Volumetric Representation of Human Body Parts Using Superquadrics

Ryo Hachiuma, Hideo Saito
Keio University
{ryo-hachiuma, hs}@keio.jp

Abstract

The idea of automatically sensing the 3D human body has been of interest in computer vision. The 3D human body can be represented in many ways, such as meshes, parametric models and joint skeletons. We represent the human body into multiple 3D primitive shapes.

Method

To represent each joint by superquadrics, the point cloud of each joint must be extracted. Therefore, we first semantically segment human body into joint parts. Next, we estimate superquadric parameters of joints.

1. Body joints segmentation

We employed Light-weight RefineNet [5] trained on PASCAL Person-Part dataset [6]. To segment left/right limbs, we extract 3D joints by the method proposed by Shotton et al. [7]. For each pixel labeled as a limb, we calculate 3D Euclidean distance between 3D joints and each pixel. If the closest joint is left limb, the pixel is labeled as left part, and vice versa.

2. Superquadric estimation

Superquadric and its pose parameters \( \Lambda(e_1, e_2, s_1, s_2, s_3, t_1, t_2, t_3, \theta_1, \theta_2, \theta_3) \) are estimated from object point cloud \( \{(x_q, y_q, z_q)\} \) \( 0 \leq k \leq M \) by Levenberg-Marquardt algorithm.

As the minimization function is not a convex function, the initial parameters determine which local minimum the minimization converges to.

We propose a novel approach to estimate initial parameters using 3D skeleton joints to find optimal parameter.

Translation:

Jaklic et.al. [8]: the 3D centroid of 3D points.
Ours: the 3D centroid of 3D joints.

Rotation:

Jaklic et.al. [8]: Eigen vectors of 3D points.
Ours: aligns the z-axis of superquadric surface to be parallel to the vector of two connected 3D joints in each body part.

Superquadrics

Superquadrics [1] can represent various types of primitive shapes with three scale parameters and two shape parameters.

\[
f(x_q, y_q, z_q) = \left( \left( \frac{x_q}{s_1} \right)^2 + \left( \frac{y_q}{s_2} \right)^2 + \left( \frac{z_q}{s_3} \right)^2 \right)^{e_1} + \left( \left( \frac{x_q}{s_1} \right)^2 + \left( \frac{y_q}{s_2} \right)^2 + \left( \frac{z_q}{s_3} \right)^2 \right)^{e_2} = 1
\]

\( s = (s_1, s_2, s_3) \): Scale parameters
\( e = (e_1, e_2) \): Shape parameters

Superquadrics is employed to approximate shape for object shape understanding [2], object grasping [3], collision detection [4].

Result

We record three sequences using Kinect v.2. In each sequence, the person is standing in front of the camera.

Baseline Method: We compared the estimation results with the previous work [8]. This previous method has been widely used for estimating the initial parameter estimation.

1. Qualitative results

Our proposed method successfully estimated multiple superquadric parameters which approximate the point cloud of each body part.

2. Quantitative results

We employed Chamfer distance as an evaluation metric if the estimated superquadric surfaces represented the original point cloud.

Chamfer distance \( d \text{[m]} \)

<table>
<thead>
<tr>
<th>seq1</th>
<th>seq2</th>
<th>seq3</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>1.735</td>
<td>1.671</td>
<td>1.535</td>
</tr>
<tr>
<td>proposed</td>
<td>1.105</td>
<td>1.533</td>
<td>1.349</td>
</tr>
</tbody>
</table>

We can verify that our novel initial parameter estimation method found more optimal parameters than the previous method [8].

References