SPARK+AI SUMMIT 2018

Fiducial Marker Tracking Using Machine Vision

Saurabh Ghanekar, Kavi Global
Kazutaka Takahashi, University of Chicago

#AISAIS14

Outline

- Motivation & Goals
- Approach
- Results
- Next Steps

Motivation

- Feeding is a highly complex, life-sustaining behavior, essential for survival in all species
- Certain neurological conditions such as Parkinson's disease, ALS, stroke can cause difficulty in chewing and swallowing, known as dysphagia
- Affects quality of life
- Dysphagia can lead to malnutrition, dehydration, and aspiration

End-Goal

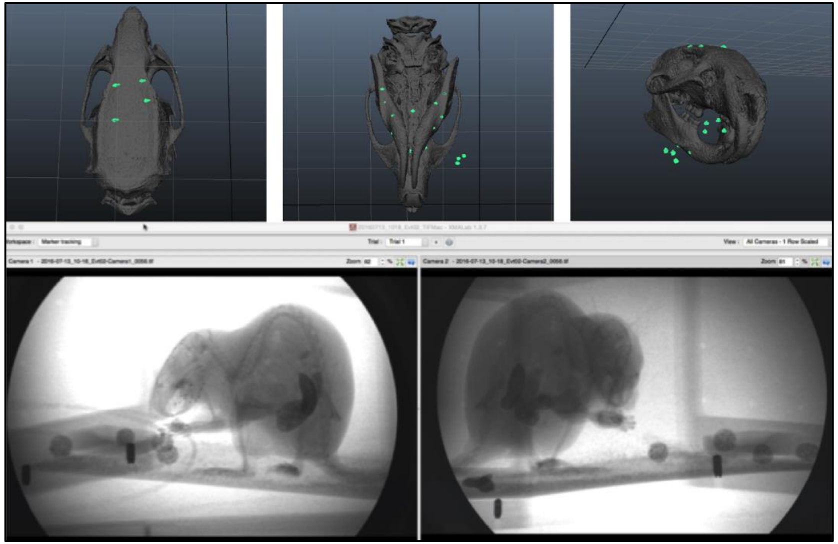
To characterize feeding dynamics and gain insights into feeding behavior changes caused by certain neurological conditions and changes in oral environment.

Current State

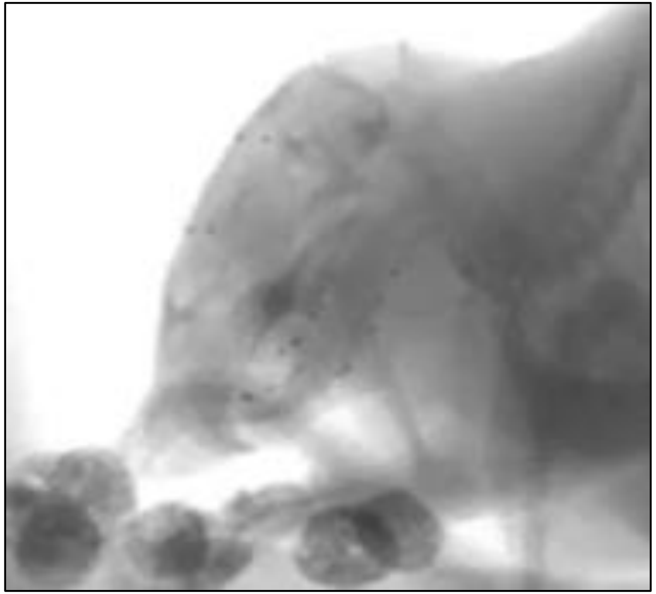
- Study focused on rodents
- X-ROMM videos of rodents feeding on kibble
- Videos recorded from 2 camera angles simultaneously
- Radio-opaque markers implanted in skull, mandible, tongue
- Movement of markers needs to be tracked and quantified
- Marker tracking process is extremely tedious as it is done using manual, frame-by-frame methods [1,2]
- Consumes valuable time, thus delaying further research

Immediate Goal

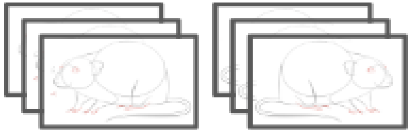
A near-automated, deep learning-based solution for detecting and tracking markers, resulting in a more efficient and robust process



(c) Bunyak et al, 2017

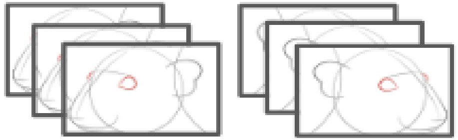


Approach: Key Steps



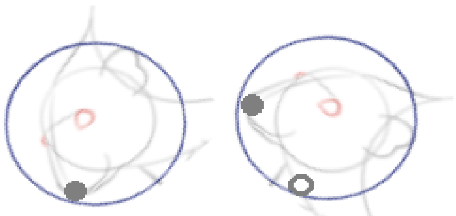
Data In:

Read in videos frame by frame for left and right cameras in 2D (x,y)



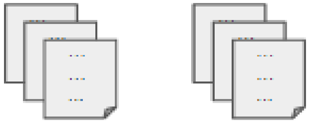
Head and Marker Detection:

Utilize neural network to identify bounding box of head and also pinpoint unlabeled markers inside the bounding box.



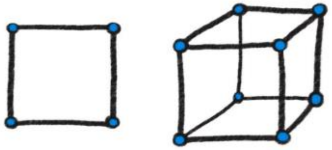
Marker Tracking:

Employ Kalman filters along with Hungarian algorithm to keep track of markers from frame to frame



Sequence Matching:

Match sequence tracks from left and right cameras



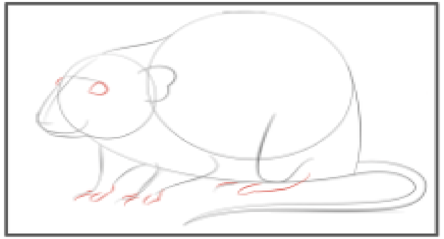
2D to 3D conversion:

Feed 2D left right coordinates along with rotational matrices and translation vector to get final 3D coordinates (x,y,z)

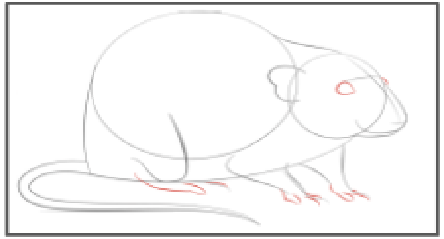
Data Description

- 13 pairs of videos (left & right camera) available for training
- 720px by 1260px videos, recorded at 250 fps, ~10 seconds each
- Head and marker coordinates per frame used for model training & evaluation
- 18-20 markers to be tracked in each video

Camera 1



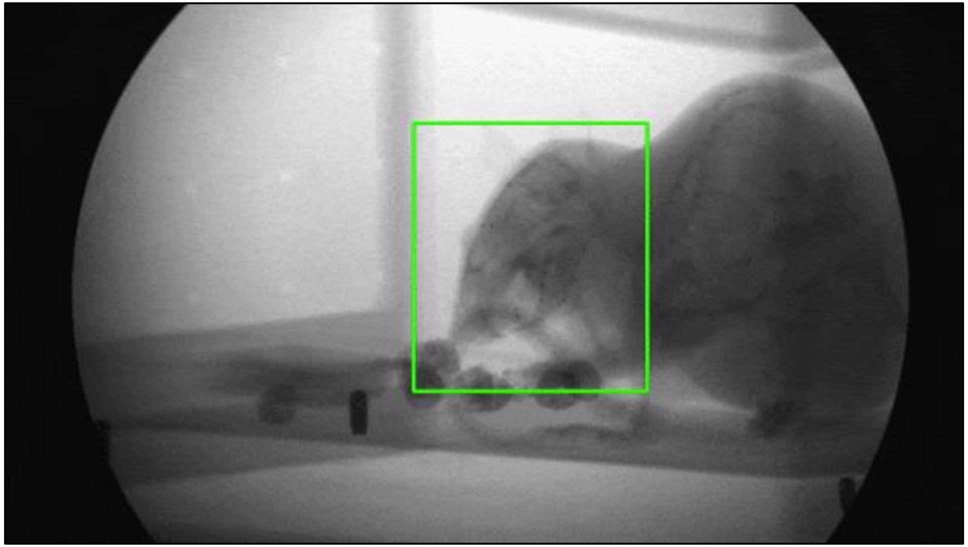
Camera 2

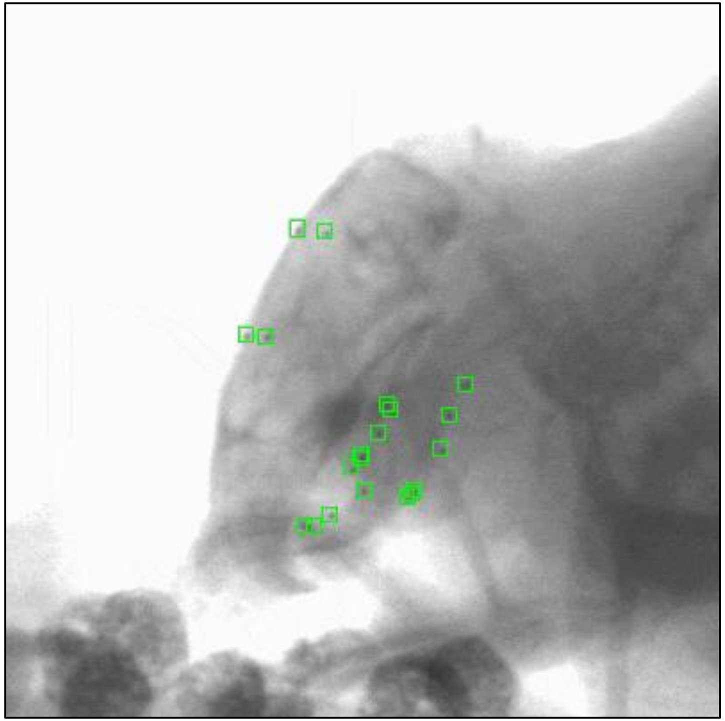


Head and Marker Detection

- TensorFlow's Object Detector API
- Single Shot Multibox Detector (SSD) with MobileNet using transfer learning from the MS COCO dataset
- Key model parameters:
  - Initial Learning Rate: 0.0004
  - Feature Extractor Type: ssd_mobilenet_v1
  - Minimum Depth: 16
  - Depth Multiplier: 1.0
  - conv_hyperparams: activation: RELU_6; regularizer: l2_regularizer; weight: 0.00004

Head and Marker Detection

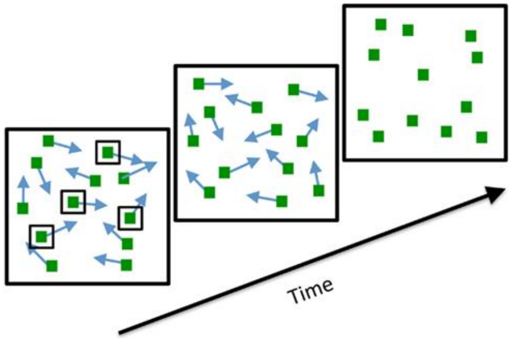




Marker Tracking

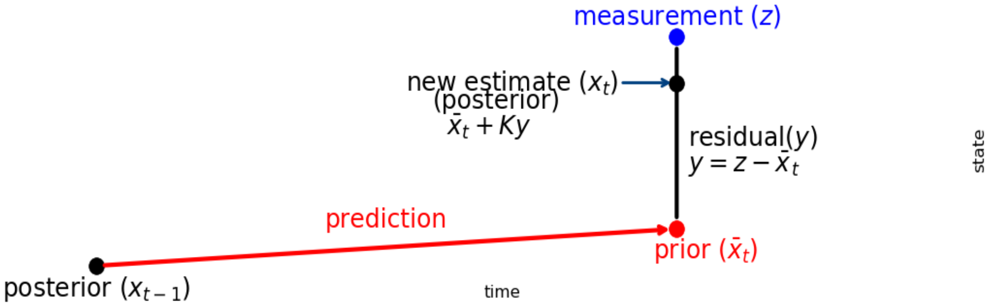
Multi-object tracking involves three key components:

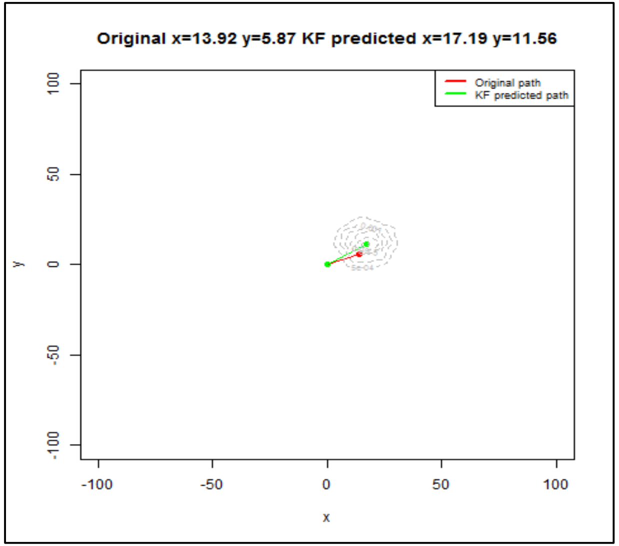
- Predicting the object location in the next frame
- Associating predictions with existing objects
- Track Management



(c) Howe, Holcombe, 2012

Prediction

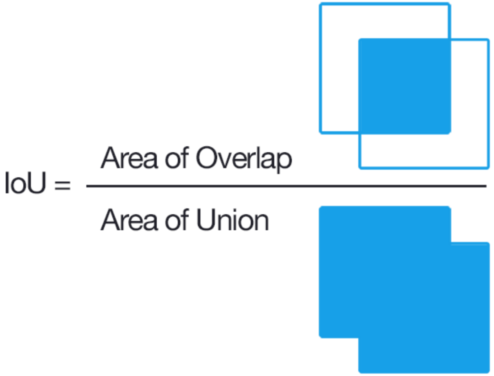




- Kalman Filter is used to predict marker location in the next frame
- Estimate position recursively in each frame, based on previous frames
- Uses Bayesian learning and estimates a joint probability distribution
- Start with initial velocity estimate & covariance matrix

Association

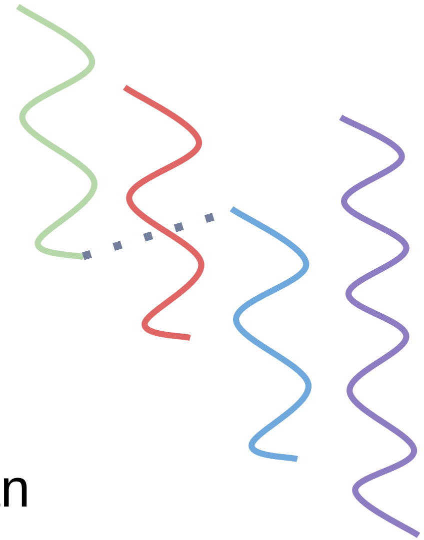
- After prediction, an assignment cost-matrix is computed from the bounding-box intersection-over-union (IoU)
- Hungarian Algorithm is used to optimally associate markers



True Marker Positions

Predicted Marker Positions	0	1	2	...
0	518	101	312	
1	24	$\boxed{963}$	225	
2	$\boxed{872}$	20	220	
...				

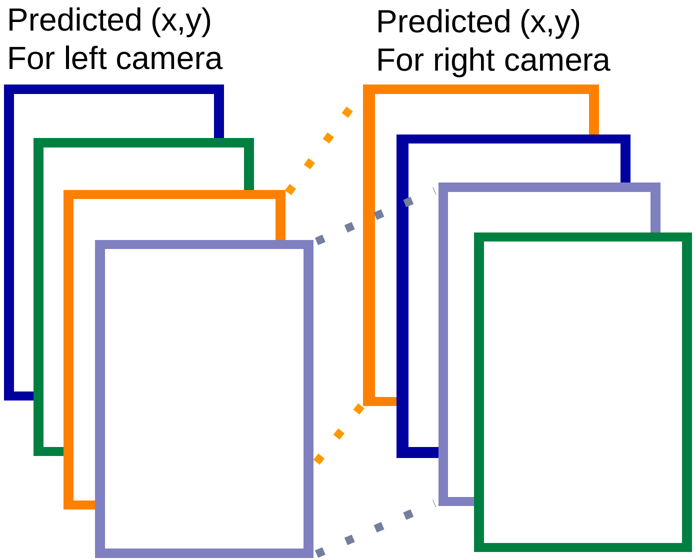
Track Management



- If IoU is below a set threshold, there is no assignment
- Also, not all potential tracks become actual tracks
- As a result, tracks may die and new ones are born
- The output of Kalman filter and Hungarian algorithm can result in a large number of discontinuous tracks
- These are “stitched” together by looking forward and backward a number of frames to find the best match based on closest Euclidean distance
- At the end, we get one track per marker

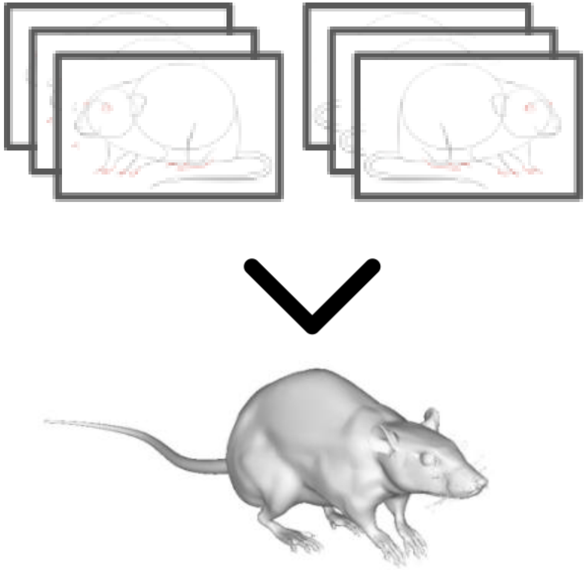
Sequence Matching

- After generating marker tracks separately for each camera, corresponding tracks from each camera must be matched
- Tried different distinct methodologies such as Time Series Clustering and different correlation measures.
- Spearman correlation on frame-to-frame changes in Y-coordinate values gave the best results (100% accuracy on manually tracked data)



2D to 3D Conversion

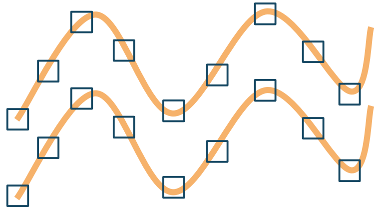
- $P = K * (R \mid T)$ - Camera Projection Matrix (3x4) for each camera
  - K = Camera Matrix (3x3)
  - R = Rotation Matrix (3x3)
  - T = Translation Vector (3x1)
- Results are a good match with actual 3D coordinates



Evaluation 1: % IoU Difference

Step 1: Calculate a perfect overlap. Sum the area of the boxes over each frame for each true stream

True Streams

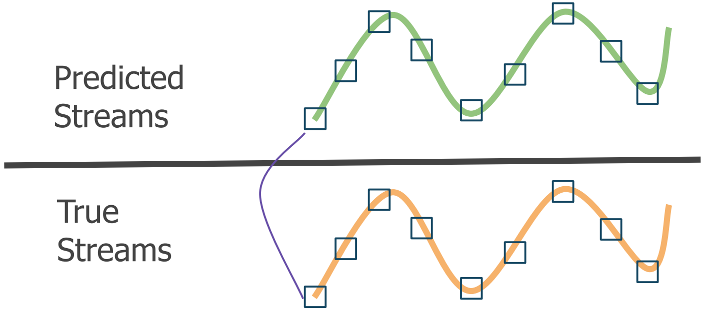


Area Stream: (box area)*(number of frames)

Step 2: Calculate the percent difference between the perfect area and actual IoU

Predicted Streams

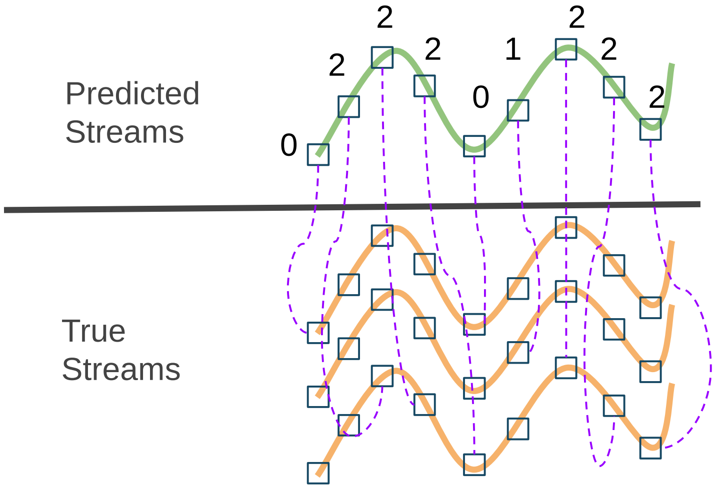
True Streams



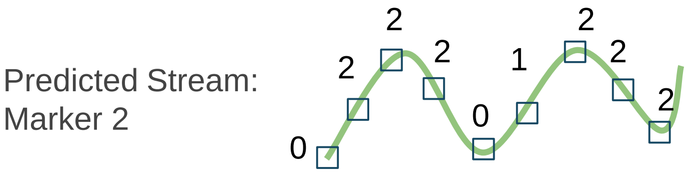
$$\%\ \text{IoU Difference} = \frac{Perfect\ Overlap\ Area - Actual\ IOU}{Perfect\ Overlap\ Area}$$

Evaluation 2: % Correctly Labeled

Step 1: Determine the best matching marker label for each frame in the predicted streams using the maximum IOU in each frame

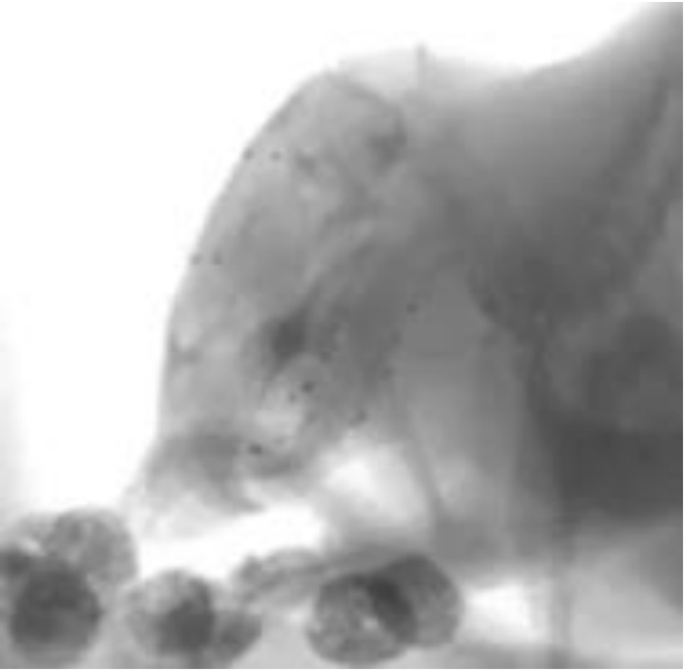


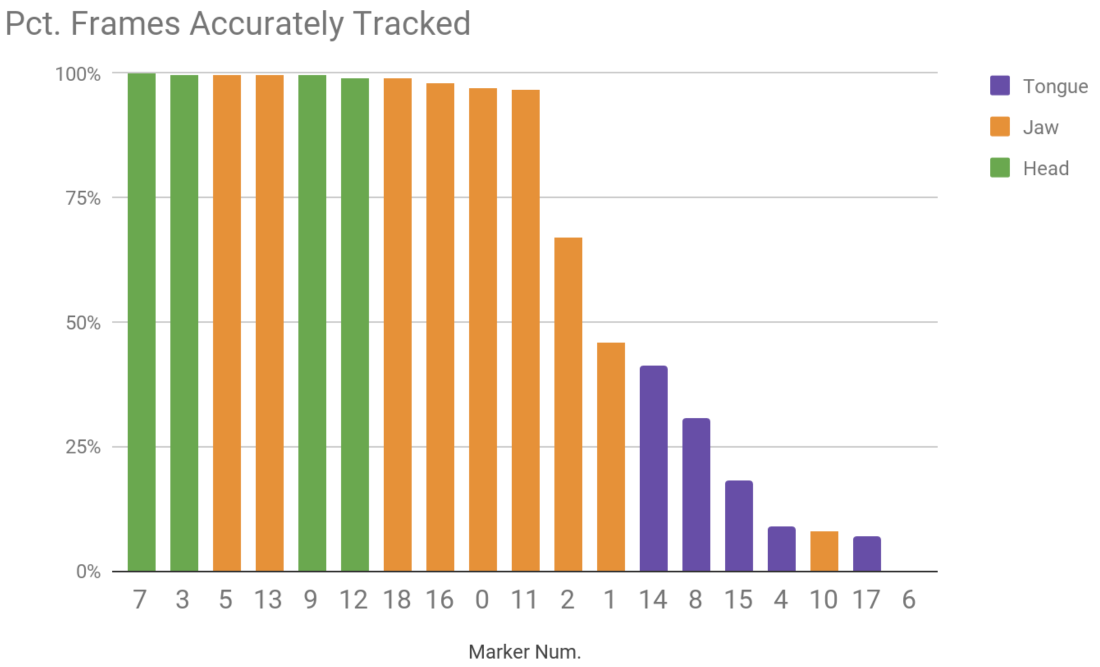
Step 2: Calculate the percent of frames labeled with the same label as the overall label given in the tracking phase



$$Percent\ Correctly\ Labeled = \frac{6}{9} \cdot 100\%$$

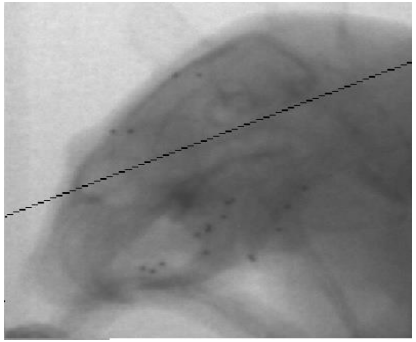
Results



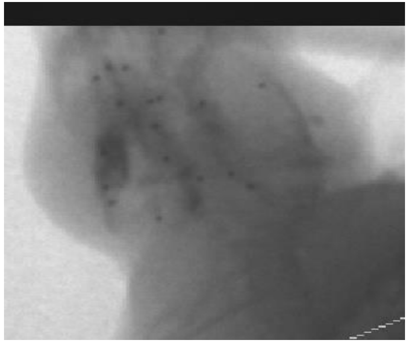


Pct. Frames Accurately Tracked
100%
75%
50%
25%
0%
7 3 5 13 9 12 18 16 0 11 2 1 14 8 15 4 10 17 6
Marker Num.
Tongue
Jaw
Head

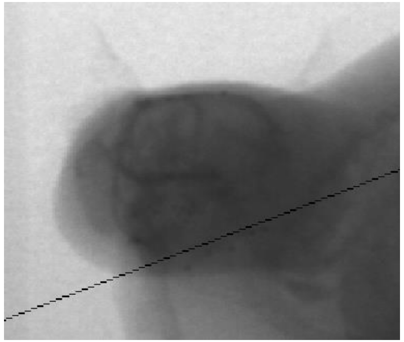
Challenges



Ideal scenario



Off-screen Markers



Occluded Markers

Even if marker detection and tracking models perform well, the above problems may negatively impact results since at some point a marker may be assigned to the incorrect track

Next Steps

- Detection
  - Tune marker detection thresholds, and marker assignment thresholds
- Kalman Filter
  - Tune initialization velocities, and acceleration and covariance matrices
  - Better initialization is known to produce better predictions
  - Non-linear methods (Extended Kalman Filters, particle filters)
- Marker Detection Assignment to Kalman Tracks
  - Currently using Hungarian Assignment. Other options include Probabilistic Assignment, Markov Chain Monte Carlo methods
- Stitching
  - Tune parameters and algorithm to better match together disparate tracks

References

[1] Bunyak F, Shiraishi N, Palaniappan K, Lever TE, Avivi-Arber L, Takahashi K. Development of semi-automatic procedure for detection and tracking of fiducial markers for orofacial kinematics during natural feeding. Conference proceedings : Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Annual Conference. 2017;2017:580-583. doi:10.1109/EMBC.2017.8036891.

[2] Best MD, Nakamura Y, Kijak NA, et al. Semiautomatic marker tracking of tongue positions captured by videofluoroscopy during primate feeding. Conference proceedings: . Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Annual Conference. 2015;2015:5347-5350. doi:10.1109/EMBC.2015.7319599.

[3] Howe PDL and Holcombe AO (2012) The effect of visual distinctiveness on multiple object tracking performance. Front. Psychology 3:307. doi: 10.3389/fpsyg.2012.00307

Thank You!

Saurabh Ghanekar
Principal Consultant
Kavi Global
saurabh@kaviglobal.com

Kazutaka Takahashi, Ph.D.
Research Assistant Professor
Department of Organismal Biology and Anatomy
University of Chicago
kazutaka@uchicago.edu

Funding Information:

- National Center for Advancing Translational Sciences of the National Institutes of Health (UL1 TR000430)
- JSPS The Strategic Young Researcher Overseas Visits Program for Accelerating Brain Circulation (S2504)
- JSPS KAKENHI (JP16K11589)

Acknowledgements:

- Dr. Naru Shiraishi, Niigata University, Japan (for experimental procedure development and data collection)
- Animal Research Center (ARC) staff at the University of Chicago