IMPROVING DEGRADED ANCIENT DOCUMENT IMAGES USING PHASE-BASED BINARIZATION MODEL

S. KRISHNA KUMAR, K. BALA

1PG Scholar, Dept. of Electronic and communication Engineering, Srinivasa Institute of Technology and Science Kadapa, Andhra Pradesh, India

2Associate Professor and HOD. of Electronic and communication Engineering, Srinivasa Institute of Technology and Science, Kadapa, Andhra Pradesh, India

E-mail: Sunku98@gmail.com, Kbalu443@gmail.com

Abstract—Here presenting a phase-based binarization model for ancient document images, and also a post processing method which can improve any binarization method and a ground truth generation tool. Three feature maps derived from the phase information of an input document image form the core of this binarization model. These features are the maximum moment of phase corresponding to covariance, a locally weighted mean phase angle, and a phase preserved denoised image. The proposed model consists of three standard steps: 1) preprocessing; 2) main binarization; and 3) post processing. In the preprocessing and main binarization steps, the features used are mainly phase derived, while in the post processing step, specialized adaptive Gaussian and median filters are considered. One of the outputs of the binarization step, which shows high recall performance, is used in a proposed post processing method to improve the performance of other binarization methodologies. Finally, we develop a ground truth generation tool, called Phase ground truth, to simplify and speed up the ground truth generation process for ancient document images. The comprehensive experimental results on the DIBCO’09, H-DIBCO’10, DIBCO’11, H-DIBCO’12, DIBCO’13, PHIBD’12, and BICKLEY DIARY data sets show the robustness of the proposed binarization method on various types of degradation and document images.

Key words—Binarization Model; Phase-Derived Features, Phase Ground Truthing, Document Enhancement.

I. INTRODUCTION

Historically important documents manuscripts and stored Library materials. Like all these documents accumulate a significant amount of human heritage over time. However, many environmental factors, improper handling, and the poor quality of the materials used in their creation cause them a high degree of degradation of various types. Now there are many techniques for digitization of these manuscripts to preserve their content for future generations.

The large amount of digital data produced requires automatic processing, enhancement, and recognition. A main step in all document image processing is binarization, but this is not a very sophisticated process, which is unfortunate, as its performance has a significant influence on the quality of OCR results. Many researchers have done to solve the problems that arise in the binarization of old document images characterized by many types of degradation [1],[19], like faded ink, bleed-through, show-through, uneven illumination, variations in image contrast, and deterioration of the cellulose structure [1], [20]. There are also differences in patterns of hand-written and machine-printed documents, which add to the difficulties associated with binarization of old document images.

None of the proposed methods can deal with all types of documents and degradation. For more details, see the Related Work section. Fig. 1 shows some of the degraded document images used in this paper.

Fig 1: Sample document images selected from the DIBCO’09 [21], H-DIBCO’10 [22], and DIBCO’11 datasets [23].

Presently in this paper, a robust phase-based binarization method is proposed for the binarization and enhancement of historical documents and manuscripts. The three main steps in the proposed method are: preprocessing, main binarization, and post-processing. The preprocessing step mainly involves image denoising with phase preservation, by some morphological operations. We incorporate the canny edge detector and a denoised image to obtain a binarized image in rough form.

Then, the phase congruency features used [18], [19] for the main binarization. Phase congruency is widely used in the machine vision and image processing literature [16]; palm print verification, object detection, finger-knuckle-print recognition, and biomedical applications [16] are few examples of the
use of phase congruency as a feature detector. We show that foreground of ancient documents can be modeled by phase congruency. In previous works [18], [19] show that phase congruency is a robust way to process historical documents, both handwritten and machine-printed manuscripts. After completing the three binarization steps on the input images using phase congruency features and a denoised image, the enhancement processes are applied. A median filter and a phase congruency feature are used to construct an object exclusion map image. This map is also used to remove unwanted lines and interfering patterns.

Here proposed binarization method is stable and robust to various types of degradation and to different datasets, thanks to its purpose-designed steps, and provide comprehensive experimental results to demonstrate this robustness. The method outperforms most of the algorithms entered in the DIBCO’09 [9], H-DIBCO’10 [10], DIBCO’11 [11], H-DIBCO’12 [12], DIBCO’13 [13] and PHIBC’12 [14] competitions, based on various evaluation measures, including the F-measure, NRM, PSNR, DRD, and MPM.

The second contribution of this paper is proposal of a fast, semi-automatic tool for ground truth (GT) creation using robust phase-based feature maps. Ground truth generation for degraded document images is a difficult and time-consuming task, even for experts; however, benchmark datasets are required for the evaluation of binarization methods. Therefore, methods are developed to simplify and speed up ground truth generation. It is worth noting that ground truth creation tools and methods work in both semi-automatic and manual approaches[15]. We proposed a tool, called Phase ground truth, uses information provided by user as a priori information to produce a binarized output in rough form. Then, the selected regions in this output contain binarization errors. Phase ground truth offers alternatives for those regions, which are selected. The user can also use brush tools to manually label pixels of interest. This is a very appealing option, because it saves the user time and simplifies ground truth creation at the same time.

The contributions in this paper are following:

- A 5% improvement, on average, over our earlier binarization results [19], and considerable improvement over the state of the art.
- New capabilities added to ground truth generation tool, Phase ground truth [15], along with modifications to further simplify and accelerate the ground truth creation task.

In section II, we describe related work and in section III, Phase derived feature maps which are used in this paper is presented. In section IV, the flowchart of the proposed binarization model is presented, followed by a description of each step of the binarization and its impact. The ground truth generation tool, Phase ground truth, is described in section V. Section VI provides comprehensive experimental results of binarization method. Finally, section VII presents the conclusions and some directions for future research.

The notations used throughout this paper are listed below:

- $A_p(x)$ Local amplitude;
- $E_p(x)$ Real part of the complex-valued wavelet response;
- $E(A_p)$ Expected value of the Rayleigh distribution at scale $\rho$;
- Filter Log-Gabor filter;
- $I$ Gray-level input image;
- $I_{bwout}$ Final binarized output;
- $ICC$ Connected components of IOEM;
- $ID$ Denoised image;
- $ID, bw$ Binary image corresponding to $ID$;
- $ID, N, bw$ Binary image corresponding to normalized $ID$;
- $IL$ Local weighted mean phase angle (LWMPA);
- $IL, bw$ Binary image corresponding to $IL$;
- $IM$ Maximum moment of phase congruency covariance (MMPC);
- $IM, bw$ Binarized $IM$;
- $IM, F$ Filled image of $IM$;
- $IM, F, bw$ Binarized $IM,F$;
- $I_{Main}$ Output of the main binarization step;
- $I_{Med}$ Median filtered binary map of $I$;
- $IOEM$ Object exclusion map image;
- $I_{Otsu}, bw$ Otsu’s output when applied on $I$;
- $I_{Pre}$ Output of preprocessing step;
- $k$ Number of standard deviations of noises;
- $M^e_\rho$ Even symmetric wavelet at scale $\rho$;
- $M^o_\rho$ Odd symmetric wavelet at scale $\rho$;
- $O_p(x)$ Imaginary part of the complex-valued wavelet response;
- $PC_{1D}$ One-dimensional phase congruency;
- $PC_{2D}$ Two-dimensional phase congruency;
- $\rho$ Index over filter scales;
- $r$ Index over filter orientations;
- $N_p$ Number of filter scales;
- $N_r$ Number of filter orientations;
- $S(x)$ Spread function;
- $T$ Estimated noise threshold;
- $W(x)$ Weighting function;
- $\Phi_{\rho}(x)$ Local phase;
- $\Delta \Phi_{\rho}$ Phase deviation function;
- $\mu_R$ Mean of Rayleigh distribution;
- $\sigma_R$ Standard deviation of the Rayleigh distribution;
- $\sigma_G$ Parameter of the Rayleigh distribution.

II. RELATED WORK

In this, the related work described some selected
binarization methods. Gatos et al. [5] propose an adaptive binarization method based on low-pass filtering, foreground estimation, background surface computation, and a combination of these. An initial binary map is get using the multi-scale Sauvola’s method [1], and then statistical methods are used to recover the missed strokes and sub-strokes. In [8], Valizadehet al. map input images into a two-dimensional feature space in which foreground and background regions can be differentiated. Then, they partition this feature space into several small regions. Lu et al. [6] propose a binarization method based mainly on background estimation and stroke width estimation. First, the background of the document is estimated as a one-dimensional iterative Gaussian smoothing procedure. Then, for accurate binarization of strokes and sub-strokes, an L1-norm gradient image is used. This method placed in 43 algorithms submitted to the DIBCO’09 competition [9]. Su et al. [7] use local maximum and minimum to build a local contrast image. Then, a sliding window is applied to that image to determine local thresholds. Related to this method, algorithms entered in the H-DIBCO’10 contest [10]. In [2], a local contrast image is combined with a canny edge map to produce a more robust feature map. This method performs better than those in [6] and [7].

A multi-scale binarization method in which the input document is binarized several times using different scales in [1]. Then, these output images are combined to form the final output image. It uses different parameters for Sauvola’s method to produce output images at same size, but at different scales. In contrast, Lazzara propose a multi-scale Sauvola’s method which binarizes different scales of the input image with the same binarization parameters. Then, binary images with different scales are combined in other way to obtain the final results.

Combination methods also present a great deal of interest, and provided effective results. The approach of combining existing methods is to improve the output based on assumption that different methods complement one another. In [11], several of these methods are mixed based on a vote on the outputs of each. In [8], a combination of global and local adaptive binarization methods applied on an implanted image is used to binarize handwritten document images. This gives extremely well results, but it is limited to binarizing handwritten document images only.

In previous learning-based methods have also been proposed. These methods are use full to improve the outputs of other binarization methods based on a feature map [12],[14], or by finding the optimal parameters of binarization methods for each image [15], [16]. In [12] and [14], a self-training document binarization method is proposed. The input pixels, depending on the binarization method(s) used, are divided into three categories: foreground, background, and uncertain, based on a priori knowledge about the behavior of every method used. Then, foreground and background pixels are clustered into different classes using the k-means algorithm or random Markov field [12], [14]. Finally, uncertain pixels are classified with the label of their nearest neighboring cluster. The features used for the final decision are pixel intensity and local image contrast.

Another combined method based on a modified contrast feature is proposed in [13]. Lelore and Bouchara also classify pixels into three categories using a coarse thresholding method, where uncertain pixels are classified based on super resolution of likelihood of foreground. A method to optimize the global energy function based on a Laplacian image. In this process, a set of training images is used for optimization. In [15], Howe improved this method by tuning two parameters for each image. In [16], it is proposed to automatically determine the optimal parameters of any binarization method for each document image. After extracting the features and determining the optimal parameters, the relation between the features and the optimal parameters is known. As show in the Experimental Results and Discussion section, a problem associated with all these algorithms is that they are not reliable for all types of degradation and with different datasets.

In contrast to the considerable effort expanded on binarization methods, it has been paid to the development of ground truth creation tools. A semi-automatic method is proposed for document ground truthing. In this method, an initial binarized map is generated using an adaptive binarization method. Due to classification errors of the adaptive binarization method used, some errors remain in the initial map and consequently in the skeletonized image also. To remove these errors, a manual correction must need in this step. After computing the edges of texts using the Canny edge detector and manual modification of edge errors, a dilation operator based on this edge image is applied on the skeleton image to achieve a final binarized ground truth image. This method is a well known ground truth creation method, and an attempt to standardize the ground truth creation process. An application called PixLabeler is developed to help users create ground truth images. PixLabeler is GUI-based software which allows users to manually select individual foreground and background pixels. In case of highly degraded images and blurred images, users have difficulty selecting real edges. Both these ground truth creation methods are time consuming and require a great deal of manual selection and correction. In section V, we discuss the advantages of the proposed Phase ground truth tool over the currently available ground truth creation methods.

III. PHASE-DERIVED FEATURES

Three phase-derived feature maps of the input document image are used here. They are two phase
congruency feature maps and a denoised image.

A. Phase Congruency-Based Feature Maps

In this phase information is most important feature of images. The feature maps are extracted from input images and are based on the kovese’s phase congruency model. Another approach to the phase based processing of images can be the monogenic scale-space method.

In Phase congruency, the pixels of interest are at those points where the phase of the Fourier components is at its maximal. Let $M_x$ and $M_y$ denote that the even symmetries and odd symmetric log-Gabor wavelets at a scale $p$, which are known in literature as quadratic pairs. By considering $f(x)$ as a one-dimensional signal, the response of each quadratic pair of filters at each image point $x$ forms a response vector by convolving with $f(x)$:

$$[c(x),o(x)] = [f(x), f(x) *]$$ (1)

Where values $c(x)$ and $o(x)$ are real and imaginary parts of a complex-valued wavelet response. Now compute the local phase $\phi(x)$ and the local amplitude $A(x)$ of the transform at a given wavelet scale $p$

$$\phi(x) = \arctan2[\phi(x), c(x)]$$ (2)

$$A(x) = \sqrt{c(x)^2 + o(x)^2}$$ (3)

The fractional measure of spread $s(x)$ and phase congruency weighting mean function $W(x)$ are defined as:

$$s(x) = \frac{A(x)}{A_{\text{max}}(x)}$$ (4)

$$W(x) = \frac{1}{N}$$ (5)

where $N$ denotes the total number of filter scales; $A_{\text{max}}(x)$ denotes the amplitude of the filter pair with the maximum response; $W(x)$ is constructed by applying a sigmoid function to the filter response spread; $c$ is a cut-off value of the filter response spread below which phase congruency values are penalized; and $\gamma$ is a gain factor that controls the sharpness of the cutoff. $\rho$ which is a sensitive phase deviation function is defined as follows:

$$\phi(x) = \cos(\rho(x)) - \rho(x) \cdot [\sin(\rho(x)) - \rho(x)]$$ (6)

Where $\rho(x) = \phi(x)$ is the phase deviation at scale $p$; and $\theta(x)$ indicates mean phase angle. Let $PC_{1D}$ denote the one-dimensional phase congruency

$$PC_{1D}(x) = \frac{1}{s(x)}$$ (7)

Where $A(x)$ is the local amplitude at a given scale $p$ expresses the equality of the enclosed quantity to itself when its value is positive, and zero otherwise. This definition of $PC_{1D}$ is highly sensitive to noise. To overcome this issue, Rayleigh distribution can be used for modeling the distribution of noise energy:

$$\sigma_G$$ (8)

Where $\sigma_G$ denotes parameter of the Rayleigh distribution. The mean $\mu_R$, the standard deviation $\sigma_R$, and the median $R$ of the Rayleigh distribution can be expressed based on $\sigma_G$.

$$\mu_R = \frac{\sigma_G}{\sqrt{2}}$$ (9)

$$\sigma_R = \frac{\sigma_G}{\sqrt{2\pi}}$$ (10)

$$R = \frac{\sigma_G}{\sqrt{2}}$$ (11)

The median $R$, and all the other parameters of the Rayleigh distribution, can be estimated using the expected value of the magnitude response of the smallest filter scale:

$$\sigma_G = \frac{\sqrt{2\pi}}{\sqrt{2\pi}}$$ (12)

Consequently, $\mu_R$ and $\sigma_R$ can be computed.

In this, a noise threshold of the following form is used:

$$T = \mu_R + k\sigma_R$$ (13)

Where $k$ is number of $\sigma_R$ to be used. By applying the noise threshold in equation (7), we have

$$PC_{1D}(x) = \frac{1}{s(x)}$$ (14)

Where $T$ is the estimated noise modified by the Rayleigh distribution, equation (14). Here, the extension of the phase deviation function to two dimensions is presented by taking both the scale ($\rho$) and the orientation ($\theta$) indices of the wavelet coefficients:

$$\phi(x) = \cos(\rho(x)) - \rho(x) \cdot [\sin(\rho(x)) - \rho(x)]$$ (15)

Using equation (15), two-dimensional phase congruency is obtained by:

$$PC_{2D}(x) = \frac{1}{s(x)}$$ (16)

And the maximum moment of phase congruency covariance ($IM$) can be defined as:

$$\mu_{\text{re}} = \frac{\text{PC}_{2D}(x)}{\text{PC}_{2D}(x)}$$ (17)

The $IM$ map is a measure of edge strength. It takes values in the range $[0, 1]$, where a larger value means a stronger edge. The two-dimensional locally weighted mean phase angle ($IL$) is obtained using the summation of all filter responses over all possible orientations and scales:

$$\phi(x) = \arctan2[\phi(x), c(x)]$$ (18)

The $IL$ pixels of $IL$ are taken between $\pm \pi/2$ (a dark line) and $\pi/2$ (a bright line) that if we use the same definition of local phase as that of equation (2) for $IL$, the $IL$ values would be in the interval of $[0, \pi]$.

There are several parameters that are to be considered in the calculation of $IM$ and $IL$. Unfortunately, these parameters are set based on experiments in the literature. We also performed set of experiments to determine the best fixed values for computing $IM$. On the other, in which only the $IM$ feature of phase congruency is used [16], we used the $IL$ map as well, especially as it results in better binarization of strokes and sub-strokes.

B. Phase Preserving Denoising

An image de-noising method is proposed by kovese, which is based on assumption that phase information is the most important feature of images. It also attempts to preserve the perceptuality important phase information in the signal. It uses non-orthogonal, complex valued log-Gabor wavelets, which extract local phase and amplitude information at each point.
in the image. The denoising process contain a noise threshold at each scale and shrinking the magnitudes of filter response vector, while leaving the phase unchanged. Automatic estimation of these noise thresholds, using the statistics of the smallest filter scale response, is the very important part of denoising. These statistics are used to estimate the distribution the noise amplitude, because they give the strongest noise response. Then, the noise amplitude distribution of other can be estimated proportionally.

Here $\mu_R$ denotes the mean and $\sigma^2_R$ the variance of the Rayleigh distribution, the noise shrinkage threshold can be computed using equation (14). For each orientation, noise responses from the smallest scale filter pair are estimated and a noise threshold is obtained. This noise response distribution is used to estimate noise amplitude distribution of other filter scales using some constant. Finally, based on the noise thresholds obtained, the magnitudes of filter response vectors shrink, and they do so by soft thresholding, while leaving the phase unchanged. Fig. 2 shows two examples of IM, IL, and ID maps, where ID denotes the denoised image.

V. BINARIZATION MODEL

The binarized output is obtained by processing the input image in three steps: preprocessing, main binarization and post processing. In the binarization model a denoised image added, it is one of the feature of binarization model, and achieved 5% improvement, on average. The flow chart of the proposed binarization method is shown in Fig. 3.

A. Preprocessing

In this preprocessing step, A denoised image is used instead of the original image to obtain a binarized image in rough form. The image denoising method discussed in section II is applied to preprocess the binarization output. A number of parameters impact the quality of the denoised output image (ID), the key ones being the noise standard deviation threshold to be rejected ($k$), the number of filter scales ($N_p$) and the number of orientations ($N_r$) to be used. The $N_p$ parameter controls the extent to which low frequencies is covered. The higher $N_p$ is, the lower frequencies, which means that there, call value remains optimal or near optimal. Based on experiments, $N_p = 5$ is the appropriate choice in this case. Therefore, to preserve all the foreground pixels, set the parameters in the experiments as follows: $k = 1$, $N_p = 5$ and $N_r = 3$. By using Otsu’s method on the normalized denoised image, where normalized denoised image is obtained by applying a linear image transform on the denoised image. This approach will also remove noisy and degraded parts of images, because the denoising method attempts to shrink the amplitude information of the noise component. From Fig. 4(d) for the output of Otsu’s method when it is applied on a normalized denoised image. The problem with this approach is that it misses weak strokes and sub-strokes, which means that we cannot rely on its output. To solve this problem, we combine this binarized image with an
edge map obtained using the Canny operator. Canny operator is applied on the original document image and for combination those edges without any reference in the aforementioned binarized image are removed (Fig. 4(f)). After then compute a convex hull image of the combined image. Fig. 4 shows an example of this procedure. At the end of this step, the structure of foreground and text is determined. However, the image is still noisy, and the strokes and sub-strokes have not been accurately binarized. Also, the binarization output is affected by some types of degradation. We therefore include additional steps to deal with them.

Fig 4: Example of the steps used in the pre-processing phase of the proposed method. a) Denoised image. b) Normalized denoised image. c) Binarization of the original image using Otsu’s method. d) Binarization of the normalized denoised image using Otsu’s method. e) Edge image using the Canny operator. f) Combination of (d) and (e). g) Convex hull image of (f). h) Combination

B. Main Binarization

Main binarization, which is based on phase congruency features: i) the maximum moment of phase congruency covariance (IM); and ii) the locally weighted mean phase angle (IL). 1) IM: Here, IM is used to separate the background from potential foreground parts. It performs very well, even in badly degraded documents, where it can reject a majority of badly degraded background pixels by means of a noise modeling method. To achieve this, we set the number of two-dimensional log-Gabor filter scales $\rho$ to 2, and use 10 orientations of a two-dimensional log-Gabor filters $r$. In addition with this, the number of standard deviations $k$ used to reject noises is estimated as follows:

$$K = 2 \cdot \{\text{a} \cdot \left(\mu_{Otsu,bw} + \alpha \cdot I_{Otsu,bw} \right)\}$$

Where $\alpha$ is a constant (we are using $\alpha = 0.5$); $Otsu,bw$ is the binarization result of Otsu’s method of the input image; $IPre$ is the output of the preprocessing. Here, the minimum possible value for $k$ is 2. Fig. 5 shows output of $IM$ with and without using equation (20) to compute $k$.

Fig 5: The output of IM with a) fixed noise parameter [18], [19], and b) adaptive noise parameter estimation

Note the different values used for setting the phase congruency feature and denoised image parameters. Fig. 6 shows an example of how we use $IM$ to remove a majority of background pixels.

2) IL: Consider the following assumption in classifying foreground and background pixels using $IL$.

$$P(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot e^{-\frac{1}{2} \cdot \left(\frac{x - \mu}{\sigma}\right)^2}$$

Where $P(x)$ denoted one image pixel; and $Otsu,bw$ denotes the binarized image using Otsu’s method. Because of the parameters used to obtain the $IM$ and $IL$ maps, $IL$ produces some classification errors on the inner pixels of large foreground images. By using more filter scales would solve this problem, but reduce the performance of $IL$ on the strokes. Also, $IL$ impacts the quality of the $IM$ edge map, of course requires more computational time.

Nevertheless, the results of using Otsu’s method to binarize the large foreground objects are of interest. Consequently, we used the $Otsu,bw$ image to overcome the problem.

Fig 6: A degraded document image and its binarized image using phase Congruency. A) Original degraded document image. B) Edge image obtained by phase congruency (IM). C) Filled image of IM. D) Binarization of (c) using Otsu’s method. E) Denoised image and f) the result of main binarization. C. Post processing
In post processing step, applying the enhancement processes. First, a bleed through removal process is applied. Then, a Gaussian filter is used to further enhance the binarization output and to separate background from foreground, also an exclusion process is applied, based on a median filter and IM maps, to remove background noise and objects. Finally, a further enhancement process is applied to the denoised image. The individual steps are follows as.

1) Global Bleed-Through Exclusion: Bleed-through degradation is a common interfering pattern and a significant problem in old and historical document images. In this, bleed-through is categorized in two classes: i) local bleed through and ii) global bleed-through. Local bleed-through involves pixels located under or near foreground pixels, while global bleed-through involves pixels located far away the foreground pixels. Global bleed-through is one of most challenging forms of degradation, because there is no local to enable true text to be distinguished from bleed-through. In this stage, we found the possibility of the existence global bleed-through. If it does exist, the parameters of the Canny edge detector are chosen to ensure that the output edge map contains only edges of text regions which expect to be located in a specific part, or parts, of the image. The existence of bleed-through is established by comparing the Otsu’s result and binary output obtained so far [19]. If there is a noticeable difference between these two binary images, we apply a global bleed-through exclusion method. Fig. 7 provides two examples of the global bleed-through exclusion process.

![Fig 7: Effect of using the proposed global bleed-through exclusion is shown in column (c). The left image (b) is the binarized image before the global bleed-through exclusion step has been applied.](image)

2) Adaptive Gaussian Filter: In this, a similar approach one used in [17], except that a Gaussian smoothing filter is used to obtain a local weighted mean as the rotationally symmetric Gaussian low-pass filter \((G)\) of size \(S\) value, estimated based on average stroke-width, where \(\sigma\) is the standard deviation. This is a modification of the fixed \(S\) value used in [19]. The value for \(S\) is the most important parameter in this approach. Local thresholds can be computed using the following two-dimensional correlation.

\[
T(x, y) = \frac{1}{\pi \sigma^2} \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)
\]

Where \(I(x, y)\) is gray level of a input image. The result must be a filtered image \(T(x, y)\) which stores local thresholds. A pixel is set to 0 (dark) if the value of that pixel in the input image is less than 95% of the corresponding threshold value \(T(x, y)\) and it is set to 1 (white) otherwise. We increased the value 85% [17] to 95%, to obtain a near optimal recall value.

Some sub steps of proposed binarization method work on objects rather than on individual pixels, and so it is important to separate foreground text from the background. The Gaussian filter described above is one of the methods used to achieve this. This filter also applied to the equalized adaptive histogram image instead of the original image, in order to preserve weak strokes. The average stroke width is computed, order to set \(S\). There are various methods for computing stroke width [1], [3], [16]. In this paper, a very rapid approach, based on the reverse Euclidean distance transformation [48] of the rough binary image obtained so far, is used to estimate the average stroke width. This will depend on the quality of the rough binary image, which has the potential to produce errors; however, it is a very fast way to calculate stroke width, and provides a good estimate of the average width.

a) Document type detection: In this step, we have needed to determine the type of input document we are dealing with. To apply the enhancement processes that is after this step to the handwritten documents only and not to machine printed documents. The method we propose for detecting the type of document is straightforward and fast. We use the standard deviation of the orientation image that was produced during calculation of the phase congruency features. This image takes positive anticlockwise values between 0 and 180. A value of 90 corresponds to a horizontal edge, and a value of 0 indicates a vertical edge. By considering the foreground pixels of the output binary image obtained so far, we see that the standard deviation value of the orientations for these pixels is low for handwritten document images and higher for b-machined-printed documents. The reason for this is the different orientation values for interior pixels and edges. This approach works well for almost all the images we tested, including 21 machine-printed images and 60 handwritten document images, and only one classification error was found. It can be seen from the Fig.8 that the histogram of orientations of a handwritten document follows a U-shape behavior.
Improving Degraded Ancient Document Images Using Phase-Based Binarization Model

Note that even if this approach fails to accurately detect the type of document, it nevertheless produces satisfactory output.

4) Majority Criterion: Here a majority criterion based on a denoised image, ID. A majority criterion supposes that early binarization steps provide an optimal or near optimal recall value. Then, based on the fact that a foreground pixel should have a lower value than its adjacent background pixels, exclusion over the foreground pixels is performed. A majority criterion works as follows. For each foreground pixel in Ibwout, its 5 × 5 adjacent background pixels in ID are checked, and that pixel is removed from the foreground if its value in ID is less than that of any of the 5×5 background values in ID. The criterion works very well on noise, unwanted lines, and strokes and sub-strokes. Algorithm 1 provides the pseudo code of proposed binarization method.

Algorithm 1 The Pseudo Code of the Proposed Binarization Method. Note: Foreground and Background Pixels Take the Values of “1” and “0”, Respectively (BW10 Representation [1]). Also, ∧ (∨) Denotes Pixel-Wise AND (OR), Respectively.

1. PROEDURE PHASE-BASED BINARIZATION(I) START
2. Calculate ID, ID bw, ID N, bw, IC, IOEM;
3. ICH=Convex Hull (ID, N, bw, IC);
4. IPre=ID bw ∧ ICH;
5. CALCULATE IM, IL, IL bw, IM F, IM F bw;
6. ITemp1=IPre ∧ IM F,bw;
7. ITemp2=IL bw v IOtsu,bw;
8. imain=ITemp1 ∧ ITemp2;
9. Calculate IG, Imed, IM bw, IOEM;
10. If (global Bleed-Through) Then
11. Apply Global Bleed Through Removing Algorithm;
12. end If
13. Ibwout=Imain ∧ IM;
14. IOEM=Imed ∧ IM bw;
15. Label Each Connected Component In Ibwout(Icc1….N)
16. For (i=1 To N) Do
17. If (ICci ∈ IOEM) Then
18. Ibwout=Ibwout−Icc;
19. End If
20. End For
21. If (handwritten) then
22. Apply Enhancement Criteria On Ibwout;
23. end if
24. return ibwout
25. end procedure

We propose an object exclusion map image (IOEM) based on a combination of median filter and a binary map of IM (see Algorithm 1). Any object without a reference in this binary map will be removed from the final binarization results. This approach can remove, local bleed-through, and interfering patterns. It is known that a median filter can reject salt-and-pepper noise in the presence of edges. Like the method used in the previous section for the Gaussian filter, local thresholds computed by applying an S × S symmetric median filter each pixel in the input image. The value for S is estimated on the average stroke width, instead of taking a fixed value as in [19]. In turn, a filtered image equal in size to the input image is produced (Med), where its pixel values are local thresholds. A pixel is set to 0 (dark) if the value of that pixel in the input image is less than 90% of corresponding pixel value in Med, and it is set to 1 (white) otherwise it is called IMed.

V. PHASEGT: AN EFFICIENT GROUND TRUTHING TOOL

We introduce a ground truthing application; called Phase ground truth, use it to generate ground truth images. Phase ground truth is mainly based on phase congruency features and a phase-denoised image. Fig. 9 shows how Phase ground truth produces ground truth images in three steps:
Improving Degraded Ancient Document Images Using Phase-Based Binarization Model

Fig 9: Flowchart of the proposed ground truthing tool (PhaseGT).

i) The user provides *a priori* information about the input document image. This will include the type of image example: machine-printed, handwritten, or a combination of the two. The user then asked to choose degradation types from a list. If this is not possible, Phase ground truth works in automatic mode to generate intermediate output.

ii) Phase preprocesses the input image and generates an intermediate ground truth binary image. The objective in this step is to save human interaction time.

iii) The user makes final modifications to the intermediate ground truth image. For example, can change pixels using brushing tools, also select a portion of the intermediate binary image that contains binarization errors, and choose from a number of replacement options offered by Phase ground truth. This can save the user time by avoiding the need for manual corrections. Also, an optional edge map is provided by Phase ground truth to help users choose the real edges of texts.

A. Document Type
A preprocessing step based on *a priori* information about the input document image is included by to produce an intermediate binary image. The *a priori* information is the type of document. Usually, the patterns of the texts of handwritten and machine printed documents are different. In handwritten documents, strokes and sub-strokes are the key parts, and so the preprocessing step should preserve them. In machine-printed documents, the inner parts of large texts in the document image should be preserved.

B. Degradation Types
Another piece of *a priori* information is degradation type. In this type the user can select two degradation types for foreground text: i) nebulous text, and ii) weak strokes and/or sub-strokes. More degradation options are available for background, which can select. These are global bleed-through, local bleed-through, unwanted lines or patterns, and alien ink. There are other types of degradation as well, and we plan to continue to develop Phase ground truth to enable it. For example, in the case of an ancient book in which the degradation types on its pages are similar, parameters should be tuned so, that Phase ground truth produces the overall output. In this way, ground truthing could be achieved more rapidly. Of course, this approach can also extend to binarization methods when they are applied on an archive of similar documents. As can be expected, earlier work on the detection of degradation types in documents has confirmed that manual selection of degradation types is more accurate than automatic detection of degradation. Phase ground truth can deal with noise based on *a priori* information provided and can currently detect the following degradation types: local and global bleed-through, unwanted lines and patterns, alien ink and faded ink. Phase ground truth uses the same features as used in the binarization method proposed in section IV, however, the parameters of each feature map is tuned based on the *a priori* information provided by the user.

C. Phase ground truth in Practice
In this step, the Phase ground truth tool is used to develop the PHIBD 2012 dataset, and is also tested on other datasets. Begin with one degraded document image from DIBCO'09[9] and another from PHIBD'12, as shown in Fig.10 (a). The *a priori* information for the first image is as follows: a handwritten document, weak strokes and sub-strokes, and faded ink. For the second image, the *a priori* information provided is as follows: a handwritten document, faded ink, and unwanted lines and patterns. The intermediate ground truth produced by Phase ground truth is shown in Fig. 10(b), where the edges are shown in red and other pixels are in blue. The white pixels are background pixels. It takes about five hours to produce a ground truth for the image H03 from DIBCO’09 using the Pixel Labeler tool, while it takes less than a 20 minutes when image is first processed using Phase ground truth and predefined patches.
Improving Degraded Ancient Document Images Using Phase-Based Binarization Model

Consider a machine-printed image from DIBCO’11, assuming the following a priori information: machine-printed, nebulous text. The binary image produced by Phase ground truth is shown in Fig. 11. For the handwritten document image in Fig. 11, the a priori information is the following: handwritten, weak strokes/sub-strokes, local bleed-through, and global bleed-through. The binary image produced by Phase ground truth is also shown in Fig. 11.

Here applied Phase ground truth on DIBCO’09 document images to produce their ground truth images. As expected, and has been shown earlier [52], different methods and different individuals will produce different ground truth images. This shows the need for standard criteria, or tools to be proposed or developed, so that different individuals can develop the same, or approximately the same, ground truth results. One solution could be to use different feature maps and methods based on a priori information provided by a human. Our current version of Phase ground truth is a primitive attempt to develop such a tool, which is also easier for the ground truther to use. We are working to expand and modify Phase ground truth in future work, with a view to addressing this aspect of the tool.

Although the term “ground truth” is usually understood to be an objective matter, it is highly subjective in practice. A clear example is the presence of various “ground truth” images for the DIBCO’09 database. We believe these efforts would eventually converge to a more objective ground truth by considering and developing more abstract and precise definition of the ground truth in image binarization. Although there will be no unique ground truth even when using the same definition because the final verification would be still subjectively carried out by human, the difference would be marginal in contrast to the case of today’s state-of-the-art ground truthing.

Fig. 12 shows a portion of a historical document from the DIBCO’09 and the ground truth provided by DIBCO series Also, a ground truth image from BSU and one obtained using Phase ground truth are shown. It is clear that there are many differences between these three ground truth images. For example, we can see that the ground truth image in Fig. 12(c) appears to be a dilated image of the others. This indicates that different users will produce different ground truth images for a single input image. However, user fatigue and screen resolution could also lead to different ground truth images being produced, even by the same user.

There are three aspects of our proposed ground truth creation method (Phase ground truth) that are noteworthy:

i) The near-ground truth intermediate binary image produced based on the user’s selection of degradation types, which is the main contribution
of the proposed method. The advantage of this feature is that it needs less additional interaction on the part of the user, which means that he can generate near-ground truth images for a very large dataset in a short time. It should be noted that, by using IL, we have all the strokes, even very weak ones, on the intermediate binarized output (pseudo-Recall>99%). See Figs. 10(b), 11(b), and 11(d). (This is one of the main differences between our proposed method and that proposed in [36].)

ii) The option that we give the user of using a Canny edge map or of selecting some of the edges by hand. (This is another difference between the proposed method and that proposed)

iii) The alternative patches that we offer to the user for at those locations where errors are found, in order to reduce manual correction effort. These alternative patches also are generated based on a set of predefined parameters and a priori information provided edges constitute the most time-consuming part that need to be maintained by the user, we used different parameters to generate different local weighted mean phase angle IL outputs that user can use and replace. These outputs are generated by changing the parameter $N_p$, which is the number of filter scales. Fig. 10 shows examples of these alternative patches.

**EXPERIMENTAL RESULTS**

The proposed binarization method is evaluated on a number of datasets. The following datasets are used DIBCO’09 [9], DIBCO’10 [10], DIBCO’11 [11], H-DIBCO’12 [12], [13], PHIBD’12 [14], and BICKLEYDIARY [18]. These datasets provide a collection of images which are suffered different types of degradation, and which give enough information and sufficiently challenging in terms of evaluation setup to enable a meaningful examination of various algorithms.

First, compare the subjective and objective performance of the proposed method with that of leading binarization methods in the literature. Then, compared proposed binarization method with state-of-the-art algorithms and the top ranking algorithm in each competition.

**A. Subjective Evaluation**

In this section, we compare outputs of the proposed method with those of top-placing methods in each contest, whenever possible. Our proposed method performs a smooth binarization of the document images, thanks to the use of phase congruency measures and a denoised image. In Fig. 13, we compare the proposed method with three top-placing algorithms in DIBCO’11 [11], the winning algorithm in DIBCO’09 [6], [9], and the method proposed in [10].

**B. Objective Evaluation**

Here used a well-known measures F-measure (FM), pseudo F-measure (p-FM), PSNR, distance reciprocal distortion (DRD) metric, the misclassification penalty metric (MPM), the negative rate metric (NRM) to evaluate various algorithms [9], [11]. The results of the method are compared with state-of-the-art binarization methods. Tables I-VI shows the evaluation results obtained the proposed method and state-of-the-art binarization algorithms the DIBCO’09 [9], H-DIBDO’10 [10], DIBCO’11 [11], H-DIBCO’12 [12], DIBCO’13 [13], PHIBD’12 [15], BICKLEY DIARY datasets respectively. PHIBD’12 is a dataset of historical Persian images consisting of 15 degraded document images. The result shows that the proposed method achieved, on average, a 5% improvement over our earlier results [19]. It has been seen from these experimental results that other binarization methods produce different results for different datasets whereas there is a small difference between the results we obtained using the proposed method on different datasets, which shows the robustness of proposed method. The upper bound performances of the proposed method when parameters of the proposed binarization method have been tuned are listed in Tables I-VII. Also, the best value from each contest is provided in the last row of Tables I-VI. For having a fair comparison in Table VI, the best results for each measure is not highlighted because GTs of PHIBD’12 are generated using phase-congruency features and this might boosted the performance of the proposed method.

In this paper, the standard deviation of the numerical results is considered to measure the reliability of the
various methods compared. The results in Tables I-V show that the proposed algorithm shows the lowest variation among all the methods in terms of FM variations.

<table>
<thead>
<tr>
<th>TABLE I</th>
<th>COMPARISON OF THE PERFORMANCE OF THE PROPOSED METHOD AND THE OTHERS AGAINST DIBCO’09 DATASET</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
<td>Algorithm</td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>Proposed</td>
<td>93.52 ± 2.34</td>
</tr>
<tr>
<td>Proposed (upper bound)</td>
<td>93.36 ± 1.93</td>
</tr>
<tr>
<td>BE [9]</td>
<td>91.13 ± 4.80</td>
</tr>
<tr>
<td>LMM [10]</td>
<td>90.06 ± 2.93</td>
</tr>
<tr>
<td>Gb Saurola [1]</td>
<td>89.76 ± 0.07</td>
</tr>
<tr>
<td>Saurola Mx_m [40]</td>
<td>76.85 ± 20.27</td>
</tr>
<tr>
<td>AddOut [3]</td>
<td>91.55 ± 5.09</td>
</tr>
<tr>
<td>PC [10]</td>
<td>86.37 ± 4.35</td>
</tr>
<tr>
<td>LMM+ [2]</td>
<td>93.5</td>
</tr>
<tr>
<td>1st rank of contest</td>
<td>91.13 ± 4.80</td>
</tr>
<tr>
<td>Best results of contest</td>
<td>91.13 ± 4.80</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE II</th>
<th>COMPARISON OF THE PERFORMANCE OF THE PROPOSED METHOD AND THE OTHERS AGAINST H-DIBCO’10 DATASET</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
<td>Algorithm</td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>Proposed</td>
<td>91.00 ± 1.67</td>
</tr>
<tr>
<td>Proposed (upper bound)</td>
<td>90.72 ± 1.54</td>
</tr>
<tr>
<td>BE [9]</td>
<td>87.10 ± 4.99</td>
</tr>
<tr>
<td>LMM [10]</td>
<td>85.49 ± 4.63</td>
</tr>
<tr>
<td>Gb Saurola [1]</td>
<td>87.22 ± 0.43</td>
</tr>
<tr>
<td>Saurola Mx_m [40]</td>
<td>80.03 ± 12.26</td>
</tr>
<tr>
<td>AddOut [3]</td>
<td>86.01 ± 4.81</td>
</tr>
<tr>
<td>PC [10]</td>
<td>85.32 ± 2.40</td>
</tr>
<tr>
<td>Letre [41]</td>
<td>91.78</td>
</tr>
<tr>
<td>LMM+ [2]</td>
<td>92.03</td>
</tr>
<tr>
<td>1st top rank of contest</td>
<td>-</td>
</tr>
<tr>
<td>2nd top rank of contest</td>
<td>-</td>
</tr>
<tr>
<td>Best results of contest</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE III</th>
<th>COMPARISON OF THE PERFORMANCE OF THE PROPOSED METHOD AND THE OTHERS AGAINST DIBCO’11 DATASET</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
<td>Algorithm</td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>Proposed</td>
<td>91.42 ± 2.99</td>
</tr>
<tr>
<td>Proposed (upper bound)</td>
<td>92.57 ± 2.34</td>
</tr>
<tr>
<td>BE [9]</td>
<td>81.67 ± 13.98</td>
</tr>
<tr>
<td>LMM [10]</td>
<td>85.56 ± 6.16</td>
</tr>
<tr>
<td>Gb Saurola [1]</td>
<td>84.15 ± 10.13</td>
</tr>
<tr>
<td>Saurola Mx_m [40]</td>
<td>79.70 ± 12.09</td>
</tr>
<tr>
<td>PC [10]</td>
<td>85.74 ± 4.99</td>
</tr>
<tr>
<td>LMM+ [2]</td>
<td>87.7</td>
</tr>
<tr>
<td>Gates et. al [5]</td>
<td>84.3</td>
</tr>
<tr>
<td>Howe et. al [15]</td>
<td>89.2</td>
</tr>
<tr>
<td>1st rank of context [43]</td>
<td>80.86 ± 29.11</td>
</tr>
<tr>
<td>Best results of contest</td>
<td>88.72 ± 7.28</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE IV</th>
<th>COMPARISON OF THE PERFORMANCE OF THE PROPOSED METHOD AND THE OTHERS AGAINST DIBCO’11 DATASET</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
<td>Algorithm</td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>Proposed</td>
<td>92.14 ± 2.12</td>
</tr>
<tr>
<td>Proposed (upper bound)</td>
<td>93.01 ± 1.87</td>
</tr>
<tr>
<td>Mb Gb Saurola [1]</td>
<td>87.54 ± 4.52</td>
</tr>
<tr>
<td>Saurola Mx_m [40]</td>
<td>81.19 ± 8.69</td>
</tr>
<tr>
<td>AddOut [3]</td>
<td>87.73 ± 5.42</td>
</tr>
<tr>
<td>PC [10]</td>
<td>87.21 ± 5.55</td>
</tr>
<tr>
<td>Letre [41]</td>
<td>92.88 ± 10.03</td>
</tr>
<tr>
<td>1st rank of contest</td>
<td>88.76 ± 25.31</td>
</tr>
<tr>
<td>Best results of contest</td>
<td>92.68</td>
</tr>
</tbody>
</table>

C. Enhancement of Other Binarization Methods
Preprocessing and main binarization in the proposed method are used as a mask to cross out false positive pixels on the output of other binarization methods, which resulted in an high level of improvement. This mask has a high recall with an acceptable precision value. Table VIII shows the improvement achieved over other binarization methods using the proposed mask. Compared with previous works, which are retimed at modifying other binarization methods, the proposed method shows even more improvement. For example, the improved F-Measure values of Otsu’s method for DIBCO’09, H-DIBCO’10, and DIBCO’11 are 81.98, 87.13, and 83.55 respectively.
Our improved results are 89.82, 86.49, and 88.91 respectively.

**TABLE VIII**

<table>
<thead>
<tr>
<th>F-M</th>
<th>Improved-FM</th>
<th>DIBCO’09</th>
<th>H-DIBCO’10</th>
<th>DIBCO’12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Otsu [9]</td>
<td>78.60</td>
<td>80.82</td>
<td>85.45</td>
<td>86.49</td>
</tr>
<tr>
<td>Niblack [9]</td>
<td>52.30</td>
<td>59.50</td>
<td>45.66</td>
<td>45.82</td>
</tr>
<tr>
<td>Grid-based Sauvola [11]</td>
<td>88.17</td>
<td>90.75</td>
<td>84.65</td>
<td>85.05</td>
</tr>
<tr>
<td>Multi-scale Grid-based Sauvola [1]</td>
<td>89.36</td>
<td>90.34</td>
<td>85.52</td>
<td>86.67</td>
</tr>
<tr>
<td>ESBRK [3]</td>
<td>90.16</td>
<td>96.16</td>
<td>82.62</td>
<td>87.03</td>
</tr>
<tr>
<td>Sua’s method [10]</td>
<td>91.06</td>
<td>96.33</td>
<td>85.49</td>
<td>86.61</td>
</tr>
<tr>
<td>La’s method [9]</td>
<td>91.13</td>
<td>92.32</td>
<td>87.10</td>
<td>87.92</td>
</tr>
</tbody>
</table>

**D. Time Complexity**

In this process, we evaluate the run time of proposed method, performing an experiment on a Core i7 3.4 GHz CPU with 8 GB of RAM. The algorithm is implemented in MATLAB 2012 running on Windows 7. It takes 2.04 seconds to operate it on a 0.3 megapixel image, and 20.28 seconds to produce output for a 3 megapixels image. It is worth mentioning that proposed algorithm would run faster and would require much less memory if the phase congruency features are calculated using the alternative monogenic filters.

**CONCLUSION AND FUTURE PROSPECTS**

In proposed paper, we have introduced an image binarization method that uses the phase information of the input image robust phase-based features extracted from image are used to build a model for the binarization of ancient manuscripts. Phase-preserving denoising followed by morphological operations is used to preprocess input image. Then, two phase congruency features maximum moment of phase congruency covariance and the locally weighted mean phase angle are used to perform the main binarization. For post-processing, here we have proposed a few steps to filter various types of degradation in particular a median filter has used to reject noise, unwanted lines, and interfering patterns. Because some binarization steps works with individual objects instead of pixels, a Gaussian filter used to further separate foreground from background objects to improve the final binary output. The method has been tested on various datasets covering numerous types of degradation H-DIBCO’10, DIBCO’11, H-DIBCO’12, PHBD’12 and BICKLEY DIARY. The experimental results demonstrate its promising performance, and also that of the post processing method proposed to improve other binarization algorithms.

Also proposed a rapid method to find the type of document image been verified, which will be of great interest. The behavior of ancient handwritten document images and machine-printed images shows differences in terms of binarization. The strokes and sub-strokes of handwritten images require accurate binarization, and the binarization of the interior pixels of the text of machine-printed images needs to be performed carefully. Although the proposed binarization method works better on both handwritten and machine-printed documents, better results for both types of documents are achieved, when a priori information about the type of input document is available.

Finally, an efficient ground truthing tool called Phase ground truth has been provided for degraded documents. This is designed to reduce the manual correction involved in ground truth generation.

In future work, we plan to expand the application of phase-derived features, which ensures the stable behavior of document images, to other cultural heritage fields, such as microfilm analysis and multispectral imaging.

**REFERENCES**


