# The New README

Lanqing Zhao

December 2019

## 1 Basic Functionality

This set of Matlab files and its results generated are intended to achieve two primary goals

- Read Images of test cell and gather the information necessary to characterize the printer
- Use the information gathered in the previous step to build a simulated printer

The code generates two database with some overlapping information to achieve these two objectives. One database generated by the first step is cell-based database, meaning that the information is stored based on each unique cell number;the second database generated by the second step is rearrange the order based on printer addressable points and only has information related to simulation building process.

### 1.1 The Name Convention of The Data Files

The data is stored as bmp files, and the name is as following:
$bb_c cc.bmp$
where bb is the block number from 01 to 24, and ccc is the cell number from 001 to 402.
(VERY IMPORTANT)
In this database, the set of nozzles is identified by a cell number, and an ink drop is identified by a cell number and a block number. The number of cell that has the same cell number is equal to the number of block.
When we refer an individual cell, we refer it as the cell ### from block ##.

### 1.2 The basic structure of Database

The database is in the form of matlab structure .Each structure of package represents a *cell*,and we have 402 cells in total with 24 blocks of them. Each cell with the same cell number captures the same set of printer addressable

point(the input pixel).
The fields of the package structure is as following:

- cell profile: this field includes: the cell profile such as mean and standard deviation of the individual cells across *all* blocks(in individual profile). It also contains information of displacement for each individual cell that has the same cell number(in individual profile).It also stores the displacement within cell as a pair with (hor_dis, vert_dis)Most importantly, it has the dot profile of all ink drops fired by the same individual printer addressable points.This field is generated by dot_profile.m and displacement3.m. .
- Basic Data: this one includes the basic data such as row and column centroid, filename.This one is generated by data_extract.m
- Horizontal/Vertical: This field stores information of horizontal and vertical displacement of two methods(Linear Regression and Grid Fit) . They include the basis or slope and intercept of each column or row. They also contain square error and displacement(L1). They are stored as one of the same nozzle. This part also has ideal (predicted) points in the cell.This field is generated by displacement3.m
- Horizontal/Vertical Estimate: This part stores histogram information (edge and normalized density) and original displacement data for each nozzles. These two fields are generated by create_dis.m . This field contains model object of KDE,GMM3,4,5,6,7,and 8.

### 1.3 Basic Structure of Nozzle_Bag database

This database inherits *only* the dot profile(mean and sd),horizontal estimation, and vertical estimation , which are relevant to simulated printer .

## 2 Files

### 2.1 preliminary

- binary_mask: find binary mask via otsu(in progress for separation of dots)
- sub_interpol2: perform subpixel interpolation

### 2.2 Data process:

- data_process: presumed main function of data process. This file packs the package struct of dot and cell database.It returns a total mean and sd profile and plots them as contour
- data_extract: This file packs BasicData of image data
- dot_profile: This file establish dot profile( mean and sd) of cells and individual nozzles

- displacement3: this function calculates fitted line of vertical and horizontal direction based on 2 methods and computes the displacement for each nozzle, and it also updates the cell profile to add information of displacement of each individual cell.
- create_dist: it creates the data of histogram for each nozzle within one cell number across all blocks
- add_id: add cell id for horizontal and vertical estimation and log likelihood for KDE method

## 2.3 Interface(v1 minor to our objectives)

:

- plot_fitline2:plot fit lines horizontally and vertically separately for linear regression methods, and plot the grid fit
  Some note: Horizontally fitted line is in vertical direction; vertically fitted line is horizontal direction
- plot_profile: plot dot profile( mean and SD) for cells , nozzles, or total cells
- nozzle_info: read and extract nozzle information of each nozzle including horizontal displacement and dot profile
- RMSE_table: create the rmse table for all 4 methods with format: inital of each method and inital letter v for vertical h for horizontal
- read_displacement:read the displacement of each cell, nozzle, and row , and store them as .csv. The data is stored in three folders:cell_displacement, nozzle_displacement, and row_displacement filename: gf/lg-hd/vd: for linear regression/grid fit horizontal or vertical ;lg/gf-cellnumber-nozzle/row number
- plot_centroid: plot the centroid of each cell
- read_gmm: this function reads the mean and bic of mixture gaussian models
- read_loglikelihood : this function reads the log likelihood of all model used
- read_mean_dis: this function reads mean of KDE.

Requirement to run the files : The user must have run data process files first.

## 2.4 The main function for data processing

In order for user to operate, there is a main the user can run, and it generates the database,read the displacement, and generate a table for all RMSE for 402*24 cells. It also reads mean value,log likelihood, bic, and add some id to or from the database

### 2.5 Simulator

The simulator part is our second part of objective. This contains file:

- simulator_builder:This function rearranges the nozzle order as what printer naturally has
- simulated_printer:This function resembles a simulated 7200 dpi printer with ideal,GMM3,4,5,6, and KDE. This function "prints" all five model with both linear regression and grid fit

This part has mini-main function.

## 3 Some simple explanation of the algorithm

### 3.1 Data Extraction

In data extraction, the main algorithm to generate binary mask is Otsu's method, and there are 3 steps to solve the separated dot issue

- regular Otsu method to generate binary mask, and then to check whether there are correct number of dots. If correct, then skip 2 and 3.
- If the number is over the supposed number, the secondary binary mask algorithm is invoked.This algorithm first filters out background pixel values and apparent satellite, and then it performs dilation to combine the separated dots in the presumed printer addressable points. Check again if the number of dots is correct, and if correct skip step 3
- If failed again, it will skip the dots with smaller area within the presumed printer addressable point. Then the number of dots will be checked again

If the number of dots is less than supposed number, we will have to retake the image because the dots may be missed, or two dots may be connected across adjacent rows or columns.

### 3.2 Displacement

Displacement uses linear regression and orthogonal transformation method.

### 3.3 PDF Estimation

In this part, the code has Gaussian Mixture Model with k=3,4,5,6,7,and 8 as well as KDE. The algorithm comes from built-in function from matlab

### 3.4 Simulated printer

What requires some notice in this part is that the printer generates dot displacement altogether and generates random muliplier for dot profile variants from uniform distribution from -1.5 to 1.5

# 4 Some results and time used

## 4.1 The Results and Data

After the data is processed, we have a data base. The results also include the displacement csv files , the plots of each cells (optional since the time to run is long), the histogram plot(optional, the time to run is even longer), and total dot profile contour.
The RMSE table is used to check if the data runs correctly, and in addition, the data also includes the cells that do not pass the primary binary mask and the cells that do not pass the secondary binary mask. If they fail the third one, the program will stop and print out its cell number and block number, the user has to take care of it. The running time is about 7800 seconds for first main function, and the data extraction takes 1500 to 1600 seconds to go over 402*24 cells. Other data including displacement ,mean of PDF, and bic for each model is also generated.
For simulator, the database of Nozzle_Bag is generated to perform simulated printer . Simulated printer generates images from linear regression and grid fit for GMM3,4,5,6, and KDE.It also prints ideal printer.

## 4.2 The sample

In order for user to try, I selected 48 samples of 12 cells from 4 blocks, and the user should uncomment the real code to try it.