EE604A: Project Presentation

Automatic Grain Quality Assessment

Prakhar K, Satyam Dwivedi, Anurendra Kumar

Instructor: Prof. Tanaya Guha
Dept. of EE, IIT Kanpur

03-11-2016

Outline

Problem Statement

"Given an image of a fistful of wheat grains spread evenly on a mono color cloth, identify the full grain, broken grain and foreign particles to give a quality estimate of the sample."

(a)

(b)

Dataset

- No existing grain dataset for this task, so created our own dataset
- Currently focusing on Wheat grain only.
- Collected 8 samples of different qualities of Wheat grain from Anaj Mandi, Kanpur.
- Grains of each sample were manually separated into full grain, broken grain and foreign particles categories, by the Mandi staff.

Dataset

- For each of the 8 samples, high resolution pics were clicked (13-16mp) by spreading the grains on a green background:
  - Full grain: 16 images (8 non-overlapping + 8 overlapping)
  - Impurities: 4 images for each kind
  - Broken Particles: 4 images
- In each set of 4 images, 1 is taken directly from above, and remaining are taken from random angles.
- The classification results are obtained on following task at first: Given an overhead image of partially touching particles, classify each particle as grain/impurity.

Dataset

Some sample images are given below:

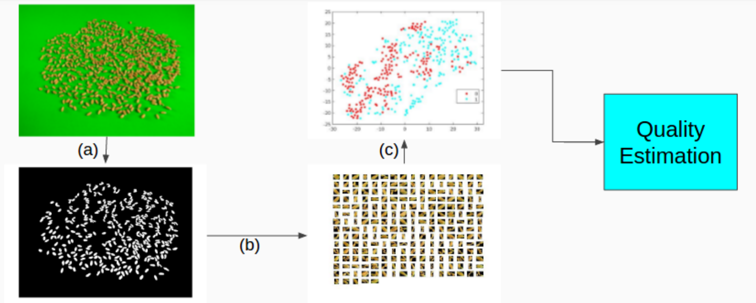
Figure: Fig: (a) is the image of the full grains and (b) is the image of impurities

(a)

(b)

Overview of Pipeline

Figure: Fig: Input image is first Pre-Processed(a) to give a binary image, which is then Segmented(b) to give different particles, which are then Classified(c) as grain/impurity.

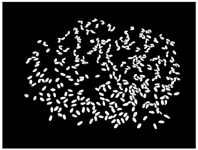


Pre-Processing

- Removal of shadows of grains: Due to green background, Red channel was free of shadows. Hence we used the Red channel.
- Noise removal and sharpening.
- Convert to binary based on a threshold.
- Use morphological opening to remove stray dots, and to open up slightly touching particles.

Figure: (a) is the input image and (b) is the binary image

(a)

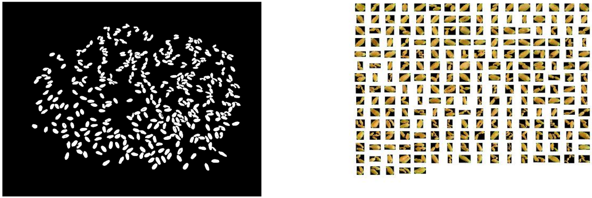


(b)

I Level Segmentation

- Find all the connected components in the binary image.
- Remove all components with pixel area less than a threshold.
- Each remaining component is a particle segment.

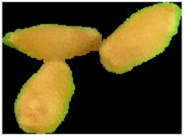
Figure: (a) is the binary image and (b) shows the obtained segments



(a) (b)

II Level Segmentation

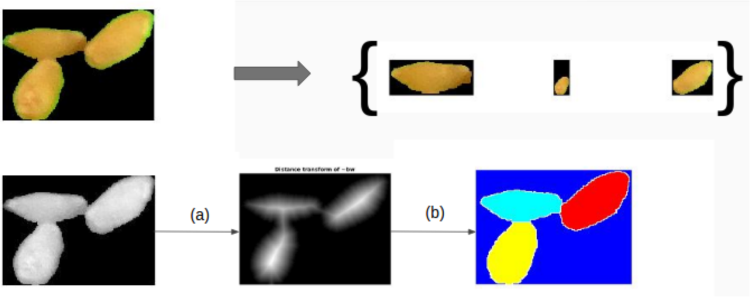
- As seen in image below, I Level Segmentation is not able to segment overlapping grains..



- To resolve this issue we propose a II level Segmentation:
  - Take the binary image of segment in question.
  - Obtain its distance transform.
  - Invert the DT and threshold to remove unwanted minimas.
  - Apply WaterShed Segmentation.

II Level Segmentation

Figure: Second Level Segmentation: Take the binary image corresponding to segment, take its Distance Transform(a), invert it then use watershed segmentation (b) after minima suppression.



Feature Extraction

Seven simple features were extracted for this task:

- **Color**: Mean [R,G,B] component of the segment (particle).
- **PCA components**: $[\lambda_1/mA, \lambda_2/mA]$, where $[\lambda_1, \lambda_2]$ the two eigenvalues for 2 component PCA on segment pixel coordinates, and mA is the mean segment area of image.
- **Eccentricity**: Ratio λ_1/λ_2
- **Size**: Area Ratio $=$ (pixel area of particle)$/mA$.

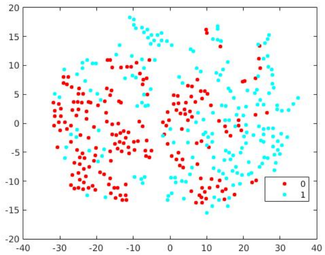
Training and Testing

- In our collection of segments of grain/impurities, no. of impurity particles is 197 while that of full grain particles is 4014.
- This will create problem during training our model.
- Solution: Randomly sample 197 full-grain points, and use 80% of these along with 80% of impurities to train the Classifier.
- Normalize the training data (mean zero and variance 1) before training the model.
- Test the Classifier on remaining 20% points AND on the 4014 superset of full-grain points.

Observations

- We plotted the 2-D TSNE projection of the training data and concluded that the distribution of two classes in feature space is Non-Linear:

Figure: TSNE plot of Training data, red for full grain and blue for impurities



- Hence, we have to use a Non-Linear Classifier.

Observations

Binary SVM, 'rbf' kernel:

Accuracy on test data	88.89%
Accuracy on 4014 full-grains	87.12%

	g	i
g	37	4
i	5	35

KNN Classifier, nNeighbors=5:

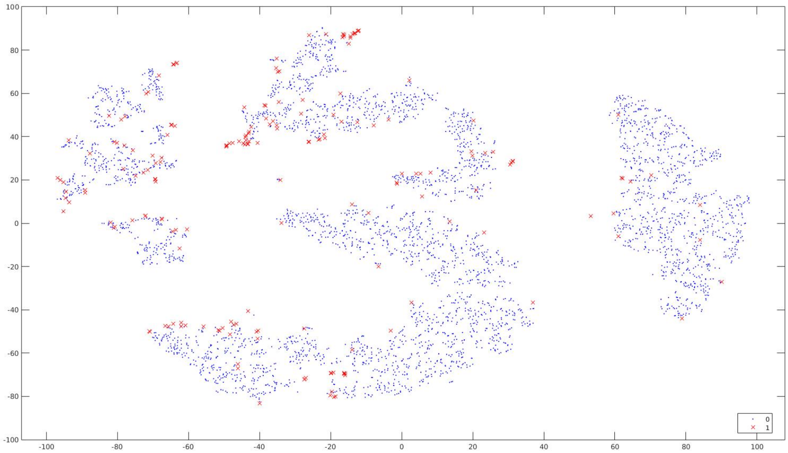
Accuracy on test data	86.42%
Accuracy on 4014 full-grains	90.76%

	g	i
g	39	2
i	9	31

(Here, g: Grain i: Impurity)

TSne Plot

Figure: TSne plot of the full 'non-overlapping' data



Challanges:

- Despite the accuracy, model still doesn't give satisfactory classification on a grain+impurities image.
- Wheat grains occur in clusters while impurities seem to be highly scattered.
- Impurities are less in number and highly varying, due to which Simple discriminitave model is not performing properly.
- Each part of the pipeline has to be separately tweaked.

Future Work

- We have to improve our feature representation and our Classifier.
- Improving contour formation during Segmentation to differentiate between pure/broken grain.
- We probably need decision tree based models where each features will come from statistics of training data.

References

[1] Marçal, André RS. "Alternative methods for counting overlapping grains in digital images." International Conference Image Analysis and Recognition. Springer Berlin Heidelberg, 2008.

[2] Harshwardhan kakkar , Jaspreet Kaur , Amandeep Singh, "DETECTION OF GOOD QUALITY WHEAT GRAINS USING IMAGE PROCESSING", Research Cell: An International Journal of Engineering Sciences , Vol17, Jan2016

[3] Github link, https://github.com/dhishku/MachineLearningforGrainAssaying

[4] The MathWorks Inc. MATLAB R2016A (9.0) and Image Processing Toolbox. Natick, Massachusetts, United States. Documentation available online: (http://in.mathworks.com/help/matlab/).

End of Presentation

Thank you.