

Enhancement of Images Inside PDF File

Tawfiq A. Al-Asadi a

Technology collage, Babylon University

Zaid Abdul Wahid Abod Salih

computer science, Babylon University

tawfiqasadi63@yahoo.com

Abstract:

In this paper, we present a new algorithm to open PDF file and extracting the images for applying enhancement filters. The proposed system consist of three stages, First stage for open PDF file and extract images from it, Second stage apply enhancement filter on extracted image depending on the nature of image, and the Third stage resave the enhancement images in original PDF file without any change in properties of PDF file.

Keywords: PDF, PDF File, image enhancement Filter, Open File.

الخلاصة

في هذا البحث، نقدم خوارزمية جديدة لفتح ملف PDF واستخراج الصور الموجودة داخله لتطبيق تحسين الصور على الصور المستخرجة، النظام المقترح يتكون من ثلاثة مراحل، المرحلة الأولى فتح الملف واستخراج الصور منه، المرحلة الثانية تطبيق فلتر تحسين الصور بالاعتماد على طبيعة الصور المستخرجة، المرحلة الثالثة إعادة حفظ الصور المحسنة في الملف الاصلي دون اي تغيير في خصائص الملف.
الكلمات المفتاحية: PDF، ملف، صورة تعزيز تصفية، فتح ملف.

1-Introduction

The Adobe Portable Document Format (PDF) is the native file format of the Adobe Acrobat family of products. The goal of these products is to enable users to exchange and view electronic documents easily and reliably, independently of the environment in which they were created. PDF relies on the same imaging model as the PostScript page description language to describe text and graphics in a device-independent and resolution-independent manner. To improve performance for interactive viewing, PDF defines a more structured format than that used by most PostScript language programs. PDF also includes objects, such as annotations and hypertext links that are not part of the page itself but are useful for interactive viewing and document interchange.

2-Background

We will divide this step to two parts:

A- The PDF file Consists of collection of objects that describe data, image, properties, and other information that together describe the appearance of one or more *pages* in PDF file. The component of PDF files as shown in Figure (1), (2) and (3) [Adobe team, 2010]:

1- Objects. A PDF document is a data structure composed from a small set of basic types of data object. Section 3.1, "Lexical Conventions," describes the character set used to write objects and other syntactic elements. "Objects," describes the syntax and essential properties of the objects themselves. "Stream Objects," provides complete details of the most complex data type, the stream object.

2- File structure. The PDF file structure determines how objects are stored in a PDF file, how they are accessed, and how they are updated. This structure is independent of the semantics of the objects, "File Structure," describes the file structure "Encryption," describes a file-level mechanism for protecting a document's contents from unauthorized access.

3- Document structure. The PDF document structure specifies how the basic object types are used to represent components of a PDF document: pages, fonts, annotations,

and so forth. “Document Structure,” describes the overall document structure; later chapters address the detailed semantics of the components.

4- *Content streams.* A PDF *content stream* contains a sequence of instructions describing the appearance of a page or other graphical entity. These instructions, while also represented as objects, are conceptually distinct from the objects that represent the document structure and are described separately. “Content Streams and Resources,” discusses PDF content streams and their associated resources.

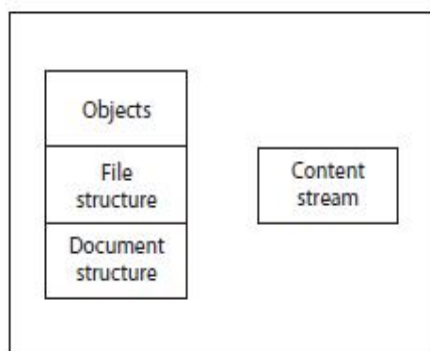


Figure (1) Describes the body
of PDF file

```
2 0 obj
<</Type/Pages/Count 1/Kids[ 3 0 R] >>
endobj
3 0 obj
<</Type/Page/Parent 2 0 R/Resources<<XObject<</Image5 5 0 R/Image6 6 0 R/Image7 7 0 R/Image8 8 0 R>>/Font
endobj
4 0 obj
<</Filter/FlateDecode/Length 214>>
stream
xœuêY'ëBtâñ#çl'zJq%ٲك'فهاء«UٲwJgZجندٲttC èll-q+Ex†ws9g:èl@-ف-pld.g@Fai
```

Figure (2) Explain of data in PDF

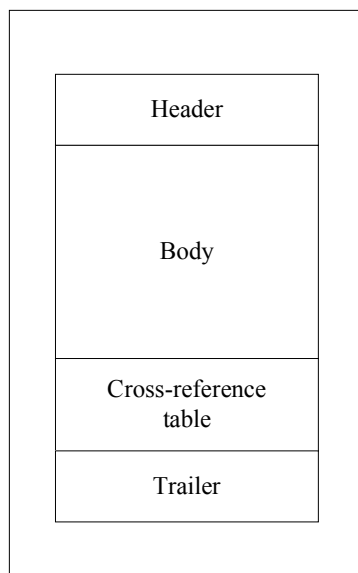


Figure (3): PDF Components.

B- Image enhancement processes consist of a collection of techniques that seek to improve the visual appearance of an image or to convert the image to a form better suited for analysis by a human or a machine. In an image enhancement system, there is no conscious effort to improve the fidelity of a reproduced image with regard to some ideal form of the image, as is done in image restoration. Actually, there is some evidence to indicate that often a distorted image, for example, an image with amplitude overshoot and undershoot about its object edges, is more subjectively pleasing than a perfectly reproduced original.[William, 2001]

Image enhancement is basically improving the interpretability or perception of information in images for human viewers and providing 'better' input for other

automated image processing techniques. The principal objective of image enhancement is to modify attributes of an image to make it more suitable for a given task and a specific observer. During this process, one or more attributes of the image are modified. The choice of attributes and the way they are modified are specific to a given task. Moreover, observer-specific factors, such as the human visual system and the observer's experience, will introduce a great deal of subjectivity into the choice of image enhancement methods. There exist many techniques that can enhance a digital image without spoiling it. The enhancement methods can broadly be divided in to the following two categories:

1. Spatial Domain Methods
2. Frequency Domain Methods

In spatial domain techniques, we directly deal with the image pixels. The pixel values are manipulated to achieve desired enhancement. In frequency domain methods, the image is first transferred in to frequency domain. It means that, the Fourier Transform of the image is computed first. All the enhancement operations are performed on the Fourier transform of the image and then the Inverse Fourier transform is performed to get the resultant image. These enhancement operations are performed in order to modify the image brightness, contrast or the distribution of the grey levels. As a consequence the pixel value (intensities) of the output image will be modified according to the transformation function applied on the input values.

Image enhancement is applied in every field where images are ought to be understood and analyzed. For example, medical image analysis, analysis of images from satellites etc. Image enhancement simply means, transforming an image f into image g using T . (Where T is the transformation. The values of pixels in images f and g are denoted by r and s , respectively. As said, the pixel values r and s are related by the expression, Figure (4) Showing the effect of Image Enhancement. [Raman,2010].



Figure 4 .the effect of Image Enhancement

3-Proposed System

In general, the proposed system as illustrated in figure(5) and (6) consist of three stages, first one is open PDF file and extract images from it, Second stage apply enhancement filter on extracted image, Third stage resave the enhancement images in original PDF file.

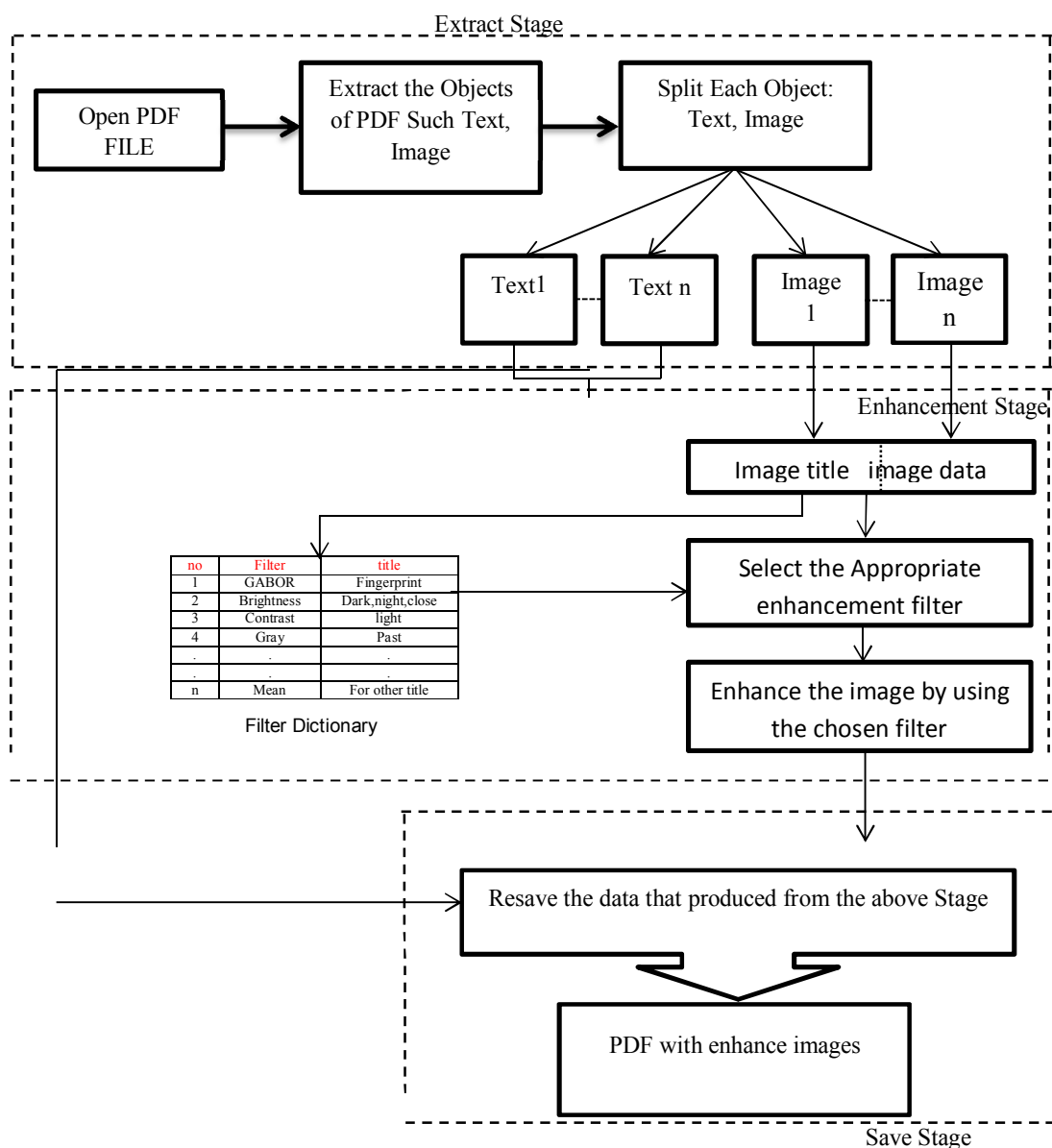


Figure (5) Block Diagram of Proposed System

3.1- Extract Stage

this stage is divided into the following steps:-

- 1-Open PDF as File, the Data Structure of the PDF File will be loading
- 2- Extract the Objects of PDF Such Text, Image, the data of image and text will extract from the PDF data.
- 3- Split Each Object: Text, Image. The data of image and text will be splitting from each other.

The Algorithm for Extract Stage:

INPUT: PDF File

OUTPUT:all the image in the PDF File,PDF Text

1-open pdf file

2-read the header of pdf file to get the number of images in the file, the right name of objects for images data in the file, and the Data of Text.

3-for i=1 to the no. of image

Get the data of image (i) and save it ,to be exist for the next stage

Endfor

4- get the data of text

End algorithm

3.2 - Apply Enhancement Filter on Extracted Image

From the above stage we have the data of all images in PDF file, now we can apply any enhancement filter for any image extracted from the PDF file.

Let we have the data of image 1 that extracted from pdf file,now we can apply on it any filter we want to be aplynig on this image , assume we want to aply the contrast filter to this image:

3.2.1 get the data of image(1).

3.2.2 apply contrast filter to it.

3.2.3 save the new data, to use it in the next stage.

3.3 - Resave the Enhancement Images in Original PDF File

After applying the second stage now we can save the enhance images to the PDF file on the same location and properties. See Figure (6) .

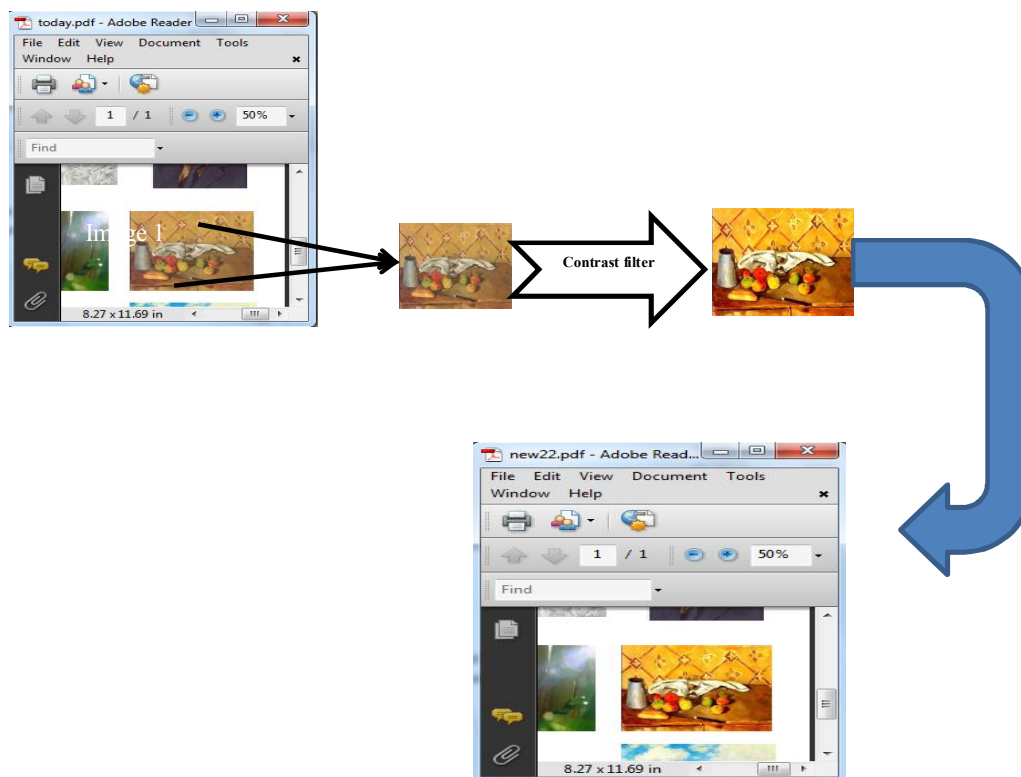


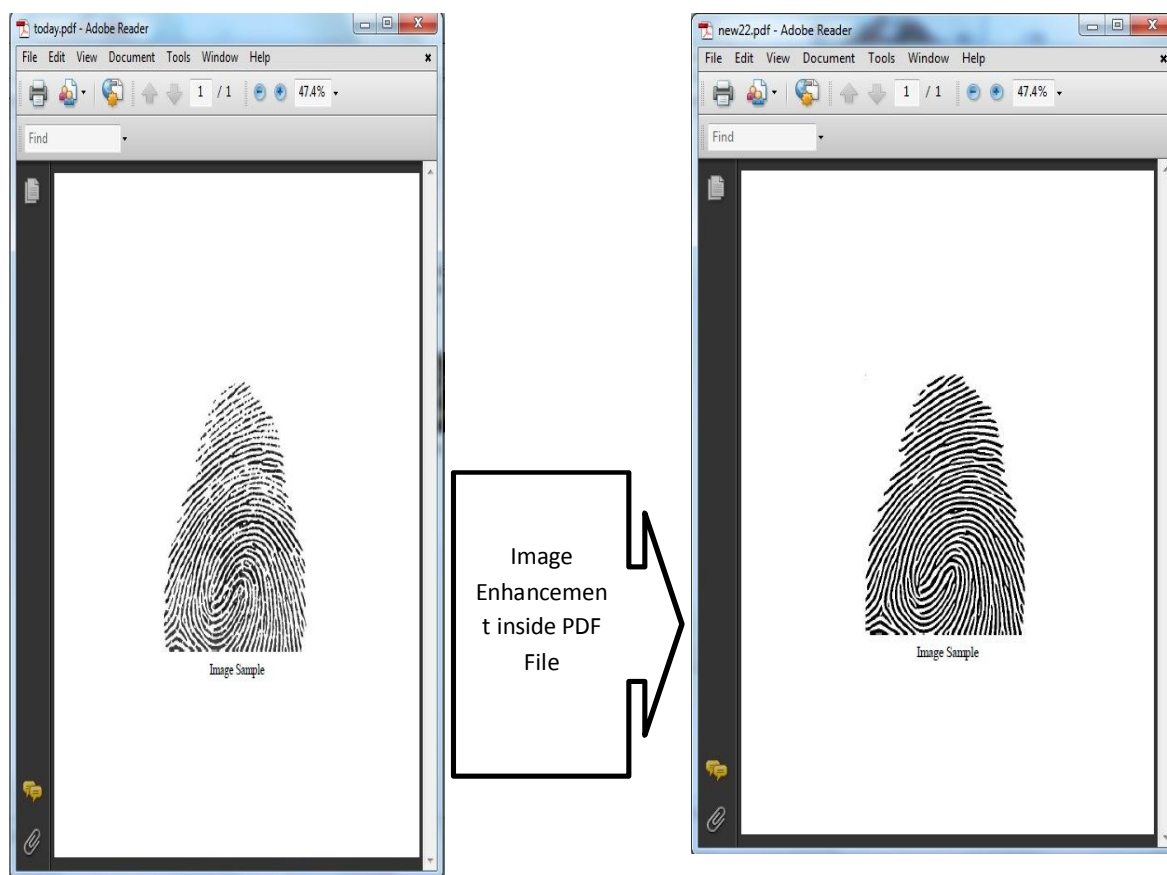
Figure (6) show the enhancement image inside Pdf by using the contrast filter

4-Result

The proposed system is applied on samples of the PDF files as follows:

Case (1)

The PDF file is Contain a single image represent fingerprint image with size of (460×472), we will take this image and apply the Gabor filter on it and replace the old image with new one “enhanced image”, as illustrated in figure (7).



(a)Original PDF

(b) PDF after the image enhancement

Figure (7) Results of case (1)

The resulted PDF file document has the same image but it enhanced with gabor filter, the size of resulted image is the same size of Original image (460×472), and with same format. Peak signal to noise ratio is (33.17) and this is an accepted ratio; this ratio is calculated by using equation (1)

$$SNR_{peak} = 10 \log_{10} \frac{(l-1)^2}{\frac{1}{n \times m} \sum_{r=0}^{n-1} \sum_{c=0}^{m-1} [g(r,c) - I(r,c)]^2} \dots (1)$$

Where, l is the number of gray levels in fingerprint image.

$I(r,c)$ are the pixels before embedding process in fingerprint image.

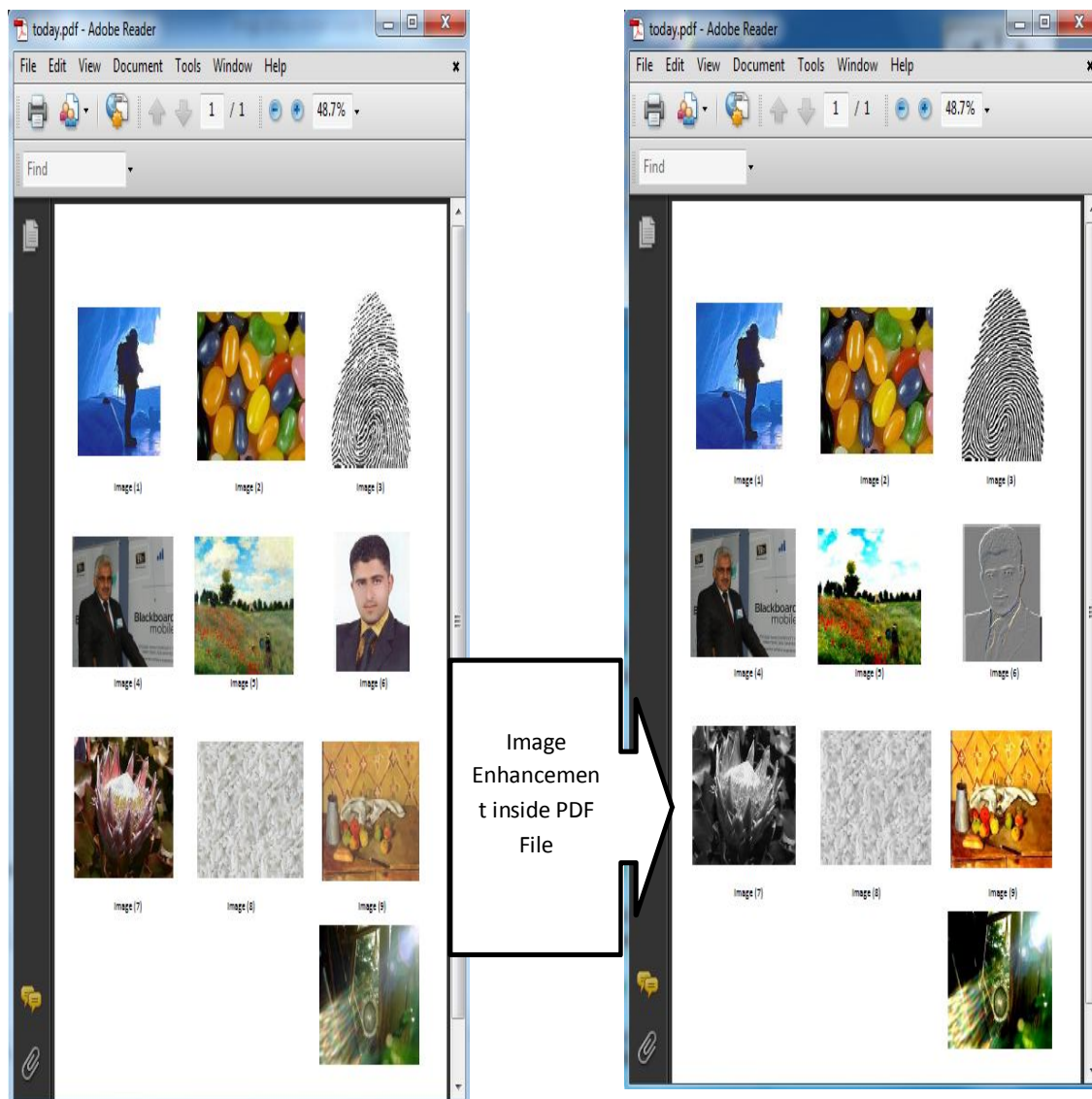
$g(r,c)$ are the pixels after embedding process in fingerprint image.

n, m are the width and height of fingerprint image respectively.

Case (2)

The PDF file is Contain (10) images represent deferent image with a deferent sizes, we will take these images and apply the deferent filter on them such as Gabor, embossing, contrast, diffuse, gray level, smooth, sharpen, Median, mean, and others filters, and replace the old images with new one “enhanced images”, as illustrated in figure (8).

The resulted PDF file document has the same images but it enhanced with deferent filters, the size of resulted images is the same sizes of Original images, and with same format.



(a) Original PDF

(b) PDF after the image enhancement

Figure (8) Results of case (2)

5- Conclusion

The image enhancement in PDF file has been done by using a new algorithm, The proposed system consist of three stages ,First stage for open PDF file and extract images from it, Second stage apply enhancement filter on extracted image, Third stage resave the enhancement images in original PDF file without any change in properties of PDF file.

References

- Adobe team 2010 Systems Incorporated," PDF Reference ", USA
bernd jahne, 2002" Digital Image Processing", (Springer),
Marcus Sachs 2011 "Malicious PDF Documents Explained", (iEEE),
Raman Maini and Himanshu Aggarwal 2010," A Comprehensive Review of Image Enhancement Techniques", (Journal of Computing)
SumitTandon, 2005 " edge detection", university of Texas at Arlington,
William K. Pratt, 2001"Digital Image Processing: *PIKS Inside, Third Edition*", USA,