumption, despite the fact the many of the algorithms are very sensitive to deviations from CIB. In Cox and Hingorani (1995) a study revealed that after an examination of 49 images pairs contained in the SRI JISCT stereo database (Bolles, Baker, & Hannah, 1993), a dataset that includes images provided by research groups at JPL, INRIA, SRI, CMU, and Teleos, the constant image brightness assumption is indeed often false. We have also found the same issue in our set of stereo images affecting the matching process necessary for terrain reconstruction.

* Corresponding author. Tel.: +34 685850240.
E-mail address: raulcorreal@hotmail.com (R. Correal).

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.09.003

Fig. 1. Stereo pair of a natural terrain.

The problem of correspondence in real (non-simulated) stereoscopic systems stems from the fact that images from cameras, although similar, show different intensity levels for the same physical entity in the 3D scene. An important reason for this feature lies in the different response from the camera sensors to the signal light from the scene and also from the different mapping of the scene over each image due to their different locations. That makes necessary to devote a major research effort to correct these deviations typical of all natural and real stereo system, as this problem is not yet satisfactorily solved, particularly in unstructured and uncontrolled environments. Fig. 1 shows an example of a stereo pair taken with the onboard navigation camera, a Videre stereoscopic system of parallel optical axes separated 9 cm, STH-DCSG-9 color, 640×480 pixels resolution and 3 mm miniature lenses. It can be appreciated both images expose slightly different color tones, although they have been simultaneously captured by two supposedly identical cameras (in next sections the histograms of these images is shown, where differences can be more easily observed).

Most matching algorithms are based on the minimization of differences, using correlation between brightness (intensity) patterns in the local neighborhood of a pixel in one image with respect the other (Baker, 1982). These differences in illumination may lead to a poor and inaccurate computation of disparities, and therefore to an incorrect terrain reconstruction.

We propose a new automatic method based on several sequential stages for 3D terrain reconstruction from stereoscopic data applied to robotic navigation in natural terrain, where the initial phase, correction of images, is based on the application of the human expert knowledge. This leads to the design of the proposed expert system, gaining an important advantage with regard to other approaches, as no training is required and it can be directly applied to the stereo pair under processing becoming independent from other images. The design of this automatic expert system makes the main contribution of this paper.

1.2. Revision of methods

Several strategies have been proposed to correct intensity values (brightness) in images. Next, some of the most commonly used techniques, which have been in addition employed in stereo vision applications, are commented:

(1) Homomorphic filtering (Correal, Pajares, & Ruz, 2013; Gonzalez & Woods, 2008; Pajares & de la Cruz, 2007) is a procedure based on the fact that each image is formed by the concurrence of two-component image: reflectance (r) and illumination (i). The illumination component comes from the light conditions in the scene when the image is captured and may change as the light conditions also change. The reflectance component depends on how the objects in the scene reflect light, which is determined by the intrinsic properties of the objects themselves, which (usually) do not change. In many practical applications it is useful to enhance the reflectance component, while the illumination is reduced. Homomorphic filtering is a filtering process in the frequency domain to compress the brightness based on the lighting conditions, while enhancing the contrast from the reflectance properties of the objects. This approach is based on the fact that an image $f(x, y)$ can be expressed in terms of illumination $i(x, y)$ and reflectance $r(x, y)$ components by the relationship, $f(x, y) = i(x, y)r(x, y)$. Illumination is associated with low frequencies of the Fourier transform and reflectance with high frequencies. The above equation cannot be used directly to operate separately in the illumination and reflectance components because the Fourier transform of the product of two functions is not separable; however, if we define $z(x, y) = \ln f(x, y) = \ln i(x, y) + \ln r(x, y)$, then $\zeta\{z(x, y)\} = \zeta\{\ln f(x, y)\} = \zeta\{\ln i(x, y)\} + \zeta\{\ln r(x, y)\}$, or $Z(u, v) = I(u, v) + R(u, v)$ where $I(u, v)$ and $R(u, v)$ are the Fourier transforms of $\ln i(x, y)$ and $\ln r(x, y)$ respectively. If we process $Z(u, v)$ by a filter function $H(u, v)$ then we obtain, $S(u, v) = H(u, v)\, Z(u, v) = H(u, v)\, I(u, v) + H(u, v)\, R(u, v)$, where $S(u, v)$ is the Fourier transform of the result. In the spatial domain, $S(x, y) = \zeta^{-1}\{S(u, v)\} = \zeta^{-1}\{H(u, v)\, I(u, v)\} + \zeta^{-1}\{H(u, v)\, R(u, v)\} = i'(x, y) + r'(x, y)$. Finally, as $z(x, y)$ is obtained by taking the logarithm of the original image $f(x, y)$, the reverse operation provides the desired enhanced image $g(x, y)$, namely $g(x, y) = \exp[s(x, y)] = \exp[i'(x, y)]\, \exp[r'(x, y)] = \mathrm{i}_0(x, y)\, \mathrm{r}_0(x, y)$, where $i_0(x, y)$ and $r_0(x, y)$ are the illumination and reflectance components of the output image. Further details of the application of this process to stereo images and results are described in Correal et al. (2013).

(2) Histogram equalization is a method used in image processing for contrast adjustment using the image's histogram (Laia, Chunga, Chena, Lina, & Wangb, 2012; Pajares & de la Cruz, 2007). This method usually increases the global contrast, especially when the usable data of the image is represented by close contrast values. Through this adjustment, the intensities can be better distributed on the histogram. This allows for areas of lower local contrast to gain a higher contrast. Histogram equalization accomplishes this by effectively spreading out the most frequent intensity values. Consider a discrete grayscale image $\{x\}$ and let n_i be the number of occurrences of gray level i. The probability of an occurrence of a pixel of level i in the image is $p_x(i) = \mathrm{p}(x = i) = n_i/n$, $0 \leqslant i < L$, L being the total number of gray levels in the image, n being the total number of pixels in the image, and $p_x(i)$ being in fact the image's histogram for pixel value i, normalized to [0, 1]. The cumulative distribution function corresponding to p_x is $\mathrm{cdf}_x(i) = \Sigma\ p_x(j)$, which is also the image's accumulated normalized histogram. Then, a transformation of the form $y = T(x)$ produces a new image $\{y\}$, such that its CDF will be linearized across the value range, i.e. $\mathrm{cdf}_y(i) = iK$ for some constant K. The properties of the CDF allows to perform such a transform; it is defined as $y = T(x) = \mathrm{cdf}_x(x)$. The T function maps the levels into the range [0, 1]. In order to map the values back into their original range, the following transformation needs to be applied on the result: $y' = y\, (\max\{x\} - \min\{x\}) + \min\{x\}$. This method can also be used on color images by applying the same method separately to the Red, Green and Blue components of the RGB color values of the image. Different histogram equalization methods have been applied in stereo vision (Cox & Hingorani, 1995; Liling, Yuhui, Quansen, & Deshen, 2012; Nalpantidis & Gasteratos, 2010; Zhang, Lafruit, Lauwereins, & Van Gool, 2010).

(3) In Kawai and Tomita (1998) authors propose a method to calibrate intensity for images based on segment correspondence. First, the edges are detected in each image. Each