Pill Image Binarization for Detecting Text Imprints

Siroratt Suntronsuk
Department of Computer Science
Faculty of Science and Technology, Thammasat University
Pathum Thani, Thailand
siroratt_s@sci.tu.ac.th

Sukanya Ratanotayanon
Department of Computer Science
Faculty of Science and Technology, Thammasat University
Pathum Thani, Thailand
sratanot@cs.tu.ac.th

***Abstract*—Identifying pill is a major concern for both patients and pharmacists nowadays. Being able to extract pill information automatically will benefit both pill information retrieval and indexing processes. We aim to extract texts from pill imprints so that they can be used to search existing pill databases. This paper presents approaches for extracting texts from pill images based on i) a technique presented by Kasar[1] and ii) processing edge masks of imprints. We also compared different thresholds for binarizing extracted text area so that it can be used with optical character recognition (OCR). The result showed that the method based on using edge mask performed better, and the Otsu threshold gave the best results for binarization of the imprint area.**

Keywords—image processing; pill identification; text detection, binarization

I. Introduction

Due to the sheer number of available pills in the market, identifying pills by their appearances is currently an important issue for both pharmacists and patients. There are several circumstances in which pill identification is needed. For example, when a patient transfers to a new medical facility, a doctor must know what kind of medication he is on to efficiently continue his treatment. In some cases, a patient may simply wish to know what kinds of medicine he is taking.

Currently, there are several pill databases that users can input pill features via online forms to searched for matched pills [2], [3], [4], [5], [6]. However, being able to search with the pill image directly would be more convenient for users. In order to achieve this, techniques for accurately extracting pill appearance information such as size, color, shape and imprints are very important. Current extraction techniques are focusing on extracting image descriptors which are used to directly match a queried image with cataloged images. However, such descriptors cannot be used with the existing pill databases.

In this paper, we are focusing on extracting characters on the imprints, which are texts or symbols printed on the surface of a pill or a capsule, to help identifying it. Most existing pill databases include imprints as search criteria. Therefore, the extracted texts of an imprint can be used along with other pill features to search for matching pills in existing databases.

Extracting text from imprints is not a trivial task as most pill images do not work well with OCR for several reasons. Imprints in some images cannot be seen clearly because they are over/under exposure. Capsule pills may have multiple-color background. Also engraved imprints on tablet pills are hard to extract because the colors of imprints and background are very similar and the color of a character is not uniform.

We aimed to find an approach for binarizing text imprints so that the results can be passed through OCR to obtain correct text characters of the imprints. Two approaches were experimented and compared. The first approach is based on a modification of the technique that Kasar reported in [1]. Kasar's technique creates and calculates a threshold for binarizing each character individually using the intensity of edges of the character. The other approach creates an edge box containing the whole imprint area using a mask obtained from morphology of an edge map of the imprint. The area is then cropped and processed separately. Then, we experimented on applying different thresholds to the imprint area to binarize it.

We performed an experiment to evaluate these methods using 448 pill images obtained from online pill database "drug.com"[5] and U.S. National Library of Medicine (NLM) system[7]. The results showed that the method based using the edge mask along with Otsu thresholding gave the best results for binarization of the imprint area. However, the approach still need much improvement.

II. Related Work

Many organizations are aware of the concern for pill identification and establish online pill databases that users can use for this purpose. Online pill databases such as WebMD [2], RxList[3], GoodRX[4], Drug.Com[5], DailyMed[8] and Pillbox[6] need a user to input appearances information of a pill such as shape, color, size and imprint to search for its information via online forms. RXImage[9] even provides a public application programming interface (API) via web services so that programmers can query for pill images with pill appearances. Note that these databases do not contain information of medicines that are local made in Thailand as they are focusing on medicines produced in the United States. MIMS Thailand[10] does contain some Thai medicines made by big companies, but information of many local made pills are still not available online. FDA Thailand provides a system to search for information of all drugs available in Thailand, but it

978-1-5090-2033-1/16/$31.00 ©2016 IEEE

978-1-5090-2033-1/16/$31.00 ©2016 IEEE

does not contain appearance information or images of drugs. Therefore, it cannot be used for the purpose of pill identification.

Several existing works aimed to allow users to query for matching pill information directly using pill images. Such techniques would automatically extract required information from the image instead of having a user fill in feature formation manually. However, this is challenging as a pill image taken in different environments and different angles would be very different in many aspects, such as in terms of size and orientation. Images can also be blurry, over exposure, under exposure, or having unwanted shadows or reflections. Wang[11] proposed an approach that used a specialized device to capture image of pills in a medicine bag. His approach removed specular reflection of the medical bag and separated pills so that each pill image can be passed as input to other algorithms.

Other techniques aimed to extract pill descriptors using its features. HelpmePills[12] proposed a method that profiled a pill image to create a marker of the image. It took into consideration dimension, color, and shape of a pill. PillID[13], [14] proposed an approach to identify illegal drugs using size shape, color and imprint. A query image was computed to find a featured vector which can be used to compare against those of reference images in a gallery. Its later work employed Multi-scale Local Binary Pattern (MLBP)[15] and Scale Invariant Feature Transform (SIFT)[16] to find the key point mapping of the pill imprint. Yu et al.[17]proposed a method to extract pill imprint using modified stroke width transform (MSWT) which enhanced stroke width transform (SWT) with the switch function and accumulated gradients. Later on, they proposed an approach to create an image descriptor of a pill imprint using 2-step sampling distance set[18], and used it along with shape and color to identify a pill.

Although previous works dealt with extracting imprint information from pills, they were more concern with creating image descriptors for searching against image databases. However, such descriptors will not work against available pill databases and APIs that accept text input.

III. Pill Image Preparation Process

We randomly sampled images from “drug.com” which is an online pill database and from U.S. National Library of Medicine (NLM) system to use in this study. We have a total of 488 images. 40 images are used for a preliminary experiment to tune some parameters. The other 448 images are used in the evaluation that will be discussed in section V.

Since the obtained images have varied resolutions, we needed to perform image size normalization on them. We first found pill region by identifying the largest contour in an image and cropped the image to fit the pill region. Then, while maintaining the aspect ratio of the image, we resized the cropped image so that its height is 500 pixels.

Fig. 1. Original images. Top left) CD18E6D7.jpg, top right) 77233BF9.jpg, bottom left) 991D4CFA.jpg and bottom right) 4629A34D.jpg

Fig. 2. The same images after size normalization

IV. Binarization Methods

We explored two binarization approaches. The first approach is based on a modification of the technique that Kasar reported in [1]. However, we used a mask obtained from morphology of an edge map to directly identify a bounding box of a character. The second approach combined the identified bounding boxes to find imprint area and binarized only the selected area.

A. Kasar’s technique

Kasar proposed a technique to binarize color documents using a local threshold for each character. The technique first created an edge map of the image. Then, 8-connected component labeling was computed to find edge-box of each character. Some criteria based on size of the boxes were applied to filter out non-text boxes. Then it used pixel intensity along the edge in an edge-box to compute a local threshold that will be used for binarizing pixels in the box. Also, the technique judged whether the text is light-on-dark and dark-on-light automatically and binarized the pixels accordingly which should be very useful when binarizing capsules that usually contain different colors on each side.

However, results from our preliminary experiment showed that the nature of text in pill images broke several assumptions of the approach. For example, the pixel intensity in the text character was not uniform, and the size of characters was too big compared to the image size. Example results of binarizing pill images in Fig. 1with this method are shown in Fig. 3

Fig. 3. Kasar's binarization outputs of images in Fig.1

B. *Modified Kasar's technique*

We seek to enhance the results from Kasar's method. The bad results were usually caused by not finding the edge boxes for characters, or finding incorrect ones. Therefore, we aimed to better identify an edge box for each character. The steps we performed to find edge boxes were as follows.

1) *Enhancing detail of the imprint*

In the preliminary experiment, we found that the edge map produced from an engraved imprint often missed edges for text due to the similarity of the background and text colors. Therefore we applied interpolated convolution (IC) filter [19] with $\sigma_s = 10$ and $\sigma_r = 0.15$ to enhance the edge of the imprint. The enhanced results are shown in Fig. 4.

Fig. 4. Output enhancement after size normalization

2) *Edge preserve blur*

As the detail of the image increase, the noises are also increase. Therefore, we applied Normalized Convolution (NC) filter presented in [19] with $\sigma_s = 60$ and $\sigma_r = 0.4$ to blur the noises while still preserving the edges. The results are shown in Fig. 5.

Fig. 5. Edge preserve blur outputs

3) *Edge detection*

Once, we enhanced the imprint image to have better edges, we performed Canny's edge detection on separate channel of the image to create edge map E. Also we use Otsu Threshold for finding edges instead of using fixing global values.

$$E = E_R \vee E_G \vee E_B \tag{1}$$

4) *Morphology*

We used morphology to enlarge the edge in step 3) so that we could use it as a mask to find candidate areas of characters. First, we dilated the edges to make them thicker, and then we applied closing to reduce holes inside each character.

Fig. 6. Morphology outputs

5) *Finding character bounding boxes*

We identified all contours in the processed edge map. Then, for each contour, we used its bounding boxes to determine whether they are character contours. The criteria that we used are:

- Boxes that have both their width and height less than 20 pixels are rejected as noises.
- Boxes that either have its width or its height larger than half of the image's width and image's height, respectively, are also rejected. These boxes are mostly part of the contour of the pill.
- Aspect ratio can also be used to determine whether a box is a character box. A box that has its width more than 3 times larger than its height is rejected. This is because we usually do not have a character that is much wider than its height.
- All boxes that are enclosed in another boxes are rejected
- Lastly, once candidate bounding boxes are determined based on the above criteria, usually the largest box belong to one of the characters. Therefore, we further removed false positive by rejecting boxes that has less than half the area than the largest box.

6) *Computing threshold of each edge box*

Once we had boxes of each character identified, for each of them, we followed Kasar's algorithm to find threshold of each box. The foreground intensity was calculated from the average value of the pixels of the edges in the box, and background intensity was calculated from the average of intensity of points at the corner of the box. The foreground intensity was used as the threshold. The background intensity was used to determine whether the text is light-on-dark or dark-on-light.

The pixels outside of character boxes are assign value of 255 as they are background. The results after binarizing the images with this method are shown in Fig. 7.

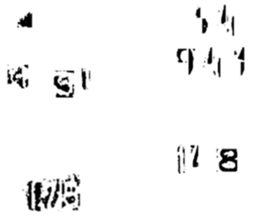

Fig. 7. Modified Kasar's output

C. Identifying and binarizing whole imprint area

We experimented on isolating imprint area and finding a suitable threshold to binarize it. To find the whole imprint area, we combined all character bounding boxes to create another bounding box that enclosed all of them as shown in Fig 8.

.

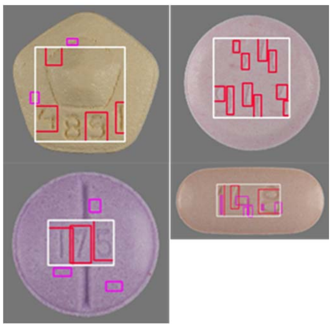

Fig. 8. The white boxes represent final text area for cropping

Then we used this box to crop the imprint area out of the pill image. Using only the imprint area, we used only pixel information in this area to find an appropriate threshold to binarize the extracted text area. The candidate thresholding methods were as follows:

1) Local threshold computed using mean adaptive thresholding[20]

This type of threshold was created by splitting an image into small regions and using a mean value of neighborhood pixels in each region as a threshold value. The binarized results are shown in Fig. 9.

Fig. 9. Imprint Area + Adaptive mean outputs

2) Global threshold computed using Otsu tresholding[20]

This threshold was computed from choosing a value that lied in the middle between the two peaks in the image histogram. So this type of threshold worked well with bimodal images. The binarized results are shown in Fig. 10.

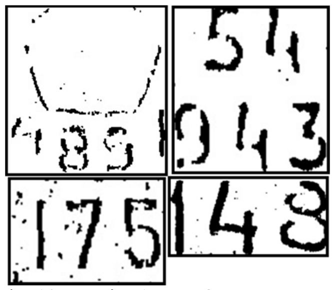

Fig. 10. Imprint Area + Otsu outputs

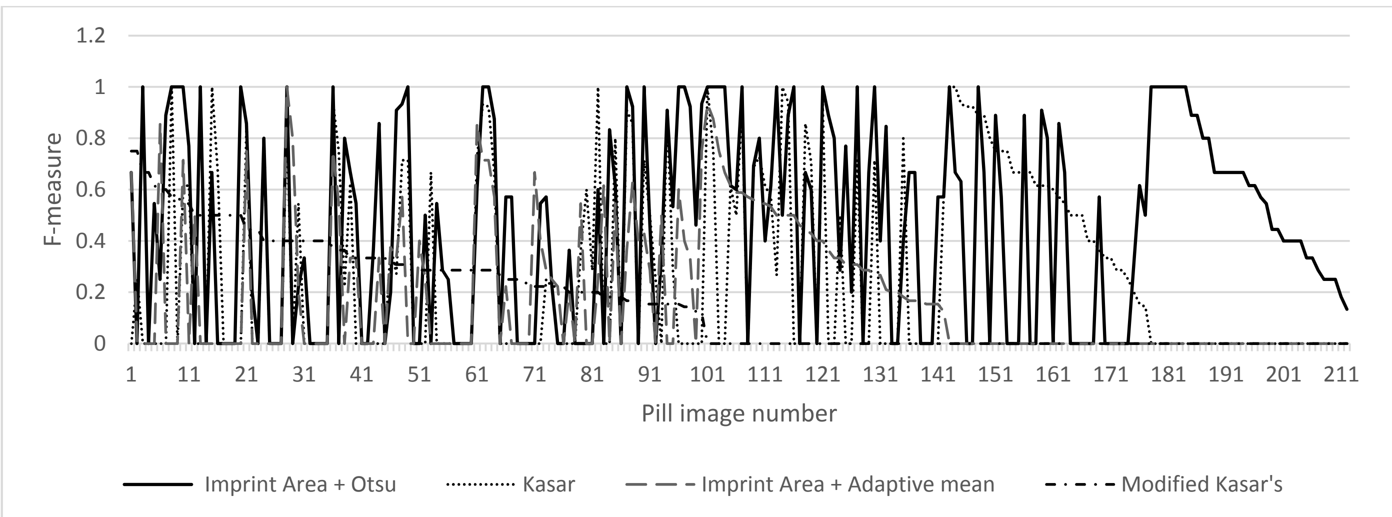

Fig. 11. F-Measure of each method (excluding images that every method fails)

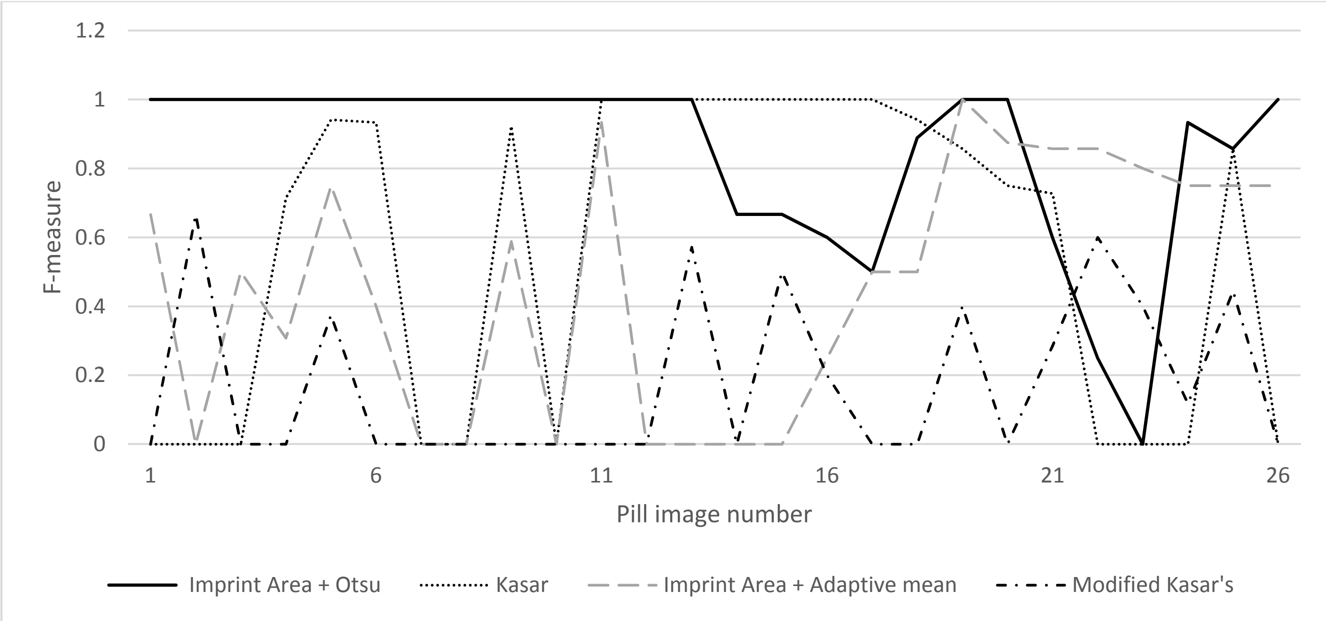

Fig. 12. Top 10 F-Measrure of each method

V. EVALUATION

We performed an experiment to evaluate methods presented in section IV. using 448 pill images obtained from online pill database "drug.com"[5] and U.S. National Library of Medicine (NLM) system[7]. Among them, 120 images contain printed imprints and 328 images contain engraved imprints. Each approach is implemented using OpenCV3.0 and Python2.7.

Two aspects were evaluated. An ability to detect an edge box of a character is very important to identify text area. Therefore, first, we compared the ability to detect character edge boxes of Kasar's approach and our edge mask approach. Next, we compared the performance of each approach in term of how well their results work with an OCR.

A. Identifying Text Area

In this experiment, we evaluated the edge mask approach and the connected components approach used by Kasar for identifying text boxes. We judged the accuracy of the result with the percentage of the correctly detected text boxes divided by the number of total characters on the imprint. The average accuracy of each approach is shown in TABLE I.

TABLE I. TEXT BOX DETECTION RESULTS

	Edge Mask	Kasar
Text box detection accuracy (%)	56.67	21.21

We saw performance improvement of the edge mask approach over the connected components used by Kasar. The reason for the failure of Kasar's approach should be because the approach had mismatch assumptions about characteristic of characters. The errors made by both approaches may cause by the noises from shadows or non-uniform nature of the engraved text.

B. OCR the Porcessed Text Area

The main objective of this study is to find an approach to binarize imprint images to use with an OCR. Therefore, this part of the study evaluated how well the OCR can process the binarized images produced by each approach. The OCR used in this study is Tesseract which is a well-know open-source OCR. It is considered one of the most accurate open source OCR engines nowadays. We passed the results of each approach to Tesseract.

TABLE II. AVERAGE PRECISION, RECALL AND F-MEASURE OF OCR RESULTS OF BINARIZED IMAGES CREATED BY EACH APPROACH

	Print Imprints				Engraved Imprints			
	Imprint Area + Otsu	Imprint Area + Adaptive mean	Modified Kasar's	Kasar	Imprint Area + Otsu	Imprint Area + Adaptive mean	Modified Kasar's	Kasar
Precision	0.710	0.291	0.114	0.405	0.111	0.038	0.075	0.059
Recall	0.610	0.282	0.114	0.419	0.113	0.040	0.079	0.067
f-measure	0.656	0.286	0.114	0.412	0.112	0.039	0.077	0.062

The output from Tesseract was evaluated using precision, recall and f-value. The definition of each are as follows.

$$precision = \frac{|\{relevant\ items\} \cap \{retrieved\ items\}|}{|\{retrived\ items\}|} \quad (2)$$

$$recall = \frac{|\{relevant\ items\} \cap \{retrieved\ items\}|}{|\{relevant\ items\}|} \quad (3)$$

$$F - Measure = \frac{2 \times precision \times recall}{precision + recall} \quad (4)$$

We considered each character in the Tesseract output of each imprint as a retrieved item. Each character of the actual text imprint is considered as a relevant item. The average precision, recall, and f-measure of result for each approach are presented in TABLE III.

TABLE III. AVERAGE PRECISION, RECALL AND F-MEASURE OF EACH APPROACH

	Imprint Area + Otsu	Imprint Area + Adaptive mean	Modified Kasar's	Kasar
Precision	0.258	0.100	0.084	0.144
Recall	0.237	0.097	0.087	0.1523
F-measure	0.247	0.099	0.086	0.148

Fig. 11 shows the f-measure of all images used in this study excluding ones that all methods failed to detect any texts. Fig. 12 shows the f-measure of images that were among the top-10 of each method. The results showed that the edge mask method for detecting imprint area along with Otsu thresholding performed the best. This may be because the color of the background of imprint image and the text are very similar. Also, there are noises that cause the local thresholding to perform poorly. Anyway, the result from all approaches still need much improvement.

VI. CONCLUSION

The proposed text imprint extraction method has 56.67% accuracy for detecting edge mask (twice the accuracy of Kasar method.) We also experimented applying different thresholds for binarization and found that Otsu thresholding perform best in our test set. We saw significant improvement in binarizing both printed imprints and engraved imprints. However, the binarize method should be improved further especially in term of Engraved imprint.

VII. FUTURE WORK

For engraved pills, the resulted texts from binarization is still not suitable for OCR as the characters' areas do not have uniform colors like printed texts, which may cause broken character images. Therefore, filling in the character area with uniform colors might help improve the OCR results. In addition, the final results can still be improved such as by training Tesseract for the fonts used by pills.

VIII. REFERENCES

[1] T Kasar, J Kumar, and A G Ramakrishnan, "Font and background color independent text binarization," 2007.

[2] "WebMD pill identification tool." [Online]. Available: http://www.webmd.com/pill-identification/. [Accessed: 14-Jan-2016].

[3] "RxList pill identification tool." [Online]. Available: http://www.rxlist.com/pill-identification-tool/article.htm. [Accessed: 14-Jan-2016].

[4] "GoodRx." [Online]. Available: http://www.goodrx.com/pill-identifier. [Accessed: 20-Apr-2016].

[5] "Drug.com Pill Identifier." [Online]. Available: http://www.drugs.com/pill_identification.html. [Accessed: 14-Jan-2016].

[6] "'Pillbox' National Library of Medicine, National Institutes of Health, United States." [Online]. Available: http://pillbox.nlm.nih.gov/. [Accessed: 14-Jan-2016].

[7] "U.S. National Library of Medicine." [Online]. Available: https://www.nlm.nih.gov/. [Accessed: 20-Apr-2016].

[8] "DAILYMED-National library of medicine." [Online]. Available: http://dailymed.nlm.nih.gov/dailymed/index.cfm. [Accessed: 14-Jan-2016].

[9] "RxIMAGE API." [Online]. Available: https://lhncbc.nlm.nih.gov/rximage-api. [Accessed: 20-Apr-2016].

[10] "Image Advanced search-MIMS thailand." [Online]. Available: https://www.mims.com/Thailand/image/AdvancedSearch. [Accessed: 14-Jan-2016].

[11] C. Wang, S. Kamata, and L. Ma, "A fast multi-view based specular removal approach for pill extraction," 2013, pp. 4126–4130.

[12] A. Cunha, T. Adão, and P. Trigueiros, "HelpmePills: A Mobile Pill Recognition Tool for Elderly Persons," *Procedia Technol.*, vol. 16, pp. 1523–1532, 2014.

[13] Y.-B. Lee, U. Park, and A. K. Jain, "PILL-ID: Matching and Retrieval of Drug Pill Imprint Images," 2010, pp. 2632–2635.

[14] Y.-B. Lee, U. Park, A. K. Jain, and S.-W. Lee, "Pill-ID: Matching and retrieval of drug pill images," *Pattern Recognit. Lett.*, vol. 33, no. 7, pp. 904–910, May 2012.

[15] T. Mäenpää and M. Pietikäinen, "Multi-scale Binary Patterns for Texture Analysis," in *Image Analysis*, vol. 2749, J. Bigun and T. Gustavsson, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 885–892.

[16] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[17] Z. Chen, J. Yu, S. Kamata, and J. Yang, "Accurate system for automatic pill recognition using imprint information," *IET Image Process.*, vol. 9, no. 12, pp. 1039–1047, Dec. 2015.

[18] J. Yu, Z. Chen, and S. Kamata, "Pill Recognition Using Imprint Information by Two-Step Sampling Distance Sets," 2014, pp. 3156–3161.

[19] E. S. L. Gastal and M. M. Oliveira, "Domain transform for edge-aware image and video processing," *ACM Trans. Graph.*, vol. 30, no. 4, p. 1, Jul. 2011.

[20] R. C. Gonzalez and R. E. Woods, *Digital image processing*, 3rd ed. Upper Saddle River, N.J: Prentice Hall, 2008.