Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition Letters 28 (2007) 797–805
Pattern Recognition Letters
www.elsevier.com/locate/patrec

3D human model and joint parameter estimation from monocular image

Minglei Tong [a,*], Yuncai Liu [a], Thomas S. Huang [b]

[a] Institute of Image Processing and Pattern Recognition, Shanghai JiaoTong University, Shanghai, 200030, People's Republic of China
[b] Beckman Institute, University of Illinois at Urbana-Champaign, Urbana, IL 61801, United States

Received 18 October 2004; received in revised form 20 January 2006
Available online 23 January 2007

Communicated by S. Dickinson

Abstract

In this paper we present a novel class of human model described by convolution surface attached to articulated kinematics skeletons. The human pose can be estimated from silhouette in monocular images. The contribution of this paper consists of three points: First, human model of convolution surface is presented and its shape is deformable when changing polynomial parameters and radius parameters. Second, convolution surface and curve correspondence theorem is presented to give a map between 3D pose and 2D contour. Third, we model the human silhouette with convolution curve in order to estimate joint parameters from monocular images and we also give an effective constraint function. Evaluation of this approach is performed on some video frames about a walking man. The experiment result shows that our method works well without self-occlusion.

Keywords: 3D human model; 2D images; Convolution surface; Model initialization; Motion estimation

1. Introduction and related work

Over the course of the past decade, human motion analysis has received more and more attentions in computer vision research community with applications in areas like marker-less motion capture for animation and virtual reality, human computer interaction or intelligent surveillance, etc. Various model-based methods have been proposed for estimation and analysis of full-body structure, at the same time many human models deformable or rigid (DeCarlo and Metaxas, 1996; Shoji et al., 2000; Ju et al., 1996; Gavrila and Davis, 1996; Sand et al., 2003; Plankers and Fua, 2003; Sminchisescu and Telea, 2002; Zhang et al., 2002) are given for human motion capture or tracking.

Our goal in this paper is to build a class of human model using convolution surface, and the shape of this model can be modulated by polynomial function. Convolution surface was firstly introduced into computer graphics by Bloomenthal and Shoemake (1991). In fact convolution surface is a kind of special implicit surface using 3D convolution between skeleton (line or curve) and some kernels. As mentioned in (Jin and Tai, 2001), if the kernels are modulated by a polynomial function along a skeleton, different shape will be produced along the skeleton. Tai et al. (2004) presents prototype modeling from sketched silhouette based on convolution surfaces.

Another goal of the paper is to determine the human model shape and estimate the body's pose from image. We extract silhouette of a human body with different gesture and trace the boundary of the human region in image and approach the orthogonal projection curve of convolution surface with convolution curve. According to the correspondence between 3D surface and 2D contour, we can estimate the 3D joint parameters by fitting the convolution curve with human contour points. The relevant polynomial coefficients, which are used to adjust the shape of the human model, can be estimated in model initialization process.

* Corresponding author. Tel.: +86 21 3420 4028; fax: +86 21 3420 4340.
E-mail address: tongminglei@sjtu.edu.cn (M. Tong).

0167-8655/$ - see front matter
doi:10.1016/j.patrec.2006.11.011

Our main contribution consists of three folds: (1) This paper gives convolution surface human model with articulated skeletons, which is deformable while changing radius parameters and polynomial parameters. (2) This paper presents the convolution surface and curve correspondence theorem under orthogonal projection, which bring a bridge between the 3D pose and 2D contour. (3) We model human contour subtracted from image with convolution curve and estimate joint parameters by fitting the human contour.

In the past decades, much effort has been devoted to deformable human model reconstruction and human motion capture based on image or video in computer vision. Lots of methods are proposed depending on extracting silhouettes and fitting body models to them. The researches on human motion analysis with model have been reviewed in (Moeslund and Granum, 2001; Aggarwal et al., 1998; Gavrila, 1999). Traditionally human model are represented as stick figures (Shoji et al., 2000), 2D contour (Ju et al., 1996) or volumetric model (Gavrila and Davis, 1996).

Recent works are inclined to make more accurate deformable human model with ingenious computer graphics techniques. Bregler and Malik (1998) and Bregler et al. (2000) made use of Factorization techniques, and built the model with a linear combination of prototype shapes. Under this model, the tracking matrix is of higher rank, and can be factored in a three steps process to yield pose, configuration and shape. But the technique has not been applied in full body tracking. Plankers and Fua (2003) proposed "articulated soft objects", in which the skin surface is presented as a series of implicit volumetric primitives attached to skeletons. Each primitive defines a field function (soft objects) and the skin is taken to be a level set of the sum of these fields. The model is optimized by observations from silhouettes and stereo depth. Though Plankers made use of very elaborate method and got a very accurate model, the model requires too many parameters to represent. Sand et al. (2003) described a very novel method for human body reconstruction from silhouettes. They presented the human skin surface using points along needles that are rigidly attached to a skeleton. Sminchisescu and Telea (2002) proposed a human body model which consists of articulated kinematics skeletons covered by 'flesh' built from super-quadric ellipsoids with additional tapering and bending parameters.

Here we do not contribute new technique in the areas of polygon method and silhouette extraction, but in ways of modeling the human contour with convolution curve in image and estimating the joint parameters by fitting the convolution curve with a effective constraint.

The human model is described in Section 2.1; the convolution surface and curve correspondence theorem is proved in Section 2.2; image process is mentioned in Section 3; Parameters estimation method is described in Section 4. The experiments will be described in Section 5.

2. Human model

The human body model in this paper is constructed through convolution between articulated skeletons and Cauchy kernel function, and the whole body needs 32 articulated skeletons.

2.1. Convolution surface and curve

Bloomenthal and Shoemake (1991) introduced convolution surface based skeleton as a logical extension of point based implicit surface models. Jin and Tai (2001) presented an analytical solution for convolving line segment skeletons with a variable kernel modulated by a polynomial function.

A convolution surface is a set of points (x,y,z) that satisfy

$$F(x,y,z) - T = 0 \tag{1}$$

T is constant, and $F(x,y,z)$ is field function, achieved through a 3D convolution of a kernel function with a skeleton function $g(P)$, $P(x,y,z)$ is an arbitrary point on surface:

$$\int_V g(Q) f(P-Q)\,\mathrm{d}v - T = 0 \tag{2}$$

Integrating all points $Q \in V$, and V is skeleton points set

$$g(P) = \begin{cases} 1 & P \in V \\ 0 & P \notin V \end{cases} \tag{3}$$

Eq. (2) can be expressed as the following form:

$$(f \otimes g)(P) - T = 0 \tag{4}$$

A line segment of length l with a start point $\vec{b} = (b_1, b_2, b_3)$ and unit direction $\vec{a} = (a_1, a_2, a_3)$, can be presented parametrically as

$$L(t) = \vec{b} + t \cdot \vec{a}, \quad 0 \leqslant t \leqslant l \tag{5}$$

$\vec{b}$ and $\vec{a}$ are 3 dimensional vectors.

Letting $\vec{d} = \vec{P} - \vec{b} = (x - b_1, y - b_2, z - b_3)$, the squared distance from a point $\vec{P}$ to a point on the line $L(t)$ described in Fig. 1, is

$$\left\|\vec{r}(t)\right\|^2 = \left\|\vec{d}\right\|^2 + t^2 - 2t\vec{d} \cdot \vec{a}, \quad \vec{P} \notin L(t) \tag{6}$$

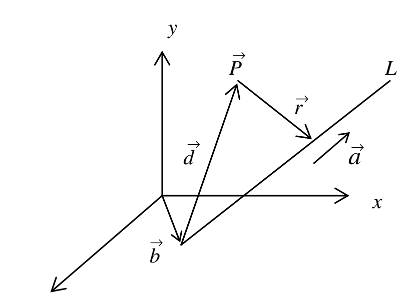

Fig. 1. Line segment L and vector $\vec{r}$.