

Chapter 28

Recovering 3D Human Body Postures from Depth Maps and Its Application in Human Activity Recognition

Nguyen Duc Thang
Kyung Hee University, Korea

Md. Zia Uddin
Kyung Hee University, Korea

Young-Koo Lee
Kyung Hee University, Korea

Sungyoung Lee
Kyung Hee University, Korea

Tae-Seong Kim
Kyung Hee University, Korea

ABSTRACT

We present an approach of how to recover 3D human body postures from depth maps captured by a stereo camera and an application of this approach to recognize human activities with the joint angles derived from the recovered body postures. With a pair of images captured with a stereo camera, first a depth map is computed to get the 3D information (i.e., 3D data) of a human subject. Separately the human body is modeled in 3D with a set of connected ellipsoids and their joints: the joint is parameterized with the kinematic angles. Then the 3D body model and 3D data are co-registered with our devised algorithm that works in two steps: the first step assigns the labels of body parts to each point of the 3D data; the second step computes the kinematic angles to fit the 3D human model to the labeled 3D data. The co-registration algorithm is iterated until it converges to a stable 3D body model that matches the 3D human posture reflected in the 3D data. We present our demonstrative results of recovering body postures in full 3D from continuous video frames of various activities with an error of about 6° - 14° in the estimated kinematic angles. Our technique requires neither markers attached to the human subject nor multiple cameras: it only requires a single stereo camera. As an application of our body posture recovery technique in 3D, we present how various human activities can be recognized with the body joint angles derived from the recovered body postures. The features of body joints angles are utilized over the conventional binary body silhouettes and Hidden Markov Models are utilized to model and recognize various human activities. Our experimental results show the presented techniques outperform the conventional human activity recognition techniques.

DOI: 10.4018/978-1-61350-326-3.ch028

INTRODUCTION

Through several million years of human evolution, stereopsis is one of the unique functions in the human vision system, allowing depth perception: it is a process of combining two images projected to two human eyes to create the visual perception of depth. Learned from the human stereoscopic system, a stereo camera was invented to synchronously capture two images of a scene with a slight difference in the view angle from which depth information of the scene can be derived. The depth information is generally reflected in a 2-D image called a depth map in which the depth information is encoded in a range of grayscale pixel values. Since its first commercial product in 1950s, *Stereo Realist*, introduced by the David White Company, there have been continuous developments of a stereo camera until now with the latest products such as a digital stereo camera, Fujifilm FinePix Real 3D W1 and a stereo webcam, Minoru 3D. Lately, 3D movies, in which depth information is added to RGB images, have received a lot of attention with the latest success of a film, *Avatar* released in 2009. Watching 3D movies and 3D TVs with the special viewing glasses is becoming a part of our lives these days.

Another area where the depth information could be valuable is the field of human computer interaction (HCI). In this area, 3D motion information of a user is utilized to better control external devices such as computers and games. In the conventional ways, capturing 3D human motion or movement (i.e., a sequence of human postures) is typically done using optical markers or motion sensors. Such systems are capable of producing some kinematic parameters of human motion with high accuracy and speed using wearable optical markers or sensors. However, it is inconvenient to a user who needs to wear specially designed optical markers or sensor-suits when running these systems. This disadvantage combined with the high cost equipment makes the systems impractical in daily use applications. In the case

of using motion sensors, a user has to hand-hold controllers equipped with accelerometers or gyroscopes. One good example is the Wii controller of Nintendo which uses optical sensors and accelerometers to recognize the hand motion of the user to control the games. Lately, some efforts are being made to capture the whole body movement without the markers or motion sensors. Using a stereo camera and its derived depth map is one of options, since depth maps may provide sufficient 3D information to derive human body motions in 3D. Although this approach should open a new possibility for various novel applications in HCI such as games and u-lifecare, obtaining human body postures in 3D directly from depth maps is not very straightforward.

There have been some attempts to develop marker-less systems to estimate human motion from a sequence of monocular images or RGB images, only reflecting 2-D information. Because the 3D information of the subject is lost, the efforts to reconstruct the 3D motion of the subject from only monocular images face difficulties with ambiguity and occlusion that lead to inaccurate results (Yang & Lee, 2007). Therefore, most marker-less systems use multiple cameras to capture 3D human motion. Through such systems, the 3D information of the observed human subject is captured from different directional views, thereby providing better results of recovered human motion in 3D (Knossow et al., 2008; Gupata et al., 2008). However, it is usually complicated to setup such a system, because it requires enough space where the cameras can be installed. Also it requires synchronization of the cameras. Thus, there are always some tradeoffs between the flexibility of using a single camera and the ability to get the 3D information using multiple cameras.

Another way of recovering a series of human postures or motion in full 3D is to utilize the information in depth maps. However, there has been little effort to recover 3D human body postures using this approach. Some conventional works to estimate human body postures from depth maps

20 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/recovering-human-body-postures-depth/60284

Related Content

Preference Extraction in Image Retrieval

Pawel Rotter and Andrzej M.J. Skulimowski (2009). *Artificial Intelligence for Maximizing Content Based Image Retrieval* (pp. 235-260).

www.irma-international.org/chapter/preference-extraction-image-retrieval/4159

Imaging by Angioscopy

Michael Rees (2012). *Intravascular Imaging: Current Applications and Research Developments* (pp. 138-146).

www.irma-international.org/chapter/imaging-angioscopy/61077

Recent Trends of Addressing COVID-19 Disease by AI/ML

Shawni Dutta, Utsab Mukherjee and Digvijay Pandey (2023). *Handbook of Research on Thrust Technologies' Effect on Image Processing* (pp. 395-419).

www.irma-international.org/chapter/recent-trends-of-addressing-covid-19-disease-by-aiml/328045

Automatic Analysis of Microscopic Images in Hematological Cytology Applications

Gloria Díaz and Antoine Manzanera (2010). *Biomedical Image Analysis and Machine Learning Technologies: Applications and Techniques* (pp. 167-196).

www.irma-international.org/chapter/automatic-analysis-microscopic-images-hematological/39560

Three-Dimensional Face Shape by Local Fitting to a Single Reference Model

Rubén García-Zurdo (2014). *International Journal of Computer Vision and Image Processing* (pp. 17-29).

www.irma-international.org/article/three-dimensional-face-shape-by-local-fitting-to-a-single-reference-model/111473