

Research Article

Online Price Recommendation System for Shopping Used Cell Phones

Amira M. Idrees and Shereen A. Taie

Faculty of Computers and Information, Fayoum University, Fayoum, Egypt

Abstract: This study introduces a model for supporting the consumer's cell phone purchase decisions by determining the phone status and recommending an estimated price for the phone. According to economic reasons, many people may have to buy used products. As the price for used cell phones is usually estimated while trading between the seller and the buyer which do not guarantee the best and fair price, this situation has emerged the idea of this study. We consider the idea of the proposed model is novel as up to our knowledge, all proposed online systems do not guarantee a fair price, they only focus on providing a link between the buyer and the seller and provide the seller's required price with no recommendation of whether this price is fair or not. The proposed model has a set of defined criteria for status estimation. The values for some of these criteria are detected by applying of digital image processing technique on the phone photos and then a classification model is applied to determine its category and its estimated price. The recommendation system has been built based on the proposed model. Moreover, a web application with a user friendly interface has been developed and a number of cases for Samsung cell phones have been conducted which proved the applicability of the model.

Keywords: Classification algorithms, e-commerce, image processing, online shopping, supervised learning

INTRODUCTION

The development of technology has growing rapidly in the last decade with the impact on the use of widgets in the daily activities. As sophisticated technology is utilized, human needs are also increasingly changing. One of the current popular human activities is online shopping which has revolutionized the way people shop (A Nielsen Report, 2010). According to economic reasons, many people may have to buy used products. However, when two individual persons need to perform a trade operation, they may not be able to detect many important characteristics of the phone to have fair price estimation. As used cell phones status usually differ according to different factors including the time used and the frequency of usage. Therefore, there is a need for a price recommendation method for used cell phones. The proposed model has proposed a solution for these situations as will be discussed in the following sections.

In many online applications extracting text from images or videos is an important problem (Jain and Yu, 1998a) Therefore, analyzing an image or a video to search for a desired text contents improves the accuracy of the system (Jain and Yu, 1998b). In addition, a larger number of approaches are introduced in several researches with high level of success, such as region

based, edge based, morphological based and texture based methods (Nayak *et al.*, 2015).

Online shopping is a growing area of technology that markets the products via a store on the Internet (Lee *et al.*, 2001). Online shopping have many impacts on sales in emerging markets embracing emerging trends and technologies helps marketers create a sustainable competitive advantage for their business (Barbuceanu and Lo, 2000). Online shopping has become more sophisticated as an effective way to create a relationship with the consumer that has depth and relevance (Kim and Ammeter, 2008). The online shopping environment allows the implementation of very high degrees of interactivity (Chevalier and Mayzlin, 2006). That characterized by the reciprocity in the exchange of information, availability of information on demand, response contingency, customization of content and real-time feedback (Pan and Zhang, 2011). Establishing a store on the Internet, allows for retailers to expand their market and reach out to consumers who may not otherwise visit the physical store (Degeratu *et al.*, 2000). The convenience of online shopping is the main attraction for the consumers (Klenow and Kryvtsov, 2008).

This study aims to introduce a recommendation system based on a classification model for used cell phones. The proposed system classifies the phone according to determined criteria and presents a

recommended price to the system user. The proposed system detects some of these criteria by processing the images of the phone and then applies a classification method for the estimation procedure. A web application has been built with a user friendly interface to apply the case study.

LITERATURE REVIEW

Different researches have been proposed in the image processing field that focused on detecting text from an image source of data. Among these researches, two different approaches have been proposed, "Connected Component Based approach" (Jung *et al.*, 2004), while "Contour-based robust algorithm" (Sushma and Padmaja, 2009) is proposed for detecting the text from a colored image, in addition to applying "binarization" before the recognition step. As discussed in Jung *et al.* (2004), optical character recognition "OCR" includes six phases; they are "detection phase, localization, tracking, extraction, enhancement and recognition". Moreover, many systems for detecting text automatically from an image are proposed as mentioned in Fabrizio *et al.* (2009), some of these systems are "optical character reading, automatic image indexing and visual impaired people assistance". These systems are based on one of the two approaches that are usually applied for image matching, they are "direct based" (Lucas and Kanade, 1981; Bergen *et al.*, 1992) and "feature based" (Förstner, 1986; Harris, 1992).

Direct based approach is based on the values of the pixels to iteratively align images, while the feature based approach is based on determined features with local information. Moreover, there are invariant features that use large amounts of local image data around salient features to form invariant descriptors for indexing and matching. Also, in a wide range of problems, invariant feature-based approaches to matching have been successfully applied, including object recognition (Lowe, 2004), structure from motion (Schaffalitzky and Zisserman, 2002) and a panoramic image stitching (Brown and Lowe, 2003). In this study we use the application of optical character reading for recognizing cell phone version from the seller uploaded image. Invariant feature-based approach was used in matching cell phone images.

One of the most significant fields that provide a successful communication among different criteria is data mining. One of the communication that data mining provide is to classify each instance of the data to its most relevant class and to assign the instance as a member in the selected class; this task is called classification data mining. As different machine learning algorithms have been presented for classification task, we have concentrated on two of those algorithms; they are K-Nearest Neighbor algorithm (Tan *et al.*, 2006; Wu *et al.*, 2008) and

Random Forest algorithm (Breiman, 2001; Fernández-Delgado *et al.*, 2014).

KNN is an algorithm that is known to be a simple algorithm, it simply detects a k number of instances in the training data, then it applies a similarity measure to find the most relevant training instances to the testing instance under examination, then it assigns the testing instance to his nearest neighbors of training instances who are a subset of the k group (Tan *et al.*, 2006). While working with KNN, there are some major factors that are discussed previously in different research, such as in Wu *et al.* (2008). These factors are training data set size as it is known that KNN's result gets more accurate as the training data gets larger, the metric that is used for measuring similarity between instances as it differs according to the type of data, the value of k instance set and the number of elements in the nearest neighbor set (Wu *et al.*, 2008).

The value of K is a critical factor as if K was too small, then noise data can affect the result, while if it is too large, then the classification may incur an incorrect result due to the incorrect neighboring instances. Also when KNN selects the class of the testing instance, it considers all attributes with the same weight which may result again to inaccurate decision. However, we have solved this issue by assigning points to each attribute according to its importance. Moreover, the selection of the similarity measure is very critical as it depends on the nature of the data. For example, for classifying documents, cosine similarity measure (Huang, 2008) is one of the best choices. While for small set of distinct attributes such as in our case, Euclidean distance measure (Advanced Projects R&D, 2005) should be selected as it is previously discussed in different research that Euclidean distance discrimination becomes poor for large number of attributes.

As discussed in Fernández-Delgado *et al.* (2014), RF is one of the best classification algorithms. RF consists of a set of trees predictors that build a set of decision trees which are interconnected to each other. Predictors depend on a vector that is randomly detected and distributed to all trees and then a vote for the most relevant class is applied to the instance on test (Breiman, 2001). RF is known to be robust to outliers, high scalable algorithm, can deal with large dataset and able to efficiently estimate missing data. In this study, the two classification algorithms are applied and the results are verified as will be discussed in section III.

PROPOSED MODEL

In this section we outline the proposed system that is presented for determining the status and recommending the price for used cell phones. The proposed model consists of three main phases. Figure 1 represents an overview of the proposed system's

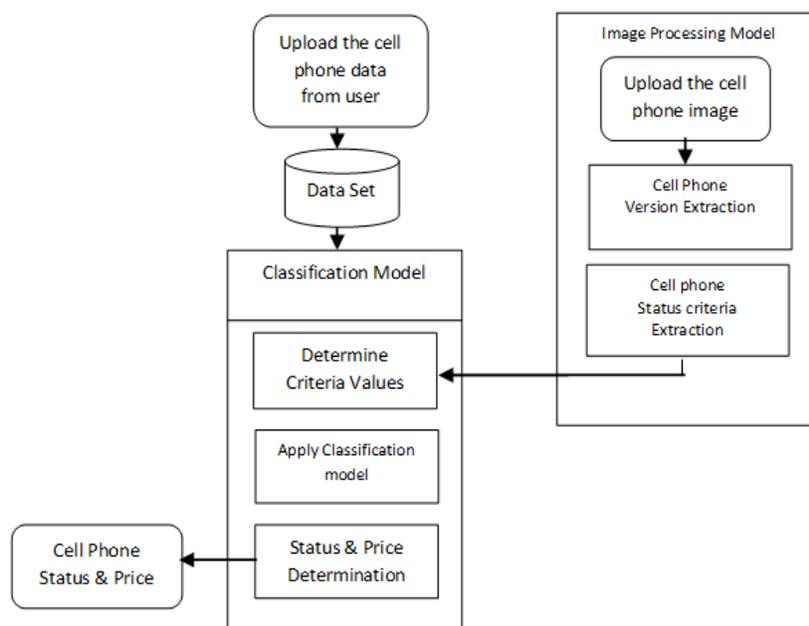


Fig. 1: Proposed system architecture

Table 1: The discrimination attributes ranges and values for each level

Attribute	Attribute Range	Attribute points	Class A	Class B	Class C
Time used	From 1 to 24 month	10 points	0 to 6	6 to 18	above 18
Congruence percentage	Percentage	10 Points	Less than 10%	From 10 % to 40%	Above 40%
Battery status	Hjgh, Medium, low	1 Point	high	High medium	N/A
Charger Exist	Yes, No	0.5 Point	Yes	N/A	No
Head Phone exist	Yes, No	0.4 Point	Yes	N/A	No
Guarantee exist	Yes, No	0.3 Point	N/A	No	No
Changed parts	From 0 to 2	1 Point	0	1	2 or more

architecture, while the following subsections illustrate the system phases in detail.

The set of criteria for cell phone classification: Due to the person’s usage of cell phones, its state normally changes from the original state, for example, its appearance may change either on the body of the phone or on the screen due to the continuous usage without protectors. Moreover, the time of usage and the usage frequency may affect the battery status, which is one of the main factors that determine the status of the phone and consequently determine the fair price.

We have conducted several meetings with the sellers’ experts who are working in the cell phone fields and tend to buy and sell the cell phones. The aim of these meetings was to gather the most important attributes that the sellers and the buyers become interested in while trading about a used cell phone. We have collected seven of these attributes for discriminating among the cell phones with three levels, which determine the usage level of the phone. The three levels are named A, B and C, where A is the highest class and C is the lowest. Table 1 demonstrates the collected attributes their ranges and the value range of each level. In Table 1, each attribute has been assigned

with points according to its importance in the experts’ point of view. For example, Congruence Percentage and the time used are attributes of highest importance while the charger existence and the headphone existence are of lower importance.

Although in our meetings, we have collected more than seven attributes, however, we have selected these seven attributes as they match with A Samsung cell phone which is the type that this research focuses on. However, more attributes have been collected but it is not applied as it do not comply with Samsung mobiles such as missing buttons as Samsung versions are usually touch screen phones. Therefore, we claim that the proposed method is able to be applied to other cell phone types with minor changes as will be discussed in the discussion part.

The value of these attributes for a cell phone that is desired to be sold is collected by one of two methods. The first method is processing the image of the cell phone; this method detects the phone version and phone’s Congruence Percentage as will be discussed later. The second method is used for the remaining of the attributes by providing a number of questions to the system user with possible answers. The second method uses a user friendly interface with some questions that

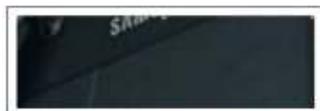


Fig. 2: Cell phone of class A



Fig. 3: Cell phone of class B



Fig. 4: Cell phone of class C

should be answered by the user and then the user presses submit. Some examples of these questions will be demonstrated in the experiment section.

The following are the questions that appear to the user for collecting the values of the required attributes:

- Enter the date of buying the phone?
- How regular is your battery charging? (once a day, Once each 12 hours, Once each 6 hours, Once each 3 hours, Once each one hour)
- How can you describe the usage level of your mobile? (Low, Medium, High)
- Is the Charger exist? (Yes, No)
- Is the Headphone exist? (Yes, No)
- Is your mobile still in the guarantee? (Yes, No)

According to these questions, some attributes directly have its values from the answer of the question such as the existence of the headphone from question number 5. However, other attributes need more calculations such as the battery status as its value is estimated from the answer of questions 2 and 3. For example the battery status is considered low if it is charged every hour although its usage is low.

Using image processing to detect phone version and status:

In this phase, we use digital image processing techniques to detect the cell phone version and evaluate the product status. This phase composed of two main steps; the first depending on using Optical Character Recognition (OCR) to detect the Samsung cell phone version. The second goes through matching the uploaded Samsung cell phone image with the standard image of the detected cell phone version, to detect the difference between two images and evaluate the cell phone status.

Detect the samsung cell phone version: Instead of manually typing the cell phone version, the proposed

system extracts the information by uploading the Samsung cell phone image while starting up to be processed and detect the cell phone version as the result to reduce fraud and mismatch. To identify and recognize the text on the cell phone image we perform a sequence of steps. At first scaling and gray scaling are applied. Then, binarization is done using a simple, efficient binarization technique (Bernsen, 1986).

Next, the black and white image is processed in the text area extraction algorithm. Detecting the appearances of text in the image by separating the area into two parts, text area and background area (Huang *et al.*, 2005). Hence, every text area is divided into smaller images for each holed character. The segmentation on text is done horizontally and vertically to dividing lines and characters respectively using histograms. Then, the text in the cell phone image is now more amenable to be recognized by the automatic optical character recognition using Google's open source OCR-Tesseract (Tesseract-ocr, 1996).

Detecting the cell phone status: The second step goes through matching the uploaded closed cell phone image with the standard image of the detected cell phone version, to detect the difference between two images and evaluate the cell phone status. In this step the cell phones defects (screen scratch) are recognized if exist to detect the cell phone status. Canny filter is used to detect edges, it is the basic algorithm for edge detection (Ding and Goshtasby, 2001). Hough Transform (HT) (Walsh and Raftery, 2002) are used to detect the details and returns the direction and the position of the scratch line, with Hough peaks 1000. Finally, we calculate the congruence rate between the used cell phone images with the standard image to classify the cell phone status. Congruence rate is the ratio of the number of coincides pixels in the used cell phone image with the corresponding pixels in the standard original image to the total number of pixels in the original image; that is, $\text{Congruence rate} = \frac{\text{Num coincides pixels}}{\text{Total Num of pixels}}$. We classify the phone status with three levels which determine the usage level of the phone. The three levels are named A, B and C, where A is the highest class and C is the lowest.

The values of Congruence rate that classified the used cell phone status from A to C classes are demonstrated as follows: $A \geq 85$, $60 \leq B < 85$ and $C < 60\%$.

Thus, the result is one of the most important parameters that is used for estimating the cell phone price as will be shown later. Figure 2 and 3 illustrate the cell phone status classes on the three different examples.

Applying classification model component (estimate phone price):

The aim of this component is to classify a subjected cell phone to one of the three classes; hence the estimated price for this phone is determined

according to this classification. In order to reach the targeted goal, we have performed several steps which are discussed in the following subsections.

Preparing the training and testing data set: We have created 483 cases that represent all possible states of the classes (A, B and C) (Fig. 4). This step is performed to create the set of cases that are used in the classification algorithm as the training set. We have implemented a code fragment for this step using Java Code.

Then we have defined three clusters, each cluster represents one of the pre-determined classes and the description of each cluster is determined according to the class description using the attributes that are discussed previously.

Finally, we have created 30241 cases that represent all possible states of any used cell phone based on the previously discussed attributes (Fig. 5). This step is performed to create the set of cases that are used in the classification algorithm as the testing set. We have also implemented a code fragment for this step using Java Code.

Applying classification algorithms: The aim of this step is to determine the class of the user's phone using the attributes' values that are collected as discussed previously by the processed phone image and the user. For this target, we have applied the previously prepared data as training and testing set on two of classification algorithms (K-Nearest Neighbor (Tan *et al.*, 2006; Wu *et al.*, 2008) and Random Forest (Breiman, 2001; Fernández-Delgado *et al.*, 2014) and classified each testing case to a determine class (Fig. 6).

However, we have performed a validation procedure for the results to select one of the algorithms to follow. A demonstration of the results of validation for both of the algorithms output is discussed as follows which reveals the selected algorithm to follow.

Validation of classification algorithms: We have conducted several meetings with the experts of the field and performed an extensive review procedure on the results of the two algorithms. Then we have applied two validation measurements for their results, they are Precision and Recall, these measure are presented in Fig. 7. The review process revealed that the results of the Random Forest Algorithm have higher Precision and Recall (95.94, 93.27%, respectively) than the k-Nearest Neighbor algorithm (83.45%, 80.35%, respectively). Therefore, we have relied on the classification results of Random Forest Algorithm which is demonstrated in Fig. 6.

Estimate phone price: According to the experts in the field, Price for each class ranges from lower to upper price bound. Table 2 demonstrates the price deduction range for each class. For example, the phone that is

Table 2: Lower and upper price deduction

Class	Upper bound	Lower bound
Class A	Deduct 5% of the original Price	Deduct 10% of the original Price
Class B	Deduct 10% of the original Price	Deduct 30% of the original Price
Class C	Deduct 30% of the original Price	Deduct 50% of the original Price

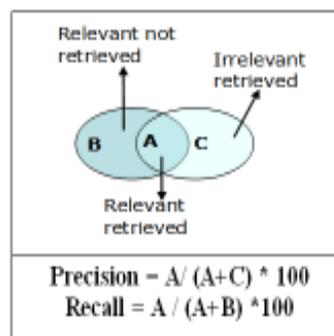


Fig. 8: Precision and recall measures

classified to class A, has a price deduction from 5% to 10 % according to its status, which means if the new phone of the same version costs 5000 LE in the market, this means that the used cell phones of class A has a price ranges from 4500 LE to 4750 LE according to its status.

In the proposed system, we have determined the estimated price of the phone according to its confidence in its related class according to equation 1. This confidence has been determined by the classification algorithm which was discussed previously.

$$E = LB + ((UB - LB) * X) \tag{1}$$

where,

E : The estimated price for the cell phone under examination.

LB : The lower Price bound for the class of the cell phone under examination

UB : The upper Price bound for the class of the cell phone under examination

X : The confidence of the cell phone under examination to belong to the determined class

For example, if the phone status has been classified to Class A with confidence 75%, this means that the estimated price of the phone will be:

$$\begin{aligned} \text{Estimated Price} &= 4500 + ((4750 - 4500) * 0.75) \\ &= 4687.5 \text{ LE} \end{aligned}$$

CASE STUDY

A web application has been built for applying the selected cases. Figure 8 present the about page of the system.

Figure 9 presents the page in which the user upload his phone image, then the system process this image



Fig. 9: About page of fairtrade.com



Fig. 10: About page of fairtrade.com

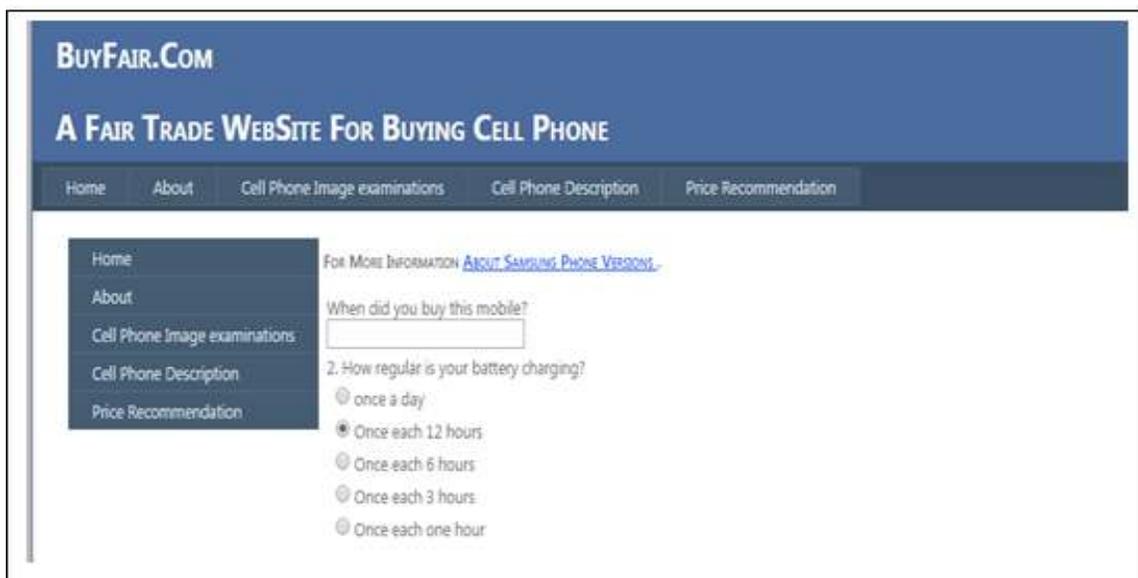


Fig. 11: About page of fairtrade.com

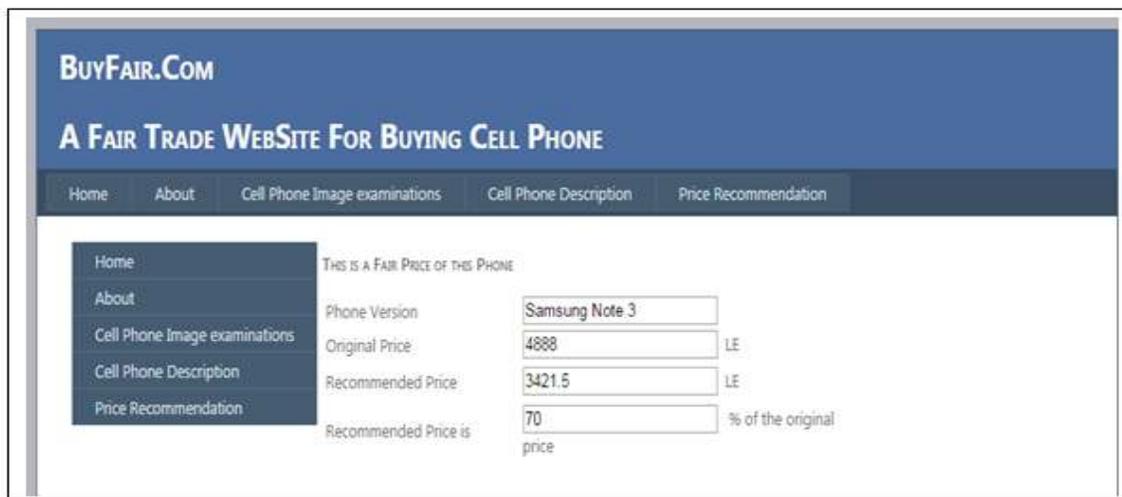


Fig. 12: About page of fairtrade.com

and extracts the values of determined criteria. One of these criteria is the version of the uploaded image phone with detecting its Congruence Percentage by comparing the phone image with the original image of that version.

Figure 10 presents a set of questions that appear to the user to get the remaining criteria values.

Finally, Fig. 11 presents the status of the phone and the recommended price

The presented case has been demonstrated to reveal the applicability of the proposed system for supporting the cell phone buyers in having a suitable price for the phone in consideration. However, we have run the system on 20 cases with different status and the results were satisfying (Fig. 12).

CONCLUSION

Fair trading is not always available due to the lack of knowledge between the seller and the buyer, especially in the market of used (second-hand) products. In this study, we presented a recommendation system for the used cell phone buyers to support them in having a fair price for the product that they are going to buy. The system detects the phone status and recommends a price for the phone according to its current status. The phone status has been detected using image processing and classification techniques. A web application has been built and a number of cases have been applied to the system for the Samsung cell phones.

However, more enhancements can be applied for more successful output; one direction is to perform a study for different classification algorithms for a final decision of the most suitable algorithm to apply according to the nature of data, as in this research as only two algorithms have been applied based on the recommendation of different researches. Another direction is to collect data and enrich the criteria for different cell phone types as this system only applied on Samsung versions.

REFERENCES

- A Nielsen Report, 2010. Global Trends in Online Shopping. Retrieved from: <http://hk.nielsen.com/documents/Q12010OnlineShoppingTrendsReport.pdf>.
- Advanced Projects R&D, 2005. Euclidean Distance Raw, Normalized, and Double-scaled Coefficients. Retrieved from: <http://www.pbarrett.net/techpapers/euclid.pdf>.
- Barbuceanu, M. and W.K. Lo, 2000. A multi-attribute utility theoretic negotiation architecture for electronic commerce. Proceeding of the 4th International Conference on Autonomous Agents, pp: 239-246.
- Bergen, J.R., P. Anandan, K.J. Hanna and R. Hingorani, 1992. Hierarchical model-based motion estimation. Proceeding of the 2nd European Conference on Computer Vision (ECCV'92), pp: 237-252.
- Bernsen, J., 1986. Dynamic thresholding of grey-level images. Proceeding of the 8th International Conference on Pattern Recognition. Paris, pp: 1251-1255.
- Breiman, L., 2001. Random forests. Mach. Learn., 45: 5-32.
- Brown, M. and D.G. Lowe, 2003. Recognising panoramas. Proceeding of the 9th International Conference on Computer Vision (ICCV'03). Nice, France, 2: 1218-1225.
- Chevalier, J.A. and D. Mayzlin, 2006. The effect of word of mouth on sales: Online book reviews. J. Marketing Res., 43(3): 345-354.
- Degeratu, A.M., A. Rangaswamy and J. Wu, 2000. Consumer choice behavior in online and traditional supermarkets: The effects of brand name, price, and other search attributes. Int. J. Res. Mark., 17(1): 55-78.

- Ding, L. and A. Goshtasby, 2001. On the canny edge detector. *Pattern Recogn.*, 34(4): 721-725.
- Fabrizio, J., M. Cord and B. Marcotegui, 2009. Text extraction from street level images. In: Stilla, U., F. Rottensteiner and N. Paparoditis (Eds.), *CMRT'09. IAPRS, Paris, France*, 38: 199-204.
- Fernández-Delgado, M., E. Cernadas, S. Barro and D. Amorim, 2014. Do we need hundreds of classifiers to solve real world classification problems? *J. Mach. Learn. Res.*, 15(1): 3133-3181.
- Förstner, W., 1986. A feature based correspondence algorithm for image matching. *Int. Arch. Photogramm. Remote Sens.*, 26(3): 150-166.
- Harris, C., 1992. *Geometry from Visual Motion*. In: Blake, A. and A. Yuille (Eds.), *Active Vision*. MIT Press, Cambridge, MA, USA, pp: 263-284.
- Huang, A., 2008. Similarity measures for text document clustering. *Proceeding of the NZCSRSC. Christchurch, New Zealand*, pp: 49-56.
- Huang, W., C.L. Tan and W.K. Leow, 2005. Associating text and graphics for scientific chart understanding. *Proceeding of the IEEE 8th International Conference on Document Analysis and Recognition*, 2: 580-584.
- Jain, A.K. and B. Yu, 1998a. Document representation and its application to page decomposition. *IEEE T. Pattern Anal.*, 20(3): 294-308.
- Jain, A.K. and B. Yu, 1998b. Automatic text location in images and video frames. *Pattern Recogn.*, 31(12): 2055-2076.
- Jung, K., K.I. Kim and A.K. Jain, 2004. Text information extraction in images and video: A survey. *Pattern Recogn.*, 37(5): 977-997.
- Kim, D. and A.P. Ammeter, 2008. Examining shifts in online purchasing behavior: Decoding the 'net generation'. *Proceeding of the Allied Academies International Conference on Academy of Information and Management Sciences*, 12.1: 7-12.
- Klenow, P.J. and O. Kryvtsov, 2008. State-dependent or time-dependent pricing: Does it matter for recent U.S. inflation? *Q. J. Econ.*, 123(3): 863-904.
- Lee, J., M. Podlaseck, E. Schonberg and R. Hoch, 2001. Visualization and analysis of clickstream data of online stores for understanding web merchandising [J]. *Data Min. Knowl. Disc.*, 5(1-2): 59-84.
- Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2): 91-110.
- Lucas, B.D. and T. Kanade, 1981. An iterative image registration technique with an application to stereo vision. *Proceeding of the 7th International Joint Conference on Artificial Intelligence*, 2: 674-679.
- Nayak, A.A., P.S., Venugopala, H. Sarojadevi and N.N. Chiplunkar, 2015. An approach to improvise canny edge detection using morphological filters. *Int. J. Comput. Appl.*, 116(9): 38-42.
- Pan, Y. and J.Q. Zhang, 2011. Born unequal: A study of the helpfulness of user-generated product reviews. *J. Retailing*, 87(4): 598-612.
- Schaffalitzky, F. and A. Zisserman, 2002. Multi-view matching for unordered image sets, or "How do I organize my holiday snaps?". *Proceeding of the 7th European Conference on Computer Vision-Part I (ECCV'02)*, pp: 414-431.
- Sushma, J. and M. Padmaja, 2009. Text detection in color images. *Proceeding of the IEEE International Conference on Intelligent Agent and Multi-Agent Systems (IAMA, 2009)*. Chennai, pp: 1-6.
- Tan, P.N., M. Steinbach and V. Kumar, 2006. *Introduction to Data Mining*. Pearson Addison Wesley, Boston, pp: 16.
- Tesseract-ocr., 1996. Tesseract-ocr. Retrieved from: <http://code.google.com/p/tesseract-ocr/>. (Accessed on: Oct 23, 2015).
- Walsh, D. and A.E. Raftery, 2002. Accurate and efficient curve detection in images: The importance sampling Hough transform. *Pattern Recogn.*, 35(7): 1421-1431.
- Wu, X., V. Kumar, J.R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G.J. McLachlan, A. Ng, B. Liu, P.S. Yu, Z.H. Zhou, M. Steinbach, D.J. Hand and D. Steinberg, 2008. Top 10 algorithms in data mining. *Knowl. Inf. Syst.*, 14: 1-37.