ABSTRACT Human action recognition using 3D camera for surveillance application is a promising alternative to the conventional 2D camera based surveillance systems. We evaluate a body part classification method using various configuration of feature extraction for the human action recognition. Experimental results shows the body part classification accuracy of 74.46% which is sufficiently high accuracy for the subsequent joint detection and human action recognition.

Keywords: Body part classification, action recognition, 3D camera, depth image.

1. INTRODUCTION

Random forest based body part classification

Random forest is an ensemble of decision trees. It classifies each pixel in depth images into one of the predefined body parts [5]. The features used in the random forest for body part classification are based on depth gradients defined as

$$
\phi(l, x) = d_l(x + \frac{\delta_1}{\alpha}) - d_l(x + \frac{\delta_2}{\alpha}), \alpha = |l, d_l(x)|,
$$

where $l$ is an image, $x$ is a pixel, $d_l(\cdot)$ is the depth value, $\delta_1$ and $\delta_2$ are the displacement from $x$, and $\alpha$ is a normalizer. Figure 1 shows a schematic of a unary and binary features taken from the center pixel of a window. For the feature extraction, pixels around a human depth image are randomly selected and one (for unary feature) or two (for binary feature) other points are selected for the feature construction. We randomly selected 2,000 pixels and used a window size of 60x60. For each window, we used a grid configuration shown in Figure 2 for the usnary feature extraction.

3. EXPERIMENTAL RESULTS

We used two synthetic 3D human models and Mocap data [6] to generate depth images of various actions. From three different actions (i.e., jump, kick, and punch) and three different viepoints (i.e., frontal, right, left), we chose frontal jumn, left kick, left punch, right punch, and frontal punch actions for the evaluation.

$$
\phi(l, x) = d_l(x + \frac{\delta_1}{\alpha}) - d_l(x + \frac{\delta_2}{\alpha}), \alpha = |l, d_l(x)|,
$$
For each sequence of actions, 100 training images and 16 test images are randomly selected. Experimental results shows that the grid configuration (1,3) is the best for the body part classification. Table 1 shows summary of body part classification for five different actions.

![Figure 1. Schematics of (a) unary, (b) binary features, and (c) grid configurations](image)

**Table 1. Body Part Recognition Accuracy for Five Actions**

<table>
<thead>
<tr>
<th>Action</th>
<th>Mean accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jump (frontal)</td>
<td>65.85</td>
</tr>
<tr>
<td>Kick (left view)</td>
<td>81.70</td>
</tr>
<tr>
<td>Punch (left view)</td>
<td>88.64</td>
</tr>
<tr>
<td>Punch (right view)</td>
<td>89.43</td>
</tr>
<tr>
<td>Punch (frontal)</td>
<td>90.10</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>74.46</strong></td>
</tr>
</tbody>
</table>

4. CONCLUSIONS

We described body part classification method using depth gradient features with various feature extraction configurations. Body part classification is a fundamental step required most of the 3D human action recognition systems. From the body part classification result, we are developing joint detection and human action recognition system.

ACKNOWLEDGMENT

This work was supported by the ICT R&D program of MSIP/IITP. [B0101-16-0525, Development of global multi-target tracking and event prediction techniques based on real-time large-scale video analysis]

REFERENCES

6. CMU Mocap Database (2016, November 21). Received from http://mocap.cs.cmu.edu/