

PAPER • OPEN ACCESS

Automated egg grading system using computer vision: Investigation on weight measure versus shape parameters

To cite this article: Ahmad Fakhri Ab Nasir *et al* 2018 *IOP Conf. Ser.: Mater. Sci. Eng.* **342** 012003

View the [article online](#) for updates and enhancements.

Related content

- [Computer vision system for egg volume prediction using backpropagation neural network](#)
J Siswanto, M Y Hillman and M Widiarsi
- [Evaluating focused ion beam patterning for position-controlled nanowire growth using computer vision](#)
A B Mosberg, S Myklebost, D Ren *et al.*
- [Image Processing, Computer Vision, and Deep Learning: new approaches to the analysis and physics interpretation of LHC events](#)
A. Schwartzman, M. Kagan, L Mackey *et al.*



IOP | ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

Automated egg grading system using computer vision: Investigation on weight measure versus shape parameters

Ahmad Fakhri Ab Nasir, Siti Suhaila Sabarudin, Anwar P P Abdul Majeed and Ahmad Shahrizan Abdul Ghani

Innovative Manufacturing, Mechatronics and Sport Laboratory, Faculty of Manufacturing Engineering, Universiti Malaysia Pahang, Pekan, Malaysia

Email: afakhri@ump.edu.my

Abstract. Chicken egg is a source of food of high demand by humans. Human operators cannot work perfectly and continuously when conducting egg grading. Instead of an egg grading system using weight measure, an automatic system for egg grading using computer vision (using egg shape parameter) can be used to improve the productivity of egg grading. However, early hypothesis has indicated that more number of egg classes will change when using egg shape parameter compared with using weight measure. This paper presents the comparison of egg classification by the two above-mentioned methods. Firstly, 120 images of chicken eggs of various grades (A–D) produced in Malaysia are captured. Then, the egg images are processed using image pre-processing techniques, such as image cropping, smoothing and segmentation. Thereafter, eight egg shape features, including area, major axis length, minor axis length, volume, diameter and perimeter, are extracted. Lastly, feature selection (information gain ratio) and feature extraction (principal component analysis) are performed using *k*-nearest neighbour classifier in the classification process. Two methods, namely, supervised learning (using weight measure as graded by egg supplier) and unsupervised learning (using egg shape parameters as graded by ourselves), are conducted to execute the experiment. Clustering results reveal many changes in egg classes after performing shape-based grading. On average, the best recognition results using shape-based grading label is 94.16% while using weight-based label is 44.17%. As conclusion, automated egg grading system using computer vision is better by implementing shape-based features since it uses image meanwhile the weight parameter is more suitable by using weight grading system.

1. Introduction

The egg production industry has become a large industry owing to the demand and requirement by consumers, particularly in Malaysia. Egg is widely used worldwide and is healthy and safe for consumers. According to the report from the Ministry of Agriculture and Agra 2011 (Report 2011), the consumption of eggs per capita in Malaysia is 295 pieces per year [1]. The average production of eggs in the country is approximately 25 million per year [1]. The production rate increases to 37 million during festival seasons, such as Hari Raya, Chinese New Year and school holidays [1]. The high demand of eggs by consumers comes with high expectation for eggs with good quality. Many production companies are competing to produce eggs with extra nutrients and good shapes. Chicken



eggs are rated and graded into AA, A, B, C, D, E and F on the basis of their weight. The grade classification based on weight is shown in Table 1 [1].

Table 1. Grade classification based on weight.

Grade	Weight (g)
AA	More than 70
A	65–70
B	60–65
C	55–60
D	50–55
E	45–50
F	Below 40

The incremental percentage of egg demand in Malaysia has required improvement in egg grading systems. Classifying eggs on the basis of their grade becomes easier than conventional method, fast and efficient by implementing an automated system. Egg grading cannot be perfectly and continuously conducted by human operators because of eye fatigue and increasing error as a result of over wearing. An automated system for egg grading can be a solution.

Many methods using computer vision are introduced to accelerate the speed of egg grading systems. These methods can be categorised as grading using egg size or shape parameter. They differ from using egg weight in egg grading by egg suppliers. Researchers have continued performing egg grading using egg size parameter despite the said difference. As mentioned earlier, many countries, including Malaysia, use weight in egg grading. Therefore, the classification process based on weight measure versus shape parameter is conducted in detail in this study.

2. Related Works

The literature on egg grading shows that computer vision systems are used for egg dirt inspection, egg blood spot detection and egg crack detection [2-8]; however, these topics are beyond the scope of the current study. Two case studies in [9] and [10] are considered. All algorithm and experimental setup details are illustrated in the next section.

An egg grading system generally consists of two sections, namely, hardware and software designs. The hardware design includes the equipment and components needed to carry eggs and capture their images, such as a conveyor to carry the eggs, a camera to capture their images and a lamp for lighting in capturing the egg images. The software design involves the algorithms that process the egg images. The software design starts when the egg image is already captured by the camera. Then, the image will be sent into the pattern recognition system for processing, and the result will determine the egg grade.

2.1. Case Study 1

In this study, a grade classification and dirt inspection system for eggs is developed using a combination of image processing techniques. A few tools are used for the hardware system. An illumination box is built to enable light adjustment and light noise removal. Two bulb lamps are attached on top of the box. The dimension of the box is 10 cm × 10 cm × 10 cm. A webcam camera is installed to capture egg images, and the lens camera is attached 200 mm from the egg. Figure 1 shows the overview of the image capturing system.

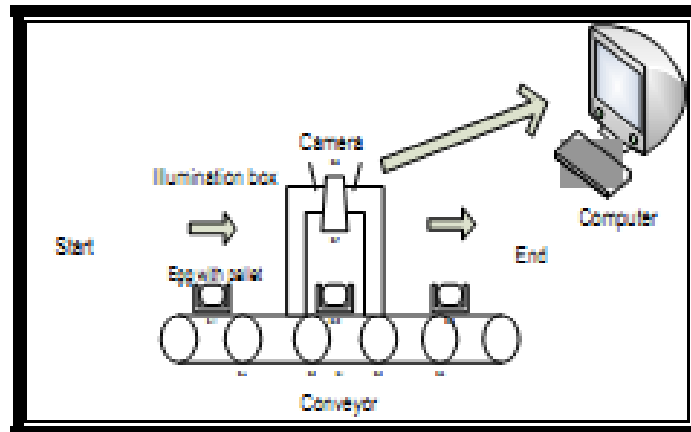


Figure 1. Overview of the system [9].

A database is developed to classify eggs. Pre-processing of images involves four steps: (i) converting colour images to greyscale; (ii) image thresholding processing (greyscale to binary); (iii) removing noises in binary images; and (iv) sharpening, smoothening and filling the region and hole of binary images. Two major features, area of white pixels and diameter of egg, are used in this system (Figure 2). A total of 60 eggs from 6 different grades (AA-E) are used in the experiment. The classifier for classification is not discussed in detail herein. The experimental results are illustrated in Figure 3. The system tests the cleanliness and dirtiness of eggs and executes classification and dirt inspection of eggs simultaneously.

Grade	Diameter (cm)	White pixel area
AA	≥ 15.0	≥ 79000
A	14.5 – 14.9	≥ 72000 - < 79000
B	14.0 – 14.4	≥ 66000 - < 72000
C	13.5 – 13.9	≥ 61000 - < 66000
D	13.0 – 13.4	≥ 56000 - < 61000
E	12.5 – 12.9	≥ 48000 - < 56000

Figure 2. Egg diameter vs. pixel area [9].

Grade	Classification (%)
AA	90%
A	90%
B	80%
C	90%
D	80%
E	80%

Figure 3. Classification results [9].

2.2. Case Study 2

An algorithm for indirect measurement of egg weight is designed using machine vision system and image analysis. The hardware setup uses a digital camera VIDO AU CC540HDN, light source system and work surface. The steps of the proposed algorithm are shown in Figure 4.

Seven features are used as geometric features of egg. Amongst them, four features, namely, Perimeter (P), Area (S), Major Axis Length (B) and Minor Axis Length (A), are calculated using the 'regionprops' function in MATLAB. Other features are executed on the basis of the previous basic features, as illustrated in the next paragraphs.

Egg volume (V_c) is calculated using the following equation:

$$V_c = \frac{2\pi}{3} AB \quad (1)$$

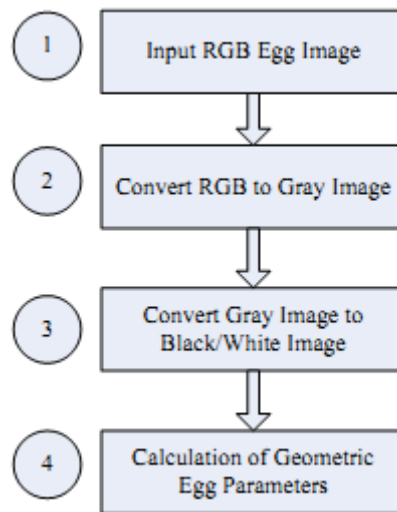


Figure 4. Block diagram of the method [10].

Two coefficients of egg shape (K_1 and K_2) are calculated using the following equation:

$$K_1 = \frac{P^2}{S} \text{ and } K_2 = \frac{A}{B} \quad (2)$$

In this study, coefficients (regression analysis) are used to measure the distance and difference of features. The results indicate that Perimeter, Major and Minor Axis Length and the two shape coefficients are significant features for indirect measurement of egg weight. Unfortunately, the features are not labelled on the basis of true weight classes. Classification, that is, assigning class labels, is essential to develop an effective automated egg grading system. After a feature extraction or selection procedure finds a proper representation, a classifier can be constructed from the labelled training data by using many possible approaches [11]. Then, this classifier is used to predict the class label of the test data. Classification is a type of supervised learning algorithm.

Supervised learning is a process of assigning a function to some desired categories as determined from the supervised training data. Here, the training data consist of a set of training examples, in which each set consists of a pair composed of an input object and a desired output value. A supervised learning algorithm learns from this training pair relationship and produces an inferred function. Specifically, in supervised learning, a teacher provides a category label or cost for each pattern in the training set that is used as a classifier. In this case, a 'true' category label is not inserted in examining the relationship between features. Therefore, significant features are only concluded on the basis of their weight without identifying the desired egg class. Hence, a complete egg grading system using statistical pattern recognition approach is developed to investigate features.

3. System Development

Firstly, digital images are taken as preliminary data. Then, image processing and egg feature measurement are conducted. Lastly, classification process is carried out.

3.1. Image Acquisition

A digital camera (a FUJIFILM camera of 8.1 MP) is used to capture images of graded eggs. An illumination box with a dimension of 100 cm x 100 cm is developed for egg placement to enable light adjustment and light noise removal. Two bulb lamps with yellow light are attached on top of the box. The distance between the lens and egg is approximately 200 mm. Figure 5 shows the overview of the hardware setup.



Figure 5. Image acquisition environment and setup.



Figure 6. Original image.

A total of 120 egg images that consisted of 30 eggs from each of four grades (A-D) are captured. The sample of image is shown in Figure 6.

3.2. Image Processing Tasks

Before geometric features can be measured, the images should be processed first to obtain the final image containing only [1, 0] binary values. Firstly, original *RGB* images are cropped to remove unwanted background. This procedure is essential because the image should show only the centre and a small amount of the total area (Figure 6). Accordingly, the amount of time executions for the next tasks can be reduced. This process finds the centre of image pixel firstly and then creates a rectangle box that covers the area of egg. Lastly, the box rectangle size is cropped, and the targeted image is saved in the database.

Secondly, the cropped image is converted to greyscale image. Thresholding process is executed on greyscale image by using manually setup level. In this case, the ideal level of 150 is used to convert to binary image. The smoothing of binary images involves two stages. The hole in the image is filled firstly and then the image is smoothed using the 'bwareaopen' function. The function checks the connection of pixels to its neighbourhood with low number of pixels. By implementing this function, some noises outside the egg can be removed given that the connected pixels present extremely lower number of pixels compared with those of the targeted image. The result of all image processing tasks is shown in Figure 7.

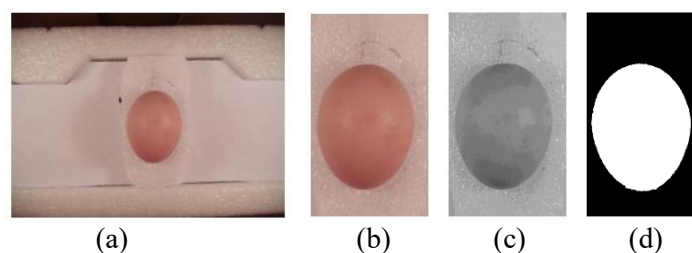


Figure 7. Processed egg image, (a) original image, (b) cropped image, (c) greyscale image, (d) final binary image.

Seven features are used as in Case Study 2. These features are Area (S) denoted as {f1}, Major Axis Length (B) denoted as {f2}, Minor Axis Length (A) denoted as {f3}, Volume (V_c) denoted as {f4}, Perimeter (P) denoted as {f6} and coefficients of egg shape (K_1 denoted as {f7} and K_2 denoted as {f8}). One feature, which is Diameter (D) denoted as {f5}, is added. Diameter is a scalar that

specifies the diameter of a circle with the same area as a region [11]. Diameter is computed as $4A/\pi$, where A is the egg area.

3.3. Classification Tasks

Two feature representation models, namely, principal component analysis (PCA) and information gain ratio (IGR), are considered.

PCA is a statistical method for reducing the dimension of variables. PCA is useful when implemented on a large number of variables with some redundancy. In this case, redundancy indicates that some of the variables are correlated with one another. The general form for the formula to compute scores on the first component extracted in PCA is expressed in the following equation:

$$C_1 = b_{11}(X_1) + b_{12}(X_2) + \dots + b_{1p}(X_p) \quad (3)$$

where C_1 is the subject score on principal component (PC) (the first component extracted); b_{1p} is the regression coefficient (or weight) for observed variable p , as used in creating first PC; and X_p is the subject score on observed variable p .

IGR is usually used to determine the capability of each single feature to separate the given dataset. Entropy is commonly used in the information theory measure and characterises the purity of an arbitrary collection of examples. If A is an attribute (feature) and C is the class, then the following equations provide the entropy of the class before and after observing the attribute:

$$H(C) = -\sum_{c \in C} p(c) \log_2 p(c) \quad (4)$$

$$H(C|A) = -\sum_{a \in A} p(a) \sum_{c \in C} p(c|a) \log_2 p(c|a) \quad (5)$$

The amount by which the entropy of the class decreases reflects the additional information about the class provided by the attribute and is called information gain. Each attribute A_i is assigned a score on the basis of the information gain between itself and the class, as follows:

$$IG_i = H(A_i) + H(C) - H(A_i, C) \quad (6)$$

k -nearest neighbour (k -NN) is used for classification purposes. k -NN is a method for classifying objects on the basis of closest training examples in the feature space. k -NN is a type of example-based learning or lazy learning where the function is only approximated locally and all the computation is deferred until classification. This algorithm is a simple machine learning algorithm in which an object is classified using a majority vote of its neighbours. The object is then assigned to the class, which is the most common amongst its k -NN. The neighbours are taken from a set of objects for which the correct label is known. The learning in this model is based on storing all the training instances that correspond to points in an n -dimensional space along with their class labels, and classification is delayed until a new instance arrives.

4. Experimental Results

A total of 120 egg images from four classes are used in the experiments with 10-fold cross validation for classifier evaluation. The performance evaluation is examined by the recognition accuracy through the presented formula:

$$RC = \frac{M_C}{M_T} \quad (7)$$

where M_C presents the number of correct classified eggs in each class, and M_T is the number of all eggs in all classes. The experiments are conducted on four different ways:

- (i) The egg features {f1:f8} are used as starting features to find the best single feature using IGR theory.
- (ii) The sorted features are {f5 = 0.655, f4 = 0.643, f3 = 0.571, f2 = 0.531, f7 = 0, f6 = 0, f8 = 0}, from the top to the bottom ranking features. Two starting features of all features (original) and

two of the best features of > 0.5 information gain value {f5, f4, f3, f2} are used for classification process.

- (iii) The feature extraction using PCA is implemented on seven features {f1:f8} for comparison with other feature representation techniques. Three PCs are retained and selected as starting features. The three PCs cover 95% of total variance for the seven features.
- (iv) The classification process uses five different distance functions in linear k -NN: (a) Euclidean distance, (b) Manhattan distance, (c) Minkowski distance, (d) Filtered distance and (e) Chebychev distance. One nearest neighbour 1-NN is set for the entire experiments.

Table 2 shows the average recognition accuracy using weight-based grading as a referred class (egg grading). As shown in Table 2, the recognition accuracy for PCA is better than using the combination of all original features {f1:f8} and selected several best single features {f5, f4, f3, f2} by at least 2%–4%. Comparing the distance function methods in k -NN shows some result differences among them, thereby increasing the recognition accuracy. However, the accuracy result for this dataset only achieves a maximum of 51.67%. As predicted at the beginning of this study, the recognition accuracy should be low because the classification process is conducted by referring to supplier's egg grading labelling, that is, using weight measure. Therefore, a new egg class labelling based on shape measure is obtained through simple k -means clustering for solving the above-mentioned problem.

Table 2. Average recognition accuracy (weight-based grading).

No.	Method	Starting Features	Recognition Accuracy (%)					Average (%)
			(a)	(b)	(c)	(d)	(e)	
(ii)	Original features	{f1:f8}	41.67	39.17	41.67	40.83	45.83	41.83
	IGR	{f5, f4, f3, f2}	41.67	37.50	41.67	38.33	43.33	40.50
(iii)	PCA	{f1:f8}	42.50	40.00	42.50	51.67	44.17	44.17

Table 3 shows that the first 30 images in the database, which are supposed to be in grade A in accordance with weight measure, are changed to multiple classes using shape measure after performing simple k -means clustering. Five eggs are clustered to class 0, two eggs to class 1, three eggs to class 3 and others (majority) are classified as class 2. Many class changes also occur for other egg grades in the database. The classification experiments are repeated using a new cluster. Table 4 shows the recognition accuracy using shape-based grading. The classification results are increased dramatically compared with the previous results of weight-based grading. The best recognition results are achieved consistently using original features compared with using PCA and single feature set. PCA can fully improve the low recognition results for this type of dataset (Table 2). When PCA is implemented on good recognition results (Table 4), the results are insignificantly improved.

For both datasets, using the best single features (based on IGR theory) also does not help improve the results. The reason is that the single feature alone is filtered without considering the combination of feature subsets. In the future, wrapper feature selection approach can to be used provide improved combination of feature subsets.

Table 3. New cluster assigned for egg grade A.

Image No.	New Cluster Assigned	Image No.	New Cluster Assigned	Image No.	New Cluster Assigned
1	0	11	2	21	2
2	0	12	2	22	2
3	2	13	3	23	2
4	2	14	2	24	0
5	2	15	2	25	2
6	2	16	2	26	3
7	2	17	2	27	3
8	2	18	2	28	0
9	2	19	2	29	0
10	2	20	1	30	1

Table 4. Average recognition accuracy (shape-based grading).

No.	Method	Starting Features	Recognition Accuracy (%)					Average (%)
			(a)	(b)	(c)	(d)	(e)	
(ii)	Original features	{f1:f8}	94.17	95.00	94.17	93.33	94.17	94.16
	IGR	{f4, f3, f2}	82.50	83.33	82.50	83.33	77.50	81.83
(iii)	PCA	{f1:f8}	91.67	92.50	91.67	96.00	90.83	92.53

5. Conclusions

In this study, an automated egg grading system using computer vision is developed. Two methods using weight and shape measure are comprehensively investigated. The experimental results show that the recognition results are better when the classification process is conducted on the basis of a new cluster (shape-cluster). By contrast, the recognition results of weight-cluster are extremely low since the shape measure is not related to weight measure (class). Previous researchers have supported the strong relation between the two. However, in this study, the relationship is poor as the classification results are extremely low by using weight-based grading. Whether the egg class will change using shape-based grading is unclear. If the answer is yes, then this system can be implemented to help accelerate the speed of grading. Otherwise, further studies are needed to find the best features (based on egg image) that are nearly perfectly correlated with weight measure.

Acknowledgments

The authors are grateful to the Universiti Malaysia Pahang for providing the funding under internal research grant [RDU1703159] to support this research.

References

- [1] Anim M 2013 Telur ayam gred dan harga <http://animhosnan.blogspot.my/2013/07/telur-ayam-gred-dan-harga.html>
- [2] Patel V C, McClendon R W and Goodrum J W 1998 Color Computer Vision and Artificial

- Neural Networks for the Detection of Defects in Poultry Eggs *Artificial Intelligence Review* **12** pp 163–176
- [3] Mertens K, De Ketelaere B, Kamers B, Bamelis F R, Kemps B J, Verhoelst E M, De Baerdemaeker J G and Decuypere E M 2005 Dirt Detection on Brown Eggs by Means of Color Computer Vision *Poultry Science* **84** pp 1653–1659
- [4] Dehrouyeh M H, Omid M, Ahmadi H, Mohtasebi S S and Jamzad M 2010 Grading and Quality Inspection of Defected Eggs Using Machine Vision *International Journal of Advance Science and Technology* **16** pp 43–50
- [5] Loredana Lunadei, Luis Ruiz-Garcia, Luigi Bodria and Riccardo Guidetti 2011 Automatic Identification of Defects on Eggshell Through a Multispectral Vision System *Food Bioprocess Technology* **5** pp 3042–3050
- [6] Yoon S C, Lawrence K C, Jones D R and Heitschmidt G W 2012 Improved modified pressure imaging and software for egg micro-crack detection and egg quality grading *Applied Engineering in Agriculture* **28** pp 283–293
- [7] Arivazhagan S, Newlin Shebiah R, Hariharan Sudharsan, Rajesh Kannan R and Ramesh R 2013 External and Internal Defect Detection of Egg using Machine Vision *Journal of Emerging Trends in Computing and Information Sciences* **4** pp 257–262
- [8] Mahmoud Omid, Mahmoud Soltani, Mohammad Hadi Dehrouyeh, Seyed Saeid Mohtasebi and Hojat Ahmadi 2013 An expert egg grading system based on machine vision and artificial intelligence techniques *Journal of Food Engineering* **118** pp 70–77
- [9] Ibrahim R, Mohd Zin Z, Nadzri N, Shamsudin M Z and Zainudin M Z 2012 Egg's Grade Classification and Dirt Inspection Using Image Processing Techniques *Proceedings of the World Congress on Engineering* vol 11 (London: Newswood Limited) pp 1–3
- [10] Alikhanov D, Penchev S, Georgieva Ts, Moldajanov A, Shynybaj Z and Daskalov P 2015 Indirect Method for Egg Weight Measurement Using Image Processing *International Journal of Emerging Technology and Advanced Engineering* **5** pp 30–34
- [11] A Fakhri A Nasir, M Nordin A Rahman, Nashriyah Mat, A Rasid Mamat and Ahmad Shahrizan Abdul Ghani 2017 Ficusdeltoidea (Jack) Moraceae Varietal Identification Using Statistical Recognition Approach *World Applied Sciences Journal* **35** pp 82–88