Greenstand Duplicate Detection and Update Feature

Problem Space and System Architecture

Data Point Capture

Data Properties

- Image
- Timestamp
- GPS Coordinate
- GPS Accuracy
- Device id
- Planter identifier

Data Facts

- Very dense GPS locations
- GPS accuracy is not good enough for unique identification

Problem Space

- Detect when the same picture has been submitted multiple times

- Detect when the same tree has been captured multiple times

- Connect multiple captures of the same tree across moderately distant timestamps

Same Image Submitted Multiple Times

- Images will never be exactly the same, but sometimes ‘cheaters’ will take the same picture multiple times

- We need to detect and flag these so they are not counted as separate trees

- Double counting in this case would mostly apply to a single individual

Detect when the same tree has been captured multiple times

- Sometimes a tree will be captured from multiple vantage points. It should not be counted more than once.

- In this case multiple planters could try to capture the same tree.

- This scenario applies to tree captures that have nearby timestamps.

Connect multiple captures of the same tree across moderately distant timestamps

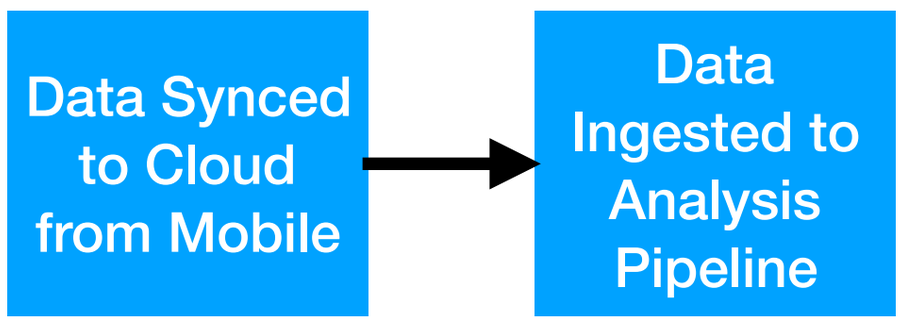
- For tree updates, additional tree captures are valid. These captures do not have nearby timestamps, they represent updated status about the tree.

- We need to make the best possible guess to connect new tree captures with the past captures.

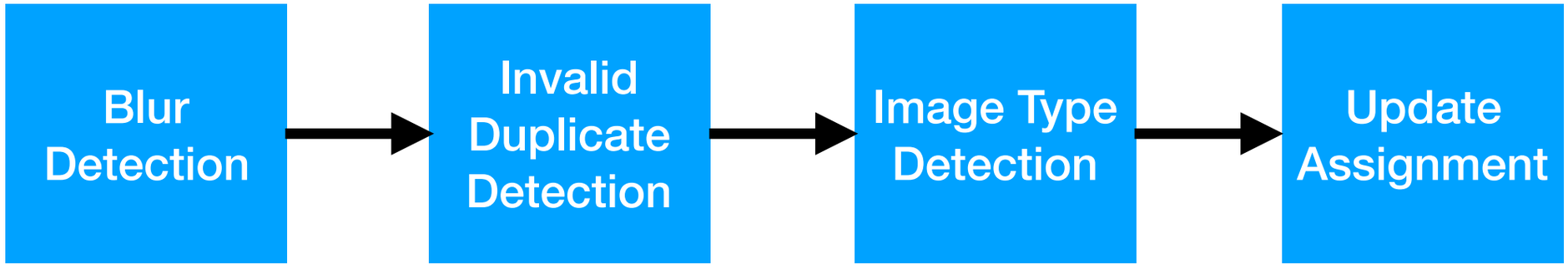
- Trees grow over time and captures are not taken from the same heading, making this a difficult problem.

Data Ingestion Pipeline

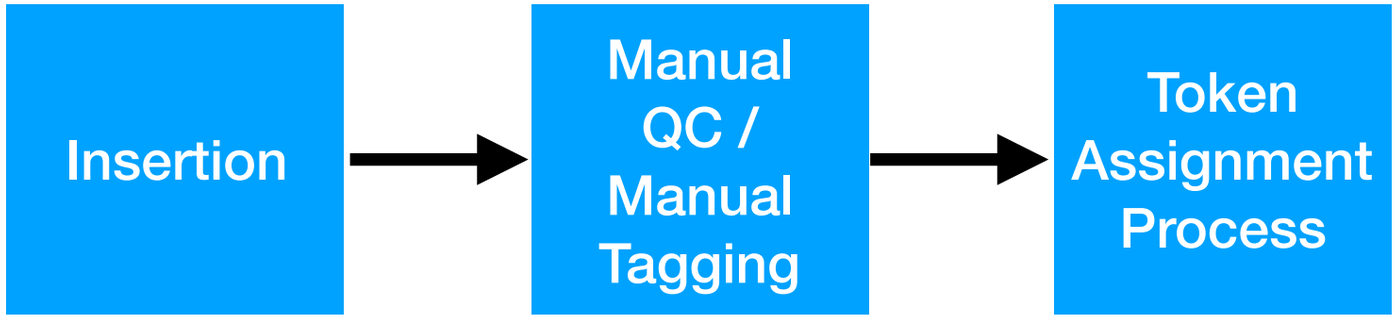
Data Sync



Data Analysis



Data Approval



Some Ideas for the Update Feature

- Improve GPS using Kalman Filter
- Force user behaviors
- Probabilistically assign multiple matches, with a match score, for manual QC
- Use proximity of multiple trees to design a ‘solver’.