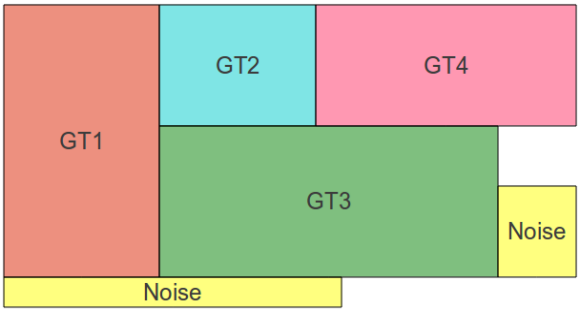
Terms and Notations Used

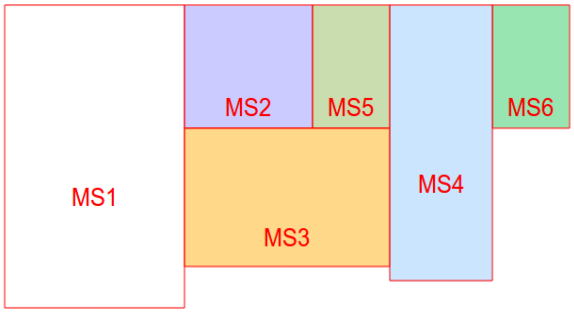
- **Original Image:** Refers to the input/scanned data for the 3D segmenteation algorithms. They would contain the depth information of every point in the image with/without additional information like RGB, intensity, etc. A 2D schematic would be:

- **Human-Segmented Image:** This would be the result of segmentation of the image into separate logical regions done by a human expert. An example of human segmentation for the earlier example would be:



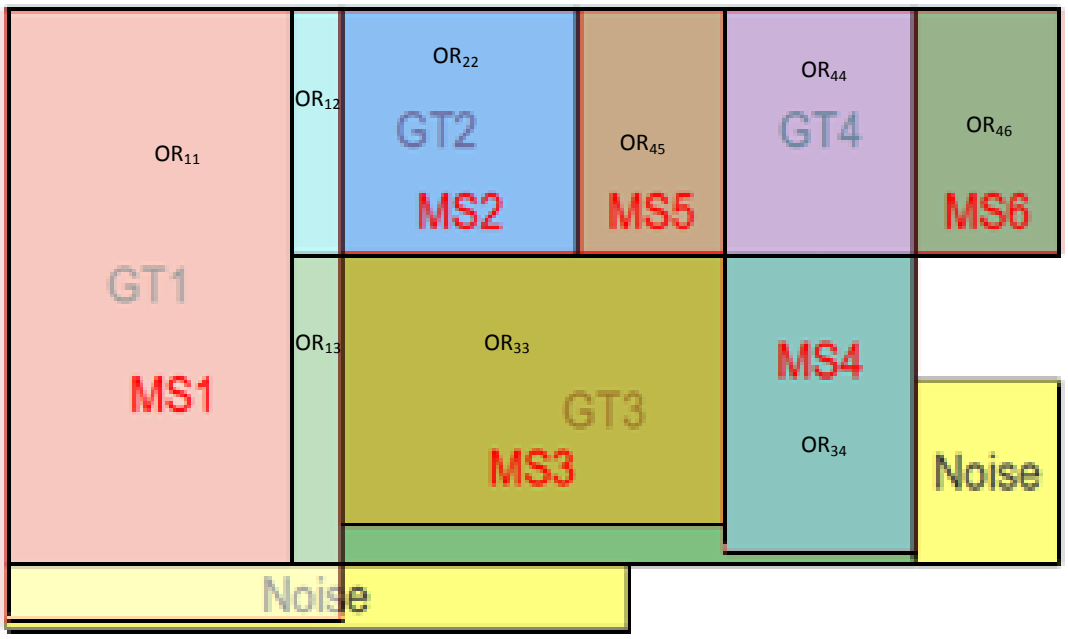
- **Ground Truth Region:** Every region in a Human-Segmented Image is denoted as a ground truth region. Each region will be denoted by (GT_i) in the rest of the article.

- **Machine Segmented Image:** The result of segmentations by the algorithm being evaluated is denoted as a Machine Segmented Image. For the original image given above, a machine-segmented result would be:



- **Machine Segmented Regions:** Every region in a Machine-Segmented Image is denoted as a machine segmented region. Each region will be denoted by (MS_j) in the rest of the article.

- **Overlap Regions:** Each intersection of GT_i and MS_j will be defined as an overlap region O_{ij}. Example overlap regions for example given before would be



Evaluation Metrics:

The idea of our evaluation framework will be based on the notion, **how well the seperate regions are classified by an algorithm**. This is formulated on the definition of 3D segmentation as proposed in [1] :

"Segmenting a range image is the process of labeling the pixels so that pixels whose measurements are of the same surface are given the same label."

The evaluation metrics are divided into three parts:

1. **Quality of identifying an individual ground truth region, GT_i**

A ground truth region (GT_i) can be a table-top, a cup, a box or any distinguishable region. It can be defined as *a cluster of pixels that have been assigned a single label* in the ground truth segmentation image by a human.

These metrics will be used to measure, how well a segmentation algorithm assigns a *single label* to the corresponding pixels of the GT in the raw data.

- **Amount of correct identification:** refers to the percentage of pixels in GT_i which are correctly identified to be GT_i. The machine segmented region MS_j with maximum number of inliers overlapping with GT_i as the correct region i.e the largest of all O_{ij}. The rest of the detected regions will be considered as over-segmented parts.

  Amount of correct detection = $[\max(\#\ O_{i1}, \#\ O_{i2}, \#\ O_{i3....}, \#\ O_{im})] / [\#\ GT_i] * 100$

- **Amount of non-classification:** Percentage of pixels in GT_i which are not labeled in any of the regions in machine segmented image.

  Amount of non-classification = $[\#GT_i - (\#\ O_{i1} + \#\ O_{i2+} \#\ O_{i3....} + \#\ O_{im})] / [\#\ GT_i]$

- **Amount of over-segmentation:** Over segmentation will be denoted by two metrics.
  - No. of over-segmented regions
  - Percentage of pixels in over-segmented regions =

$[(\#\ O_{i1} + \#\ O_{i2+} \#\ O_{i3....} + \#\ O_{im}) - \max(\#\ O_1, \#\ O_2, ..., \#\ O_n)] / [\#GT] * 100$

2. **Quality of each machine segmented region, MS_j**

- **Amount of wrong classification:** Percentage of pixels in MS_j, which does not belong to the GT_i for which it has the maximum overlap region.

  $= [\#MS_j - \max(\#O_{1j}, \#O_{2j}, \#O_{3j.....}, \#O_{nj})] / [\#MS_j] * 100$

- **Under-segmentation:** will be denoted by two metrics
  - No of other regions being merged in M_{Sj} other than GT_i for which it has the maximum overlap region.
  - Percentage of pixels in M_{Sj} which are not from than GT_i for which it has the maximum overlap region =

    $[(\#O_{1j}, \#O_{2j}, \#O_{3j.....}, \#O_{nj}) - \max(\#O_{1j}, \#O_{2j}, \#O_{3j.....}, \#O_{nj})] / [\#M_{Sj}] * 100$

- **Noise classification:** Percentage of pixels in M_{Sj} which doesn't belong to any region in the ground truth segmented image. These are the noise pixels in the raw sensor data. This metric will signify the filtering capability of the approach against noise in sensor data.

  $$= [\#MS_j - (\#O_{1j}, \#O_{2j}, \#O_{3j.....}, \#O_{nj})] / [\#MS_j] * 100$$

3. **Quality of the overall result**

   - **Time consumption**

     Time consumed to produce the machine segmented region(s).

   - **Memory consumption**

     Memory

References

[1] Adam Hoover, Gillian Jean-Baptiste, Xiaoyi Jiang, Patrick J. Flynn, Horst Bunke, Dmitry B. Goldgof, Kevin Bowyer, David W. Eggert, Andrew Fitzgibbon, and Robert B. Fisher. **An experimental comparison of range image segmentation algorithms.** *IEEE Transactions on Pattern Analysis and Machine Intelligence, 1996*

[2] http://meshlab.sourceforge.net/