Monocular Depth Cues in Computer Vision Applications

Diego Cheda

Everis, Barcelona, Spain
Advisor/s: Dr. Daniel Ponsa Mussarra and Dr. Antonio M. López Peña
Date and location of PhD thesis defense: 12 December 2012, Universitat Autònoma de Barcelona, Barcelona, Spain

---

1   Abstract

Depth perception is a key aspect of human vision. It is a routine and essential visual task that the human do effortlessly in many daily activities. This has often been associated with stereo vision, but humans have an amazing ability to perceive depth relations even from a single image by using several monocular cues.

In the computer vision field, if image depth information were available, many tasks could be posed from a different perspective for the sake of higher performance and robustness. Nevertheless, given a single image, this possibility is usually discarded, since obtaining depth information has frequently been performed by three-dimensional reconstruction techniques, requiring two or more images of the same scene taken from different viewpoints.

Recently, the problem of estimating depths from a single image has received a growing interest. Current approaches are mainly focused on the design of sophisticated features and levels of reasoning for accurate depth estimation. Their final goal is inferring the 3D structure of the scene as good as possible. However, with the aim of solving a given computer vision problem, the computation of 3D information could be just the starting step to subsequently analyze an image more reliably instead of being a goal in itself. Furthermore, even just a rough depth description could help to achieve better performances.

In our thesis, we first focus on how, by means of simple image features based on monocular cues, is possible obtaining a representation of a scene depth from a single image taken from a camera. This representation is computed by applying the learned relation between a set of visual features extracted from a scene image and scene depth. Figure 1 depicts an example of what kind of information we want to predict. With this objective in mind, we develop a low-cost method to estimate coarse depth information, but informative enough to tackle many computer vision problems from a different perspective, leading to alternative and more reliable solutions.

In particular, we explore the use of coarse depth information in three different computer vision problems: egomotion estimation, background estimation, and pedestrian candidate generation. We pose each one of these problems by proposing novel ways of addressing them.

In the first case, we have computed camera rotation mounted in a moving vehicle applying two novels methods based on distant elements in the image, where the translation component of the image flow vectors is negligible [1, 3]. In background estimation, we have proposed a novel method to reconstruct the background by penalizing close regions in a cost function, which integrates color, motion, and depth terms [2]. Finally, we have

---

■ near   ■ medium   ■ far   ■ very-far
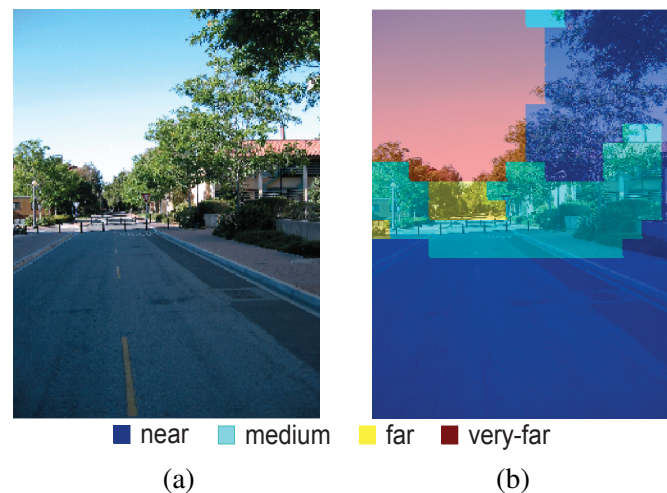
(a)   (b)

Figure 1: Example from our approach showing: (a) Original image and (b) Ideal depth segmentation considering four depth thresholds.

benefited of geometric and depth information available on single images for pedestrian candidate generation to significantly reduce the number of generated windows to be further processed by a pedestrian classifier [4]. In all cases, results have shown that our approaches contribute to better performances.

Our final intention is following one of the primary goals in computer vision, which is characterized by holistic scene understanding. In that sense, instead of explaining certain aspects of a scene in isolation, we focus on how to exploit the existent dependencies between depth and other problems to achieve better performance in different computer vision tasks.

References

[1] D. Cheda, D. Ponsa, and A. M. López. Camera Egomotion Estimation in the ADAS Context. In *IEEE Conf. Intell. Transp. Syst.*, pages 1415 –1420, 2010.

[2] D. Cheda, D. Ponsa, and A. M. López. Monocular Depth-Based Background Estimation. In *Int. Conf. Comput. Vision Theory Appl.*, pages 323–328, 2012.

[3] D. Cheda, D. Ponsa, and A. M. López. Monocular Egomotion Estimation Based On Image Matching. In *Int. Conf. Pattern Recognit. Appl. and Methods*, pages 425–430, 2012.

[4] D. Cheda, D. Ponsa, and A. M. López. Pedestrian Candidates Generation using Monocular Cues. In *IEEE Intell. Veh. Symp.*, pages 7–12, 2012.