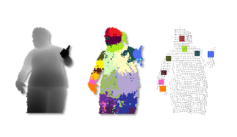
Introduction

Kinect V2 - a vision device for detecting position, shape and dimensions of an object, a controller-free game controller, tracking the body or bodies of one or more players in its field of view.

Kinect v2

overview

With the newly introduced sensors and approaches in the field of computer vision, Kinect v2 introduces interactive human body tracking which has multiple applications s including gaming, human-computer interaction, security, telepresence, and even healthcare.



(->depth image -> segmented body parts -> skeletal landmarks)

Specs

Overview

|  |  |  |
| --- | --- | --- |
| Color sensor | Camera resolution  Frame rate | 1920 x 1080  30 fps |
| Depth sensor | Camera resolution  Frame rate | 512 x 424  30 fps |
| Field of view (depth) | Horizontal  Vertical | 70 degrees  60 degrees |
| Range | 0.5-4.5 m | |
| Depth technology | Time-of-flight (tof) | |
| Tilt motor | no | |

Time of Flight

the measurement of the time taken by an object, particle, or wave to hit a medium (a laser or an LED))

Text, shape

Description automatically generated

Human Pose Recognition - Microsoft Research Cambridge & Xbox Incubation

Data

depth images

self collect a large database of human actions.

The human body is capable of an enormous range of poses which are difficult to simulate. Instead, capture a large database of motion capture (mocap) of human actions.

use clustering to get discard similiar, redudant frame ( repetitive poses)

We need not record all possible combinations of the different limbs; in practice, a wide range of poses proves sufficient.

use Autodesk MotionBuilder to render the depth image from the frame extracted.

Depth image features

A picture containing text

Description automatically generated

Depth image features. The yellow crosses indicate the pixel x being classified. The red circles indicate the offset pixels in (a), the two example features give a large depth difference response. In (b), the same two features at new image locations give a much smaller response.

At a given pixel x, the features will compute differences between the depth of a given pixel X and its offset ( u and v). If an offset pixel lies on the background or outside the bounds of the image, the depth probe is given a large positive constant value.

illustrates two features at different pixel locations x. Feature θ1 looks upwards. Hence, will give a large positive response for pixels x near the top of the body, but a value close to zero for pixels x lower down the body. Feature θ2 may instead help find thin vertical structures such as the arm.

Randomized decision forests

Randomized decision trees and forests have proven fast and effective multi-class classifiers.

Chart, line chart

Description automatically generated

forest is an ensemble of T decision trees, each consisting of split and leaf nodes. Each split node consists of a depth feature and a threshold.

To classify pixel x in image I, one starts at the root and repeatedly evaluates Eq. 1, branching left or right according to the comparison to threshold τ . At the leaf node reached in tree , a learned distribution over body part labels is stored.

The distributions are averaged together for all trees in the forest to give the final classification.

Training

Each tree is trained on a different set of randomly synthesized images. A random subset of 2000 example pixels from each image is chosen to ensure a roughly even distribution across body parts.

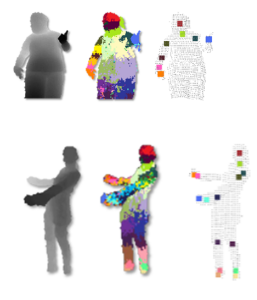
A picture containing different, several

Description automatically generated

Joint estimation

The segmented body parts and its infered information must then be pooled across pixels to generate reliable proposals for the positions of 3D skeletal joints.

Employ local mode-finding approach based on mean shift - D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis.



application in making educational game