

# Adaptive Binarization of Document Images Using Phase Congruency

Yibin TIAN\*, Wei MING\*

## Abstract

Existing adaptive binarization methods usually compute a threshold within each local window of fixed size. Such fixed window may lead to problematic division of document images.

This report presents an adaptive binarization method for document images that takes into account unique characteristics of documents. First, image phase congruency (IPC) is calculated for the image; then, connected component analysis is carried out on the IPC edges to segment out a local window for each symbol; afterwards, a local threshold is calculated for each window to binarize the corresponding portion of gray scale image. Two advantages of this method are: (1) the local windows are tailored to individual symbols; (2) it is insensitive to significant non-uniform illumination, a common issue in camera-captured images.

Using both synthetic and real camera-captured document images, we evaluated the proposed method with both psychophysical and OCR methods to show its excellent performance among adaptive binarization methods in dealing with non-uniform illumination and noise.

## 1 Introduction

In document processing, there is often a need to generate a binary image from a color or gray-scale image via binarization. For example, most Optical Character Recognition (OCR) methods utilize image binarization as the first step. Document images can be obtained via scanners or cameras. With the widespread usage of digital cameras on smart phones, a user can conveniently take a picture of a document or a product label bearing text. Unfortunately, there are a number of challenges that are unique to camera-captured document images, such as perspective distortion, non-uniform illumination, focusing blur, complex background, and so on<sup>1) 2)</sup>. Non-uniform illumination is especially detrimental to the binarization of document images.

The threshold for binarization can be selected either globally or locally. In the global approach, the entire input image is used to calculate the threshold, and subjected to the same thresholding operation. In the local approach, the threshold is computed and applied only within a local region, thus it is adaptive in nature. Obviously, local adaptive binarization methods are better suited for images degraded by non-uniform illumination.

There are a number of adaptive binarization methods in the literature, and many of them compute a local threshold based on certain image statistics of a local region. Bernsen chose the average of the maximum and minimal pixel values in the surrounding region as the local threshold<sup>3)</sup>. Niblack combined the mean and standard deviation of pixel values in a sliding window as the local threshold<sup>4)</sup>. White and Rohrer used a running average of the pixel values as the local threshold<sup>5)</sup>. Abutaleb determined the local threshold by maximizing the entropy of each local region<sup>6)</sup>. Tsai proposed to find the local threshold to preserve image moments<sup>7)</sup>. Recently, more sophisticated adaptive binarization methods have been proposed to take into account the unique characteristics of document images. Gatos and Pratikakis et al. utilized background

---

## Authors



Yibin TIAN



Wei MING

---

\* Konica Minolta Laboratory U.S.A., Inc.

removal to deal with non-uniform illumination, using stroke connectivity preservation as a constraint<sup>8</sup>. Sauvola and Pietikainen treated a document image as a collection of background, texts and pictures, and binarized text and non-text regions differently<sup>9</sup>.

Most methods described above need to specify a fixed local window size, which the quality of the binarization results depends on. Unfortunately, it is difficult to apply a predetermined local window size universally to many document images to obtain good results. In this report we propose a new approach that does not require specification of local window size, instead, it is determined in such a way that each symbol falls within its own local window. Very briefly, we first utilize image phase congruency to reduce the impact of non-uniform illumination. Then connected component analysis is applied to the edges generated from the phase congruency of the image to separate symbols and determine local windows. Afterwards, a local threshold is determined for each symbol using existing methods. Our approach takes into consideration of the structural characteristics of document images while simultaneously reducing the impact of non-uniform illumination. The proposed method is compared to representative existing methods using both synthetic and real camera-captured document images to show its advantages.

## 2 Methods

### 2.1 Image phase congruency

Image phase congruency is a measure that describes the phase relationship among frequency components of an image<sup>10</sup>. It has been used to extract edge-like image features<sup>10, 11</sup>. The use of phase congruency for detecting features has significant advantages over gradient-based methods as it is a dimensionless quantity invariant to image brightness or contrast, which results in insensitivities to non-uniform illumination in camera captured images<sup>12</sup>. Phase congruency is difficult to compute based on its definition. Here we utilize the implementation by Kovessi<sup>11</sup>, as described below.

Let  $S(x)$  denote a 1D signal and  $F_n^e(\cdot)$  and  $F_n^o(\cdot)$  the even-symmetric (cosine) and odd-symmetric (sine) log-Gabor wavelet filters at scale  $n$ . The responses of each pair of filters form a response vector

$$[e_n(x), o_n(x)] = [F_n^e\{S(x)\}, F_n^o\{S(x)\}]$$

The amplitude, phase angle and phase deviation are

$$A_n(x) = \sqrt{[e_n(x)]^2 + [o_n(x)]^2}$$

$$\Theta_n(x) = \text{atan}(e_n(x), o_n(x))$$

$$\Delta\Theta_n(x) = \cos[\Theta_n(x) - \overline{\Theta(x)}] - |\sin[\Theta_n(x) - \overline{\Theta(x)}]|$$

where  $\overline{\Theta(x)}$  is the average phase angle of all scales. Phase congruency is computed as

$$PC(x) = \frac{\sum_n w(x) f_p[A_n(x) \Delta\Theta_n(x) - T]}{\sum_n A_n(x)} \quad (1)$$

where  $f_p[u] = u$  if  $u > 0$ ,  $f_p[u] = 0$  if  $u \leq 0$ ;  $T$  is a noise threshold, and  $w(x)$  a weight function.

To extend the above definition to 2D, the phase congruency can be computed at multiple orientations of the 2D image and the results summed up as

$$PC(x, y) = \frac{\sum_{or} \sum_n w_{or,n}(x, y) f_p[A_{or,n}(x, y) \Delta\Theta_{or,n}(x, y) - T]}{\sum_n A_{or,n}(x, y)} \quad (2)$$

where  $w_{or,n}(x, y)$ ,  $A_{or,n}(x, y)$  and  $\Delta\Theta_{or,n}(x, y)$  are similar to the corresponding variables in Eq. 1 but are calculated at each orientation. Phase congruency values range from 0 to 1; for practical purposes, we can treat the computation of 2D phase congruency of an image as a non-linear image transformation. The 2D phase congruency output is of the same size as the input image. For convenience, we call this 2D transformation output of the input image as its phase congruency map.

It should be noted that this implementation has a built-in de-noising mechanism, as indicated by the noise threshold  $T$  in Eq. 2, which is adaptively determined during the phase congruency computing process<sup>11</sup>.

### 2.2 Adaptive binarization for document images

Two commonly encountered binarization issues in camera-captured document images are noise and non-uniform illumination, especially those captured by low-quality cameras on cellular phones. As described above, the phase congruency implementation used in this report can address both issues at the same time.

Our adaptive binarization method using phase congruency consists of four major steps (Fig. 1): (1) a phase congruency map is computed for input image, with de-noising carried out automatically; (2) edges are extracted from the phase congruency map via hysteresis thresholding; (3) connected components and their bounding boxes are obtained from the edges; (4) the gray input image is divided into individual symbols using the component bounding boxes, and

each symbol is binarized individually with an existing thresholding method. Each step is explained in more details below.

#### STEP 1: Computation of phase congruency map

Phase congruency map as shown in Eq.2 is computed on gray images only. Most camera-captured document images are in RGB color format. There are two options to apply Eq.2 to such images: first, converting them to gray images; second, applying Eq.2 to each of the RGB components and combining the three maps together. For a certain images, the second approach may give better results since conversion from color to gray images may lead to significant information loss, as demonstrated on focus measures by Tian et al.<sup>13)</sup>. However, application of Eq.2 is computationally expensive as the transformation is done at multiple scales and orientations, so we choose to use the conversion method for color images.

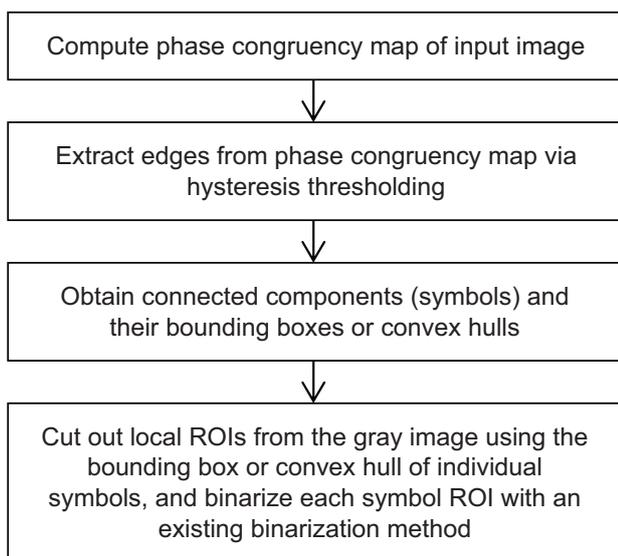


Fig. 1 Major steps of the proposed binarization method.

The numbers of scales and orientations used in Eq.2 need to be chosen carefully. Using larger number of scales and orientations may yield more accurate phase congruency map, at the cost of more computations. In this report, we choose 5 scales and 4 orientations.

As pointed out earlier, adaptive de-noising is automatically carried out as an inherent part of phase congruency computation in Eq.2. So we do not apply additional de-noising step. However, for images with low signal-to-noise ratio (SNR), an edge-preserving filtering step before applying Eq.2 may yield better edge extraction results in Step 2. Either anisotropic diffusion or bilateral filter can serve this purpose<sup>14) 15)</sup>.

#### STEP 2: Edge extraction from phase congruency map

Edge extraction from the phase congruency map is simple as edges in an image exhibit high phase congruency. In other words, a greater value at a pixel in the phase congruency map indicates a higher probability that the pixel is on an edge in the input image. Therefore we can simply binarize the phase congruency map to obtain an edge map, a binary image containing edge pixels as foreground and non-edge pixels as background. To achieve better results, hysteresis thresholding is used, similar to the thresholding in Canny edge detector<sup>16)</sup>.

#### STEP 3: Connected component analysis of edge map

Morphological operations are used to remove spurious edges and/or connect broken edges. Then standard connected component analysis is carried out on the cleaned edge map. The bounding box or convex hull of each component is saved for the next step.

Characteristics of the connected components can be analyzed to obtain statistical information of the connected components. For example, each connected component may be measured by its height and width, and the height and width distributions of all connected components may be analyzed. For a typical document image, the connected components are mostly text characters which tend to have similar heights and widths compared to those for images or graphics. Connected components for non-text (tables, graphics) also tend to be much larger than those for text characters. Based on this statistical information and the fact that characters tend to form lines, text and non-text regions can be separated and handled differently. For example, pictures and graphics may be binarized differently, as suggested by Sauvola et al.<sup>9)</sup>.

It should be pointed out that a single character may be divided into multiple connected components, some of which may be very small. With a little more processing, they can be merged together to form true characters. However, this is highly language dependent: it is easy to do in Indo-European languages (such as English, German and French), but much more challenging in East Asian languages (such as Chinese, Japanese and Korean). In this report, we do not merge connected components, and we use the term “symbols” from now on to distinguish them from characters.

#### STEP 4: Binarization of individual symbols

With the bounding box or convex hull information for each symbol from the previous step, we can cut

out individual region of interest (ROI) for each symbol from the gray image. Each bounding box or convex hulls should be extended by a few pixels as long as it does not overlap with its neighboring symbols. This can avoid cutting out some extremely simple symbols without much background.

Each gray ROI can be binarized individually using one of the conventional binarization methods described earlier. We use Niblack's method in our implementation, but only one threshold is computed for the entire ROI. Once all gray ROIs are binarized, the outputs can be assembled to form the binary output image.

### 3 Results

We evaluated the proposed method on both synthetic and real camera-captured document images, and compared it to 6 existing adaptive binarization methods. Both psychophysical and objective methods were used to evaluate the binarization results.

(a) Gray image

*Abstract*—This paper presents an evaluation of eleven locally adaptive binarization methods for gray scale images with low contrast, variable background intensity and noise. Niblack's method with the addition of the postprocessing step of Yanowitz and Bruckstein's method added performed the best, and was also one of the fastest binarization methods.

*Keywords*—Locally adaptive binarization, Thresholding, Evaluation, Utility maps, Document images.

**I. INTRODUCTION**

**T**HERE is currently a large and growing interest in the field of document image analysis. Numerous research groups are trying to design systems for extracting relevant information from such diverse documents as engineering drawings, maps, magazines and newspapers, forms, and mail envelopes [1-3]. In most of these systems, binarization of the scanned gray level image (labeling each pixel print or background) is done prior to further processing. This paper will focus on binarization of map images.

(b) Bensen's method

*Abstract*—This paper presents an evaluation of eleven locally adaptive binarization methods for gray scale images with low contrast, variable background intensity and noise. Niblack's method with the addition of the postprocessing step of Yanowitz and Bruckstein's method added performed the best, and was also one of the fastest binarization methods.

*Keywords*—Locally adaptive binarization, Thresholding, Evaluation, Utility maps, Document images.

**I. INTRODUCTION**

**T**HERE is currently a large and growing interest in the field of document image analysis. Numerous research groups are trying to design systems for extracting relevant information from such diverse documents as engineering drawings, maps, magazines and newspapers, forms, and mail envelopes [1-3]. In most of these systems, binarization of the scanned gray level image (labeling each pixel print or background) is done prior to further processing. This paper will focus on binarization of map images.

(c) Niblack's method

*Abstract*—This paper presents an evaluation of eleven locally adaptive binarization methods for gray scale images with low contrast, variable background intensity and noise. Niblack's method with the addition of the postprocessing step of Yanowitz and Bruckstein's method added performed the best, and was also one of the fastest binarization methods.

*Keywords*—Locally adaptive binarization, Thresholding, Evaluation, Utility maps, Document images.

**I. INTRODUCTION**

**T**HERE is currently a large and growing interest in the field of document image analysis. Numerous research groups are trying to design systems for extracting relevant information from such diverse documents as engineering drawings, maps, magazines and newspapers, forms, and mail envelopes [1-3]. In most of these systems, binarization of the scanned gray level image (labeling each pixel print or background) is done prior to further processing. This paper will focus on binarization of map images.

(d) White and Rohrer's method

*Abstract*—This paper presents an evaluation of eleven locally adaptive binarization methods for gray scale images with low contrast, variable background intensity and noise. Niblack's method with the addition of the postprocessing step of Yanowitz and Bruckstein's method added performed the best, and was also one of the fastest binarization methods.

*Keywords*—Locally adaptive binarization, Thresholding, Evaluation, Utility maps, Document images.

**I. INTRODUCTION**

**T**HERE is currently a large and growing interest in the field of document image analysis. Numerous research groups are trying to design systems for extracting relevant information from such diverse documents as engineering drawings, maps, magazines and newspapers, forms, and mail envelopes [1-3]. In most of these systems, binarization of the scanned gray level image (labeling each pixel print or background) is done prior to further processing. This paper will focus on binarization of map images.

(e) Abutaleb's method

*Abstract*—This paper presents an evaluation of eleven locally adaptive binarization methods for gray scale images with low contrast, variable background intensity and noise. Niblack's method with the addition of the postprocessing step of Yanowitz and Bruckstein's method added performed the best, and was also one of the fastest binarization methods.

*Keywords*—Locally adaptive binarization, Thresholding, Evaluation, Utility maps, Document images.

**I. INTRODUCTION**

**T**HERE is currently a large and growing interest in the field of document image analysis. Numerous research groups are trying to design systems for extracting relevant information from such diverse documents as engineering drawings, maps, magazines and newspapers, forms, and mail envelopes [1-3]. In most of these systems, binarization of the scanned gray level image (labeling each pixel print or background) is done prior to further processing. This paper will focus on binarization of map images.

(f) Tsai's method

*Abstract*—This paper presents an evaluation of eleven locally adaptive binarization methods for gray scale images with low contrast, variable background intensity and noise. Niblack's method with the addition of the postprocessing step of Yanowitz and Bruckstein's method added performed the best, and was also one of the fastest binarization methods.

*Keywords*—Locally adaptive binarization, Thresholding, Evaluation, Utility maps, Document images.

**I. INTRODUCTION**

**T**HERE is currently a large and growing interest in the field of document image analysis. Numerous research groups are trying to design systems for extracting relevant information from such diverse documents as engineering drawings, maps, magazines and newspapers, forms, and mail envelopes [1-3]. In most of these systems, binarization of the scanned gray level image (labeling each pixel print or background) is done prior to further processing. This paper will focus on binarization of map images.

(g) Sauvola's method

*Abstract*—This paper presents an evaluation of eleven locally adaptive binarization methods for gray scale images with low contrast, variable background intensity and noise. Niblack's method with the addition of the postprocessing step of Yanowitz and Bruckstein's method added performed the best, and was also one of the fastest binarization methods.

*Keywords*—Locally adaptive binarization, Thresholding, Evaluation, Utility maps, Document images.

**I. INTRODUCTION**

**T**HERE is currently a large and growing interest in the field of document image analysis. Numerous research groups are trying to design systems for extracting relevant information from such diverse documents as engineering drawings, maps, magazines and newspapers, forms, and mail envelopes [1-3]. In most of these systems, binarization of the scanned gray level image (labeling each pixel print or background) is done prior to further processing. This paper will focus on binarization of map images.

(h) Proposed method using IPC

*Abstract*—This paper presents an evaluation of eleven locally adaptive binarization methods for gray scale images with low contrast, variable background intensity and noise. Niblack's method with the addition of the postprocessing step of Yanowitz and Bruckstein's method added performed the best, and was also one of the fastest binarization methods.

*Keywords*—Locally adaptive binarization, Thresholding, Evaluation, Utility maps, Document images.

**I. INTRODUCTION**

**T**HERE is currently a large and growing interest in the field of document image analysis. Numerous research groups are trying to design systems for extracting relevant information from such diverse documents as engineering drawings, maps, magazines and newspapers, forms, and mail envelopes [1-3]. In most of these systems, binarization of the scanned gray level image (labeling each pixel print or background) is done prior to further processing. This paper will focus on binarization of map images.

Fig.2 Gray document image with simulated non-illumination and its binarization results of 7 methods.

### 3.1 Synthetic document images

In the first set of experiments, we artificially introduced non-uniform illumination and noise to a gray document image. The non-uniform illumination is Gaussian distributed, with its peak at the center of the image, as shown in Fig. 2 (a). The image is binarized using the proposed method and 6 existing adaptive methods<sup>2-7)9)</sup>. We used the implementation of existing methods in Gamera Python at default settings<sup>17)</sup>. The binarization results are shown in Fig. 2 (b-h).

For psychophysical evaluation, the 7 binary images were printed on high quality A4 papers using Konica Minolta bizhub C650 MFP. They were shown to each subject, who was asked to choose the two best and two worst ones from them. The subjects were not aware of the methods used.

For objective evaluation, we recognize the binary images with free OCR software from ABBYY (FineReader Online), and count the number of character errors in the output text files.

Table 1 shows the evaluation results. As shown in first two rows of Table 1, the result of the proposed method was consistently ranked as the best by all 8 subjects, while the result of Bernsen's as the worst. The proposed method is also among the best in OCR evaluation with only 1 error out of 827 characters.

The proposed method also performed well when significant Gaussian or speckle noise is present in the gray image (Fig. 3). Due to space limit, we do not show the individual binarization results of existing methods.

Table 1 Psychophysical and OCR evaluation results for existing and proposed adaptive binarization methods. For existing methods<sup>2-7)9)</sup>, the initials of the first authors are used to represent the corresponding methods, and IPC is the proposed method.

	JB	WN	JW	AA	WT	JS	IPC
Best votes	0	4	2	0	0	2	8
Worst votes	8	0	0	7	1	0	0
OCR errors*	20	5	0	28	3	1	1

\*: total number of characters is 827; errors are counted by characters.

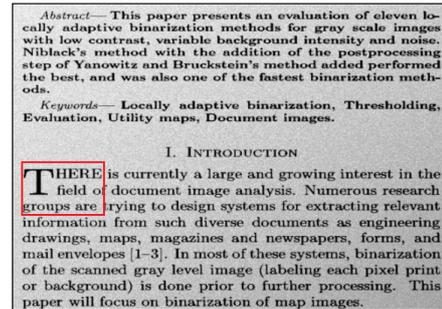
### 3.2 Real camera-captured document images

We further evaluated the proposed method on real camera-captured document images. One example is shown in Fig. 4. The original RGB image was captured with the camera on an HTC's Thunderbolt phone in office setting. The proposed method successfully binarized the entire image in the present of significant noise and non-uniform illumination, even for the "negative" (white-on-black) portion.

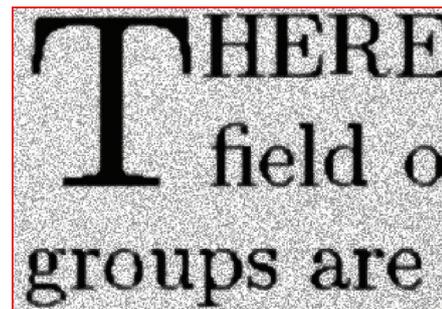
As pointed out in the Method section, during the

binarization process, we can also separate text and non-text regions of the document using the statistical information from the connected component analysis and some heuristic constraints derived from document characteristics. This topic warrants another full-length report to address. Interested readers may also refer to a United States patent application<sup>18)</sup>. Here we simply show one example of text and non-text segmentation without diving into the details (Fig. 5).

(a) Gray image



(b) Zoom-in of the red rectangle region in (a)



(c) Binarization result using IPC

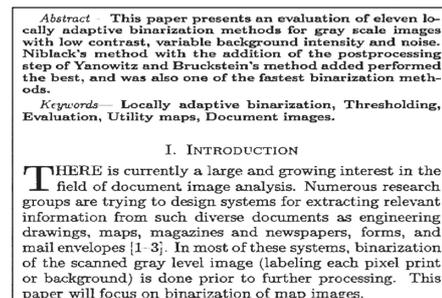


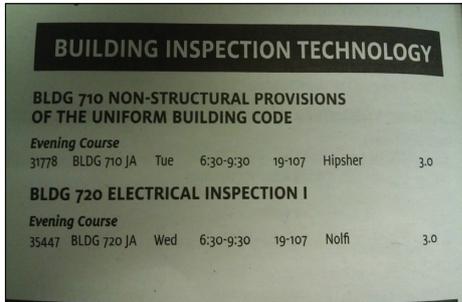
Fig. 3 Document image with simulated slanted non-illumination and Speckle noise and the binarization result using IPC.

## 4 Conclusions

An adaptive binarization method using phase congruency is proposed to deal with non-uniform illumination and take into account the unique characteristics of document images. It has been evaluated against 6 representative existing adaptive methods using both psychophysical and OCR methods, and emerged as one of the best in dealing with non-uniform

illumination and noise in both subjective and objective criteria. In addition, it has been demonstrated that this method can be extended to accomplish text and non-text segmentation as a by-product of the binarization process.

(a) Camera-captured image



(b) Binarization result using IPC

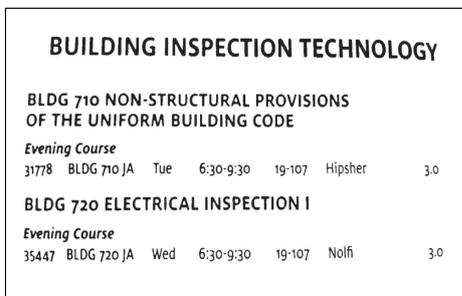


Fig. 4 A camera-captured document image and the binarization result using IPC.

(a) Camera-captured image



(b) Text segmentation result (white indicates text region)



Fig 5 A camera-captured image and the text segmentation result based on IPC and connected component analysis.

## References

- 1) J. Liang, D. Doermann, Hu. Li: Camera- based analysis of text and documents: a survey. *International Journal on Document Analysis and Recognition*, 7 (2005), 84-104,
- 2) Y. Tian, W. Ming: Adaptive deblurring for camera- based document image processing. *Lecture Notes in Computer Science*, 5876 (2009), 767-777 (Konica Minolta Technology Report, 8 (2011), 105- 110).
- 3) J. Bernsen: Dynamic thresholding of gray-level images. *The 8<sup>th</sup> International Conference on Pattern Recognition* (1986), 1251-1255.
- 4) W. Niblack: *Introduction to Digital Image Processing* (1986). Englewood Cliffs, NJ, Prentice Hall.
- 5) J.M. White, G.D. Rohrer.: Image thresholding for optical character recognition and other applications requiring character image extraction. *IBM Journal of Research and Development*, 27 (1983), 400-411.
- 6) A.S. Abutaleb: Automatic thresholding of gray-level pictures using two-dimensional entropy. *Computer Vision, Graphics and Image Processing*, 47 (1989), 22-32.
- 7) W.H. Tsai: Moment-preserving thresholding: a new approach. *Computer Vision, Graphics and Image Processing*, 29 (1985), 377-393.
- 8) B. Gatos, I. Pratikakis, J.P. Stavros: An adaptive binarization technique for low quality historical documents. *Lecture Notes in Computer Science*, 3163 (2004), 102-113.
- 9) J. Sauvola, M. Pietikainen: Adaptive document image binarization . *Pattern Recognition*, 33 (2000), 225-236.
- 10) M.C. Morrone, R.A. Owens: Feature detection from local energy. *Pattern Recognition Letter*, 6(1987), 303-313.
- 11) P. Kosevi: Image features from phase congruency. *Videre: Journal of Computer Vision Research*, 1 (1999), 1 – 26.
- 12) Y. Tian: Autofocus using image phase congruency. *Optics Express*, 19 (2011), 261-270.
- 13) Y. Tian, H. Feng, Z. Xu, S. Dai, Q. Li, M. Bian: A new autofocus technique based on analyzing the RGB components of color images. *Acta Photonica Sinica*, 31(2002), 363-366.
- 14) P. Perona, J. Malik: Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12 (1990), 629-639.
- 15) C. Tomasi, R. Manduchi: Bilateral filtering for gray and color images. *The 6<sup>th</sup> International Conference on Computer Vision* (1998), 839-846.
- 16) J. Canny: A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8 (1986), 679-698.
- 17) [www.gamera.sourceforge.net/html/binarization.html](http://www.gamera.sourceforge.net/html/binarization.html)
- 18) Y. Tian: Document image binarization and segmentation using image phase congruency. US patent application 13/607667 (2012).