Labelled Image Dataset Preparation for Rice Seed Germination Prediction and Variety Classifictaion using Low Cost Devices

S. Durai[a] and Dr. C. Mahesh[b]

[a]
Research Scholar & Assistant Professor, Department of Computer Science and Engineering, School of Computing, Vel Tech Ranagarajan Dr.Sagunthala R&D Institute of science and technology, Avadi, Chennai-60062,India.
[b]Associate Professor, Department of Information Technology, School of Computing, Vel Tech Ranagarajan Dr.Sagunthala R&D Institute of science and technology, Avadi, Chennai-60062,India.

Article History: Received: 10 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 20 April 2021

Abstract: Machine vision and Digital Image processing systems widely used in many of the applications like Quality control, Classification, Recognition, Identification, Security etc. For the implementation of machine vision system and machine learning, system must be feed by collection of images. Collecting the required images for the system are time consuming and complex task. We can collect the images from internet, research centres or existing data prepared by the researchers. Most of the systems images are not available, many of the researchers preparing the image datasets by themselves. Preparation of datasets needs costly equipments like high resolution cameras, highly configured computer systems, scanners etc. Nowadays handheld devices are equipped with high resolution cameras and cost also when compared to digital cameras very low. In this article we are going to demonstrate the step by step procedure of hardware setup, capturing image, processing images and feature extracted for those images by using mobile phone. For this implementation we have used Vivo z1pro mobile and prepared the data sets for the research of germination prediction and variety identification for the rice seeds. Preparing the datasets we have used four major rice seeds cultivated in Tamilnadu namely Andhraponni, Atchayaponni, KO50and IR20 which are collected from Agricultural university Trichirappalli,Tamilnadu,India. The prepared data set is freely available in https://github.com/duraitrichy/Riceseed.

Keywords: data set, rice seeds, germination, varietal classification, seed quality.

1. Introduction

In image processing, data set preparation is the important and core step. The accuracy of result always depends on the data sets used to train the machine. The data set [3]prepared by self contains the following important steps.

1. Hardware requirements
2. Samples
3. Hardware setup
4. Naming the images
5. Pre processing
6. Features Extraction
7. Decide on the conclusion attribute

The above steps are common for machine vision systems. Some hardware is common and must for digital image processing systems [Flat bed, camera, storage, lighting, camera holder etc] depends upon the cost and accuracy the hardware configuration may be varied. Samples are the data we are going to capture and process as an image. Samples must be carefully selected, choosing the samples following many factors, will discuss further in next chapters. Samples must be collected from any certified research centres or certified buyers. Samples must be mixed carefully for classification. Each sample must be named and noted for future reference.

Hardware setup[3,9] is the next step in the data set preparation, Digital Image Processing systems are working with features of the image , feature values are depends on the image captured, contains variety of features values like textural features, morphological features, colour features[4]. The factures may be vary depend upon variety of reasons like environment, lighting condition, camera pixel values, temperature etc. But when we are going to capture the image all the above said must be common and should be fixed in the hardware setup. The next challenge in the dataset preparation is distance between sample and capturing device. This also must be constant for the entire data set preparation.

For the identification of sample [10] whenever you are going to capture it write the same in the flat bed, for that keep separate space. Use short identifier for the same; don't keep the name as it. In proposed we explained how it was done. Before going to extract the features from the captured images, each image must be pre processed

like cropping the image in the constant dimension, image enhancement, removing unwanted objects, convert the colour image to gray scale and binary.

The next step is Feature extraction, how much attributes or features you are extracting for an image the accuracy of result will be increased. But don't extract unwanted features, include only important as well as which are giving enhanced accuracy. Based upon the type of sample you are going to process decide which type of features going to extract. Important features are morphological, colour, Textural features are common for all image processing, and apart from that some unique features also there for the type of samples you are going prepare dataset.

Once successfully decided what features are going to extract from the images taken, based upon the outcome result you are going to produce decide that is the conclusion attribute. For example classification variety name is the conclusion feature[3,4]. In the above all steps are done then decide on software or tool you are going to implement. For our project we have implemented, prepared the datasets for germination prediction and variety identification of rice seeds. In the next chapter we will explain in detail.

2. Materials used and Hardware Setup

The main objective of the project is to implement machine vision system in the field of agriculture. Because most of the work in agriculture is done manually and time consuming. Image processing techniques are used and implemented many of the fields. For our research work we are going to propose a automatic method to identification of germination prediction and varietal classification for rice seeds. For that we have proposed most economic methods to implement the same. In such a way, we have used hardware and techniques.

Capturing the image we have used VIVO Z1 Pro mobile with 16Mega Pixel camera, for storage and processing image HP Laptop with corei5 8th Gen processor, 4GB NVIDIA GEFORCE GTX , 1 TB HDD system used. The main objective of our research is to reduce the cost of data set preparation as well as easy to use. The mobile phone we have used gives descent images with good details. The flatbed scanner setup, lighting and camera setup will be explained further.

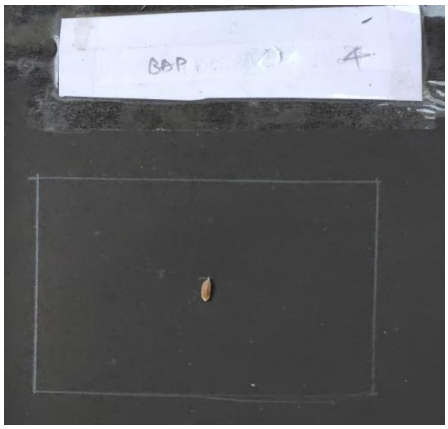

Fig 1.0 capture image of sample Bad Andhraponni

Fig 2.0 captured image of sample Good KO50

For the Flatbed scanner we have used black background, black colour sheets are available in shops with less cost. In the black sheet, we have drawn a square with 10cm side and in centre of the square mark a dot for place our sample. And in the top of the flat bed paste one white colour small sheet for naming the sample. Fixing the mobile we have used wooden materials, and distance between flatbed and camera distance is 30cm, which gives good detailed image. We have tested the same with different distances like 40cm, 20 cm and 50cm. But 30cm distance setup gives the results as expected.

In mobile capturing the image lot of options now comes with mobile, in the mobile what we have mentioned comes with photo, AI BEAUTY, NIGHT, PANO, PRO and DOC mode. We cannot use other than PRO mode, because mobile phones are having built in software to adjust the image quality, like auto contrast enhancement, brightness, colour change etc...But image processing always depends on those features only. For the same sample if you are taking multiple photos with other modes each and every image is giving different feature values. But for our system it must be constant, for that we can use only PRO mode. In PRO mode we can fix the capturing properties, once fixed the same values must be maintain for all sample capturing. In our preparation we have set EV:-1.7 and ISO:400. In the room lighting we have used Philips Stellar Bright 20-Watt Round LED Bulb and the room temperature is constantly maintained with 88degree.

3. Samples preparation

Preparing the datasets we have used four major rice seeds cultivated in Tamilnadu namely Andhraponni, Atchayaponni, KO50 and IR20 which are collected from Tamilnadu Agricultural university Trichirappalli,Tamilnadu,India. We have collected 100gm in each variety; all are certified seeds and 100% germinating viability. But for our research we need some non-germinating seeds also, so for each variety we have collected the non-germinating seeds also from the university. The Seeds are not germinated for the following reasons red rice seeds, physical damage, fungal attack and discoloration.

For each variety we have taken 20 samples , 10 are good and 10 are non germinating seeds. To name the seed the following short names are used for the entire dataset preparation shown in Table 1.0.

Variety Name	Identifier name (Good sample)	Identifier name (Bad sample)
Andhraponni	GAP	BAP
Atchayaponni	GATP	BATP
KO50	GK50	BK50
IR20	GIR20	BIR20

Table 1.0 Variety name and identifier

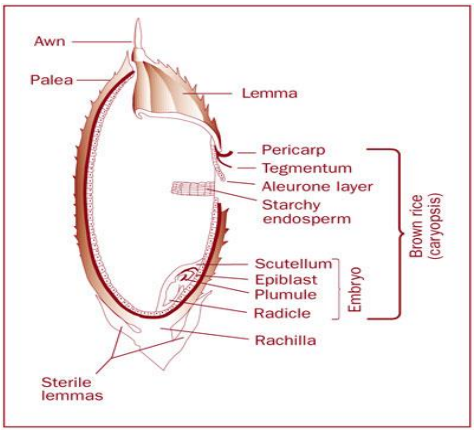

Fig 3.0 Parts of a rice seed

The above fig3.0 shows the parts of the rice seed[2]. In image processing normally images are captured for the same sample with different positions, for constant positioning and direction we have placed the each sample's awn part touched the dot which we have drawn on the flat bed. The sample name must be written on the white sheet which we pasted on the flat bed. Once sample placed on flat bed and the mobile position must be fixed we are not supposed to change the position. Samples are placed one by one and the names will be written on white sheet like GATP1,GATP2 for all samples.

4. Germination Test

Once sample captured immediately it must be kept inside a white colour cotton cloth for germination test. In the white cloth mention the name with permanent marker. In the germination test the sample will be submerged in water for 12-24 hours[1]. As per the suggestions from the experienced farmer, we have submerged the sample with cotton for 23 hours (from 6pm to next day 5.00pm). After 23hours we opened each sample cotton package, a small shoot appear on most of the samples.

For planting the samples we have used paper cups for each sample named with the same terms used to identify the samples. All the paper cups are filled with good soil for planting the seeds suggested by farmers. The same quality and quantity of soil is filled in all cups and the cups are now placed in sunlight. Every day morning the cups are filled with water. After 10 days the seeds are successfully germinated and the result of each sample is recorded, for the machine learning the germination status is the resultant attribute. The below figure shows the step by step method of germination verification samples.

a) White cotton cloth

b) GK50 Sample placed in cotton

c) BAP4 Sample placed in cotton

d) Sample are ready for submerge

e) Submerged samples

f) Planted seeds

g) Germination status after 5 days

h) germinated seeds after 10 days

Fig 4.0 Germination evaluation steps

Extracting Features

Extracting the features you can use any image processing tool, we have used matlab for extracting the features. Before processing the image the taken must be cropped. The crop will be done on the square we have drawn with same dimension for all images and save the file name as the identified term, but we are not suppose to change the colour properties. For our research work we have extracted totally 18 features shown in the below Table 2.0.

color Features	**morphological Features**	**textural Features**
Average Red color Average Green color Average Blue color	Major axis length Minor axis length orientation eccentricity area roundness aspect ratio	contrast(Horozontal&Vertical) Correlation(Horozontal&Vertical) Entropy(Horozontal&Vertical) Homogeneity(Horozontal&Vertical)

Table 2.0 Extracted Features

S.NO	VARIETY NAME	GERMINATION RESULT	RCOLOR	GCOLOR	BCOLOR	Area	Major Axis	Minor Axis	Eccentricity	Orientattion	Circularity	Aspect Ratio
1	BAP1	NOT GERMINATED	64.7758	64.621	64.3939	4150	128.0725	43.3743	0.9409	1.8311	0.5364	2.952728
2	BAP2	NOT GERMINATED	65.7698	65.5277	65.1802	4584	118.5197	51.6213	0.9002	-10.637	0.6294	2.295946
3	BAP3	GERMINATED	66.0183	65.9375	65.7054	3449	117.343	40.5705	0.9383	-4.704	0.4616	2.892323
4	BAP4	NOT GERMINATED	66.0139	65.9004	65.6484	4476	142.1091	44.059	0.9507	-5.8418	0.4211	3.225427
5	BAP5	GERMINATED	65.3832	65.2833	64.8464	4093	131.8286	41.798	0.9484	-9.986	0.338	3.153945
6	BAP6	GERMINATED	65.4816	65.3712	64.9581	5020	141.4251	46.1469	0.9453	2.74	0.389	3.064672
7	BAP7	NOT GERMINATED	65.7471	65.6659	65.2514	4732	136.2142	45.2146	0.9433	4.2724	0.4147	3.012615
8	BAP8	GERMINATED	66.2052	66.1103	65.6844	4145	125.9606	43.5817	0.9382	-13.8737	0.5729	2.890218
9	BAP9	NOT GERMINATED	65.4995	65.4122	65.1284	3949	126.4657	40.161	0.9482	-8.707	0.4533	3.148968
10	BAP10	NOT GERMINATED	66.5583	66.4476	66.0024	3685	121.4442	39.8651	0.9446	6.8553	0.4685	3.046379
11	BAP11	GERMINATED	67.6108	67.4687	67.0819	3434	111.7039	40.3037	0.9326	-5.5427	0.6447	2.771554
12	BAP12	NOT GERMINATED	66.5957	66.4796	66.2683	4310	125.3245	46.7415	0.9278	-0.8248	0.4719	2.681225
13	BAP13	NOT GERMINATED	67.1746	66.9606	66.6588	3929	133.5091	39.5692	0.9551	0.436	0.3665	3.374066
14	BAP14	NOT GERMINATED	67.5419	67.444	67.2633	3324	117.2887	37.735	0.9468	-5.568	0.395	3.10822
15	BAP15	NOT GERMINATED	68.2276	68.0514	67.7833	4638	126.5719	50.5866	0.9167	6.2308	0.3854	2.502084
16	BATP1	NOT GERMINATED	66.4123	66.2308	65.961	4680	149.4232	41.4585	0.9607	1.3633	0.4316	3.604163
17	BATP2	NOT GERMINATED	66.9279	65.8902	65.8989	3233	112.5702	37.0606	0.9443	-5.3458	0.5293	3.037463
18	BATP3	GERMINATED	65.9579	65.5383	65.1593	4064	142.2661	37.4727	0.9647	3.9024	0.501	3.796527
19	BATP4	NOT GERMINATED	66.2534	65.7915	65.5588	3985	136.7425	39.6106	0.9571	3.3697	0.2672	3.452169
20	BATP5	NOT GERMINATED	66.0688	65.4546	65.4118	3934	140.2834	36.2091	0.9661	11.4527	0.4796	3.874258

Table 3.0 Extracted Feature values

4. Conclusion

Digital Image processing systems are widely used in many applications. For the implementation of automated identification of quality and classification, the system must be trained with data sets. Data set collection is the complex, cost effective and time consuming part of the research. In recent handheld devices are equipped with good configured and giving results equal to digital cameras. Here we have proposed a method and implemented successfully , in data set preparation for rice seed germination prediction and variety classification. The system is implemented with low cost hardware and materials used. Preparing the datasets we have used four major rice seeds cultivated in Tamilnadu namely Andhraponni, Atchayaponni, KO50and IR20 which are collected from Agricultural university Trichirappalli,Tamilnadu,India. The datasets are successfully prepared and now it is implemented for the germination prediction system. By using this article researchers can prepared the dataset by themselves in their work in economic way. The prepared data set is freely available in https://github.com/duraitrichy/Riceseed.

References

1. http://naturalagreculturess.blogspot.com/2016/03/rice.html
2. http://www.knowledgebank.irri.org/
3. S.Durai, C.Mahesh, T.Sujithra, A.Suresh ," Survey of rice seed quality analysis for varietal purity
4. estimation by using image processing techniques", International Journal of Engineering &Technology, 7 (1.7) (2018) 34-37.
5. S. Durai, C. Mahesh," Research on Germination Prediction and Varietal Classification for Rice Seed Quality", International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-9 Issue-1, October 2019.
6. Archana Chaugule and Suresh N. Mali, ―Evaluation of Texture and Shape Features for Classification of Four Paddy Varieties,‖ Journal of Engineering, vol. 2014, Article ID 617263, 8 pages, 2014.
7. Lurstwut, Benjamaporn & Pornpanomchai, Chomtip. (2017). Image analysis based on color, shape and texture for rice seed (Oryza sativa L.) germination evaluation. Agriculture and Natural Resources 51(2017) 383-389.
8. Liu, Z. Y., Cheng, F., Ying, Y. B., & Rao, X. Q. (2005). Identification of rice seed varieties using neural network. Journal of Zhejiang University. Science. B, 6(11), 1095-100.

A. OuYang, R. Gao, Y. Liu, X. Sun, Y. Pan and Xiao-ling Dong, "An automatic method for identifying different variety of rice seeds using machine vision technology," 2010 Sixth International Conference on Natural Computation, Yantai, 2010, pp. 84-88.

9. Guillermo Botella, Carlos García & Uwe Meyer-Bäse , "Hardware implementation of machine vision systems: image and video processing", EURASIP Journal on Advances in Signal Processing volume 2013, Article number: 152 (2013)
10. FereshtehS.Bashiri a,b,n, EricLaRose b, PeggyPeissig b, Ahmad P.Tafti b,n,"MCIndoor20000:Afully-labeledimagedatasetto advanceindoorobjectsdetection",Data inBrief17(2018)71–75.