Toad Observation by Analysis of Depth

Armando Califano, Dr. Damian Lyons, Avishai Shuter

Fordham University Robotics and Computer Vision Lab

Objective

Working in conjunction with the Bronx Zoo, we are creating a more efficient way to monitor the Kihansi spray toads with the use of RGB-D cameras. Monitoring the toads in a 3D image will give us the ability to track the toads over a period of time, with the end goal of having automated tracking for several toads at once.

Kihansi Spray Toads

- Native only to a single waterfall in Tanzania
- Extinct in the wild but are kept in captivity, only located at the Bronx and Toledo Zoos
- Bronx Zoo holds over 3000 toads in over 30 tanks
- Adult sizes range from 1 to 3 cm

a) Close-up of Kihansi spray toad

b) Small cluster of toads

Challenges

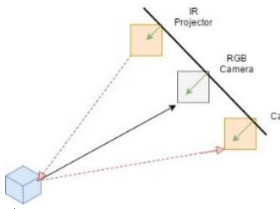
- Small size sheer quantity of toads make them hard to observe using standard recording equipment
- Quantity and size make it hard to observe with human eye
- To observe an object in 3D space, the target must be large enough for the RGB-D camera to identify it, but the small size of the toads make them very difficult to identify in a depth image.
- Once we can identify them and track them, we can automate this feature for long-term observation

RGB-D Cameras

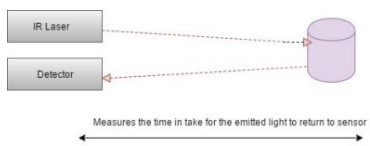
- The RGB-D cameras record normal color video along with depth readings
- Records depth with IR projector and IR camera that reads the emitted light
- We use the Microsoft Kinect for Windows v1 and v2

a) Kinect for Windows v1

b) Kinect for Windows v2



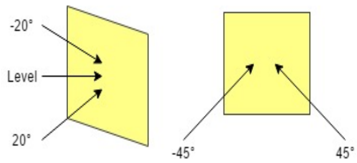
c) Kinect v1 IR projector



d) Kinect v2 time-of-flight

Data Collection

- The Kinects require a minimum distance of 40 cm from the target before the objects appear in depth
- The cameras were set up at five different angles to optimize the chances of visualizing a toad in space

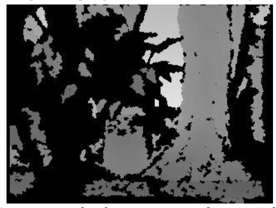


a) Five different camera angles to record RGB-D observations

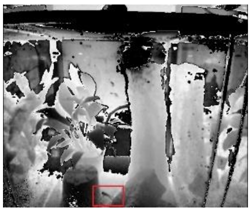
b) Different camera set-ups for the Kinect v1

Extracting Depth Image from Video Sequence

- Both the normal RGB visual frames and the depth frames were extracted
- Depth images are in greyscale, meaning objects in darker grey are closer and those in lighter grey are further away



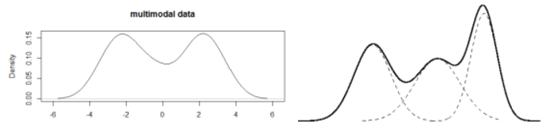
a) Kinect v1 depth image cannot locate toad



b) Kinect v2 depth image with toad

Background Segmentation using Mixed Gaussian Model

- To focus on only the toad, the background of the image needed to be removed
- Mixed Gaussian Model contains multimodal data
- Each cluster has a mean and covariance matrix



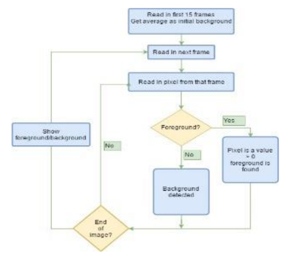
Each pixel in the scene is modelled by a mixture of K Gaussian distributions. The probability that a certain pixel has a value of $\mathbf{x}_N$ at time N can be written as

$$p(\mathbf{x}_N) = \sum_{j=1}^{K} w_j \eta(\mathbf{x}_N; \boldsymbol{\theta}_j)$$

where w_k is the weight parameter of the k^th^ Gaussian component. $\eta(\mathbf{x}; \boldsymbol{\theta}_k)$ is the Normal distribution of k^th^ component represented by

$$\eta(\mathbf{x}; \boldsymbol{\theta}_k) = \eta(\mathbf{x}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \frac{1}{(2\pi)^{\frac{D}{2}} |\boldsymbol{\Sigma}_k|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1}(\mathbf{x}-\boldsymbol{\mu}_k)},$$

where $\boldsymbol{\mu}_k$ is the mean and $\boldsymbol{\Sigma}_k = \sigma_k^2 \mathbf{I}$ is the covariance of the k^th^ component.

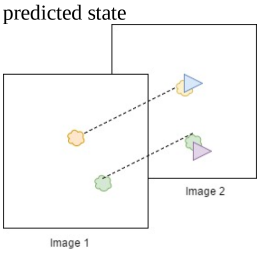


a) Original depth image

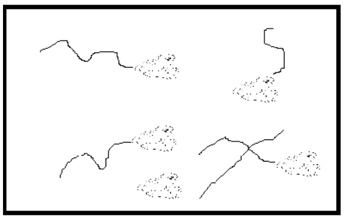
b) Foreground image

Nearest Neighbor Data Association

- Tracking method that updates the tracking data with the measurement closest to predicted state



- Find closest image to original object and map to that
- Find center of the objects that are matched together and create a track



- Final result will be a set of tracks