Phase-based binarization of ancient document images: Model and applications
Hossein Ziaei Nafchi, Reza Farrahi Moghaddam Member, IEEE., and Mohamed Cheriet, Senior Member, IEEE.

Abstract—In this paper, a phase-based binarization model for ancient document images is proposed, as well as a post-processing method that can improve any binarization method and a ground truth generation tool. Three feature maps derived from the phase information of an input document image constitute the core of this binarization model. These features are: the maximum moment of phase congruency covariance, a locally weighted mean phase angle, and a phase preserved denoised image. The proposed model consists of three standard steps: preprocessing, main binarization, and postprocessing. In the preprocessing and main binarization steps, the features used are mainly phase-derived, while in the postprocessing 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 postprocessing method to improve the performance of other binarization methodologies.

Finally, we develop a ground truth generation tool, called PhaseGT, 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 datasets show the robustness of the proposed binarization method on various types of degradation and document images.

Index Terms—Historical document binarization, Phase-derived features, Ground truthing, Document enhancement.

I. INTRODUCTION

LIBRARIES and archives around the world store an abundance of old and historically important documents and manuscripts. 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 to suffer a high degree of degradation of various types. Today, there is a strong move toward digitization of these manuscripts to preserve their content for future generations. The huge amount of digital data produced requires automatic processing, enhancement, and recognition. A key step in all document image processing workflows 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 research studies have been carried out to solve the problems that arise in the binarization of old document images characterized by many types of degradation [1]–[19], including 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 the binarization of old document images.

To the best of our knowledge, 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.

![Sample document images from DIBCO'09, H-DIBCO'10, DIBCO'11 datasets.](image)

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

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 [24], followed by some morphological operations. We incorporate the Canny edge detector [25] and a denoised image to obtain a binarized image in rough form.

Then, we use the phase congruency features [18], [19], [26] for the main binarization step. Phase congruency is widely used in the machine vision and image processing literature [27]–[30]; palmprint verification [27], object detection [28], finger-knuckle-print recognition [29], and biomedical applications [30] are just a few examples of the use of phase congruency as a feature detector. We show that the foreground of
ancient documents can be modeled by phase congruency. Our
previous works [18], [19], [31] 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 [24], 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 then used to remove unwanted lines and interfering
patterns. The effect of each step on the binarized output image
is discussed in each associated section.

The proposed binarization method is stable and robust to
various types of degradation and to different datasets, thanks to
its purpose-designed steps, and we provide comprehensive ex-
perimental results to demonstrate this robustness. The method
outperforms most of the algorithms entered in the DIBCO’09
[21], H-DIBCO’10 [22], DIBCO’11 [23], H-DIBCO’12 [32],
DIBCO’13 [33] and PHIBC’12 [34] competitions, based on
various evaluation measures, including the F-measure, NRM,
PSNR, DRD, and MPM.

The second contribution of this paper is the proposal of a
fast, semi-automatic tool for ground truth (GT) creation using
robust phase-based feature maps. GT 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 and
tools must be developed to simplify and speed up GT creation.
It is worth noting that GT creation tools and methods work
in both semi-automatic and manual approaches [35]–[38]. We
discuss these approaches in the Related Work section. The tool
we propose, called PhaseGT, uses information provided by the
user as a priori information to produce a binarized output
in rough form. Then, the user selects regions in this output
that contain binarization errors. PhaseGT offers alternatives
for those regions, which the user selects. 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 GT creation at the same time. Given our previous
works on this topic, the contributions of this paper are the
following:

- A 5% improvement, on average, over our earlier bina-
  rization results [19], and considerable improvement over
  the state of the art.
- New capabilities added to ground truth generation tool,
  PhaseGT [38], along with modifications to further sim-
  plify and accelerate the ground truth creation task.

The rest of the paper is organized as follows. In section
II, we describe related work. In section III, we present the
phase-derived feature maps used in this paper. In section IV,
the flowchart of the proposed binarization model is presented,
followed by a description of each step of the binarization chain
and a discussion of its impact. Our ground truth generation
tool, PhaseGT, is described in section V. Section VI provides
comprehensive experimental results of our binarization meth-
od. Finally, section VII presents our conclusions and some
directions for future research.

The notations used throughout the paper are listed below:

- \( e_p(x) \) Real part of the complex-valued wavelet response;
- \( E(A_p) \) Expected value of the Rayleigh distribution at scale \( \rho \);
- \( I \) Gray-level input image;
- \( I_{bw} \) Final binarized output;
- \( I_{CC} \) Connected components of \( I_{OEM} \);
- \( I_D \) Denoised image;
- \( I_{D,bw} \) Binary image corresponding to \( I_D \);
- \( I_{D,N,bw} \) Binary image corresponding to normalized \( I_D \);
- \( I_L \) Local weighted mean phase angle (LWMPA);
- \( I_{bw} \) Binary image corresponding to \( I_L \);
- \( I_M \) Maximum moment of phase congruency covariance
  (MMPCC);
- \( I_{M,bw} \) Binarized \( I_M \);
- \( I_{M,F} \) Filled image of \( I_M \);
- \( I_{M,F,bw} \) Binarized \( I_{M,F} \);
- \( I_{Main} \) Output of the main binarization step;
- \( I_{Med} \) Median filtered binary map of \( I \);
- \( I_{OEM} \) 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;
- \( M_p \) Even symmetric wavelet at scale \( \rho \);
- \( M_p^o \) Odd symmetric wavelet at scale \( \rho \);
- \( \alpha_p(x) \) Imaginary part of the complex-valued wavelet response;
- \( PC_{1D} \) One-dimensional phase congruency;
- \( PC_{2D} \) Two-dimensional phase congruency;
- \( r \) Index over filter orientations;
- \( N_p \) Number of filter scales;
- \( N_r \) Number of filter orientations;
- \( s(x) \) Spread function;
- \( \phi_p(x) \) Local phase;
- \( \Delta \Phi_p \) 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 section, we briefly describe some selected binariza-
tion 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.
In [6], an initial binary map is obtained using the multi-

dimensional Sauvola’s method [1], and then statistical methods are
used to restore the missed strokes and sub-strokes. In [8],
Valizadeh et al. map input images into a two-dimensional
feature space in which the foreground and background regions
can be distinguished. Then, they partition this feature space
into several small regions, which are classified into text and
background based on the results of applying Niblack’s method
[39].

Lu et al. [9] propose a binarization method based mainly
on background estimation and stroke width estimation. First,
the background of the document is estimated by means of
a one-dimensional iterative Gaussian smoothing procedure.
Then, for accurate binarization of strokes and sub-strokes, an
\( L_1 \)-norm gradient image is used. This method placed 1st of
43 algorithms submitted to the DIBCO’09 competition [21].
Su et al. [10] use local maximum and minimum to build
a local contrast image. Then, a sliding window is applied
across that image to determine local thresholds. A version of
this method shared 1st place with another method, out of 17 algorithms entered in the H-DIBCO’10 contest [22]. 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 [9], [10].

Farrahi Moghaddam et al. [1] propose a multi-scale binarization method in which the input document is binarized several times using different scales. Then, these output images are combined to form the final output image. This method uses different parameters for Sauvola’s method to produce output images of the same size, but at different scales. In contrast, Lazzara and Gerard [40] 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 some way to produce the final results.

Combination methods have also attracted a great deal of interest, and provided promising results. The goal 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 combined based on a vote on the outputs of each. In [7], a combination of global and local adaptive binarization methods applied on an inpainted image is used to binarize handwritten document images. The results show that this method performs extremely well; however, it is limited to binarizing handwritten document images only.

Learning-based methods have also been proposed in recent years. These methods are an attempt to improve the outputs of other binarization methods based on a feature map [12]–[14], or by determining the optimal parameters of binarization methods for each image [15], [16]. In [12], [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 the 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. In [13], another combined method based on a modified contrast feature is proposed. Lelore and Bouchara [41] also classify pixels into three categories using a coarse thresholding method, where uncertain pixels are classified based on super resolution of likelihood of foreground. Howe [17] proposes a method to optimize the global energy function based on a Laplacian image. In this method, a set of training images is used for optimization. In [15], Howe improved this method by tuning two key parameters for each image. In [16], a learning framework 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 learned. As we 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, little attention has been paid to the development of GT creation tools. In [36], a semi-automatic method is proposed for document ground truthing. In that method, an initial binarized map is generated using an adaptive binarization method. Then, a skeleton image of this map is computed. Due to classification errors of the adaptive binarization method used, some errors remain in the initial map and consequently in the skeletonized image as well. To remove these errors, a manual correction is included in this step. After computing the edges of texts using the Canny edge detector [25] 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 GT creation method, and an attempt to standardize the GT creation process. In [37], 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 the case of highly degraded images and blurred images, users have difficulty selecting real edges. Both these GT creation methods are time-consuming and require a great deal of manual selection and correction. We discuss the advantages of our proposed PhaseGT tool over the currently available GT creation methods in section V.

III. PHASE-DERIVED FEATURES

We use three phase-derived feature maps of the input document image in this paper: two phase congruency feature maps and a denoised image. The details are provided below.

A. Phase congruency-based feature maps

In [42], it is shown that the phase information of an image outweighs its magnitude information. This implicitly means that phase information is the most important feature of images. In this section, two phase congruency-based feature maps extracted from input images are discussed. These feature maps are based on the Kovaci’s phase congruency model [26]. Another approach to the phase-based processing of images could be the monogenic scale-space method of [43]. However, based on our experiments, Kovaci’s method worked better within our proposed binarization method.

In phase congruency, the pixels of interest are at those points where the phase of the Fourier components is at its maximal [26], [44]. Let \( M^r_\rho \) and \( M^o_\rho \) denote the even symmetric and odd symmetric log-Gabor wavelets at a scale \( \rho \), which are known in the literature as quadratic pairs [45]. By considering \( f(x) \) as a one-dimensional signal, the response of each quadrature pair of filters at each image point \( x \) forms a response vector by convolving with \( f(x) \):

\[
\left[ e_\rho(x), o_\rho(x) \right] = \left[ f(x) \ast M^r_\rho, f(x) \ast M^o_\rho \right].
\]

(1)

where values \( e_\rho(x) \) and \( o_\rho(x) \) are real and imaginary parts of a complex-valued wavelet response. We can now compute the local phase \( \phi_\rho(x) \) and the local amplitude \( A_\rho(x) \) of the transform at a given wavelet scale \( \rho \):

\[
\phi_\rho(x) = \arctan2 \left( o_\rho(x), e_\rho(x) \right),
\]

(2)
The fractional measure of spread \( s(x) \) and phase congruency weighting mean function \( W(x) \) are defined as follows:

\[
s(x) = \frac{1}{N} \left( \frac{\sum_{\rho} A_{\rho}(x)}{A_{\mathrm{max}}(x)} \right) ,
\]

\[
W(x) = \frac{1}{1 + e^{e^{-s(x)}}} ,
\]

where \( N \) denotes the total number of filter scales; \( A_{\mathrm{max}}(x) \) 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. \( \Delta \Phi_{\rho} \), which is a sensitive phase deviation function, is defined as follows:

\[
\Delta \Phi_{\rho} = \cos \left( \phi_{\rho}(x) - \bar{\phi}(x) \right) - \left| \sin \left( \phi_{\rho}(x) - \bar{\phi}(x) \right) \right| .
\]

where \( \phi_{\rho}(x) - \bar{\phi}(x) \) is the phase deviation at scale \( \rho \); and \( \bar{\phi}(x) \) indicates the mean phase angle. Let \( PC_{1D} \) denote the one-dimensional phase congruency:

\[
PC_{1D}(x) = \frac{\sum_{\rho} W(x) |A_{\rho}(x)\Delta \Phi_{\rho}(x)|}{\sum_{\rho} A_{\rho}(x)} ,
\]

where \( A_{\rho}(x) \) is the local amplitude at a given scale \( \rho \) and \([\cdot]\) 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:

\[
R(x) = \frac{x}{\sigma_G} \exp \frac{-x^2}{2\sigma_G^2} ,
\]

where \( \sigma_G \) denotes the parameter of the Rayleigh distribution. The mean \( \mu_R \), the standard deviation \( \sigma_R^2 \), and the median \( \tilde{R} \) of the Rayleigh distribution can be expressed based on \( \sigma_G \):

\[
\mu_R = \sigma_G \sqrt{\frac{\pi}{2}} ,
\]

\[
\sigma_R^2 = \left( 2 - \frac{\pi}{2} \right) \sigma_G^2 ,
\]

\[
\tilde{R} = \sigma_G \sqrt{\ln(4)} .
\]

The median \( \tilde{R} \), 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:

\[
E(A_{\rho_{\text{min}}}) = \frac{\tilde{R}}{2} \sqrt{\frac{\pi}{\ln(2)}} ,
\]

which results in an estimation of \( \sigma_G \):

\[
\sigma_G = \frac{E(A_{\rho_{\text{min}}})}{\sqrt{\pi/2}} .
\]

Consequently, \( \mu_R \) and \( \sigma_R \) can be computed.

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

\[
T = \mu_R + k \sigma_R ,
\]

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

\[
PC_{1D}(x) = \frac{\sum_{\rho} W(x) |A_{\rho}(x)\Delta \Phi_{\rho}(x) - T|}{\sum_{\rho} A_{\rho}(x)} .
\]

where \( T \) is the estimated noise modeled by the Rayleigh distribution, equation (14).

Here, the extension of the phase deviation function to two dimensions is presented by considering both the scale \( \rho \) and the orientation \( \gamma \) indices of the wavelet coefficients:

\[
\Delta \Phi_{\rho\gamma}(x) = \cos \left( \phi_{\rho\gamma}(x) - \bar{\phi}(x) \right) - \left| \sin \left( \phi_{\rho\gamma}(x) - \bar{\phi}(x) \right) \right| .
\]

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

\[
PC_{2D,r}(x) = \frac{\sum_{\rho} W_r(x) |A_{\rho}(x)\Delta \Phi_{\rho}(x) - T_r|}{\sum_{\rho} A_{\rho}(x)} .
\]

and the maximum moment of phase congruency covariance \( I_M \) can be defined as:

\[
I_M = \max_r PC_{2D,r}(x) .
\]

The \( I_M \) 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 \( I_L \) is obtained using the summation of all filter responses over all possible orientations and scales:

\[
I_L(x) = \arctan \left( \frac{\sum_{\rho, r} e_{\rho r}(x) \sum_{\rho} a_{\rho r}(x)}{\sum_{\rho, r} \rho r} \right) .
\]

The pixels of \( I_L \) take values between \(-\pi/2 \) (a dark line) and \(+\pi/2 \) (a bright line). It is worth mentioning that if we use the same definition of local phase as that of equation (2) for \( I_L \), the \( I_L \) values would be in the interval of \([0, \pi]\).

There are several parameters to be considered in the calculation of \( I_M \) and \( I_L \). Unfortunately, these parameters are set based on experiments in the literature. We have also performed a set of experiments to determine the best fixed values for computing \( I_M \). Unlike other research, in which only the \( I_M \) feature of phase congruency is used \([28], [30], [46]\), we used the \( I_L \) map as well, especially as it results in better binarization of strokes and sub-strokes.

### B. Phase preserving denoising

An image denoising method proposed by Kovesi \([24]\) is used in this paper, which is based on the assumption that phase information is the most important feature of images. This method also attempts to preserve the perceptually important phase information in the signal. It uses non-orthogonal, complex valued log-Gabor wavelets, which extract the local phase and amplitude information at each point in the image. The denoising process consists of determining a noise threshold at each scale and shrinking the magnitudes of the filter response vector appropriately, while leaving the phase unchanged. Automatic estimation of these noise thresholds, using the statistics of the smallest filter scale response, is the most important part of denoising. These statistics are used to estimate the distribution

\[
A_{\rho}(x) = \sqrt{e_{\rho}(x)^2 + a_{\rho}(x)^2} .
\]
of the noise amplitude, because they give the strongest noise response. Then, the noise amplitude distribution of other filter scales can be estimated proportionally.

Supposing that \( \mu_R \) denotes the mean and \( \sigma^2_R \) denotes 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 the noise amplitude distribution of other filter scales using some constant. Finally, based on the noise thresholds obtained, the magnitudes of the filter response vectors shrink appropriately, and they do so by soft thresholding, while leaving the phase unchanged. Fig. 2 shows two examples of \( I_M \), \( I_L \), and \( I_D \) maps, where \( I_D \) denotes the denoised image.

![Examples of IM, IL, and ID maps](image)

Fig. 2. Two examples of \( I_M \), \( I_L \), and \( I_D \) maps.

IV. BINARIZATION MODEL

The final binarized output image is obtained by processing the input image in three steps: preprocessing, main binarization, and postprocessing. The binarization model is an extended version of the one proposed in our previous work [19]. We have added a denoised image, which is another phase-based feature to the binarization model, and achieved 5% improvement, on average. The flowchart of the proposed binarization method is shown in Fig. 3. Each step is discussed individually in the subsections below.

A. Preprocessing

In the preprocessing step, we use a denoised image instead of the original image to obtain a binarized image in rough form. The image denoising method discussed in section III is applied to preprocess the binarization output. A number of parameters impact the quality of the denoised output image \( (I_D) \), the key ones being the noise standard deviation threshold to be rejected \( (k) \), and the number of filter scales \( (N_\rho) \) and the number of orientations \( (N_r) \) to be used. The \( N_\rho \) parameter controls the extent to which low frequencies are covered. The higher \( N_\rho \) is, the lower the frequencies, which means that the recall value remains optimal or near optimal. Based on our experiments, \( N_\rho = 5 \) is the appropriate choice in this case. Therefore, to preserve all the foreground pixels, we set the parameters in the experiments as follows: \( k = 1 \), \( N_\rho = 5 \) and \( N_r = 3 \).

We used 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 can also remove noisy and degraded parts of images, because the denoising method attempts to shrink the amplitude information of the noise component. See 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 [25]. 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)). We 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 of images (a) and (g).

B. Main binarization

The next step is the main binarization, which is based on phase congruency features: i) the maximum moment of phase congruency covariance ($I_M$); and ii) the locally weighted mean phase angle ($I_L$).

1) $I_M$: In this paper, $I_M$ is used to separate the background from potential foreground parts. This step 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 two-dimensional log-Gabor filters $r$. In addition, the number of standard deviations $k$ used to reject noises is estimated as follows:

$$k = 2 + \left[ \alpha \times \left( \frac{\sum_{n,m} I_{\text{Otsu,bw}}(n,m)}{\sum_{n,m} I_{\text{Pre}}(n,m)} \right) \right],$$

(20)

where $\alpha$ is a constant (we are using $\alpha = 0.5$); $I_{\text{Otsu,bw}}$ is the binarization result of Otsu’s method on the input image; and $I_{\text{Pre}}$ is the output of the preprocessing step. Here, the minimum possible value for $k$ is 2. Fig. 5 shows the output of $I_M$ with and without using equation (20) to compute $k$. Note the different values used for setting the phase congruency feature and denoised image parameters.

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

Fig. 6 shows an example of how we use $I_M$ to remove a majority of the background pixels.

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 ($I_M$). c) Filled image of $I_M$. d) Binarization of (c) using Otsu’s method. e) Denoised image and f) The result of main binarization.

2) $I_L$: We consider the following assumption in classifying foreground and background pixels using $I_L$:

$$P(x) = \begin{cases} 1, & I_L(x) \leq 0 \\ 0, & I_L(x) > 0 \land I_{\text{Otsu,bw}}(x) = 0 \end{cases},$$

(21)

where $P(x)$ denotes one image pixel; and $I_{\text{Otsu,bw}}$ denotes the binarized image using Otsu’s method. Because of the parameters used to obtain the $I_M$ and $I_L$ maps, $I_L$ produces some classification errors on the inner pixels of large foreground objects. Using more filter scales would solve this problem, but reduce the performance of $I_L$ on the strokes. Also, $I_L$ impacts the quality of the $I_M$ edge map, and 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 $I_{\text{Otsu,bw}}$ image to overcome the problem.
C. Postprocessing

In this step, we apply 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, and an exclusion process is applied, based on a median filter and $I_M$ maps, to remove background noise and objects. Finally, a further enhancement process is applied to the denoised image. The individual steps are as follows.

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 paper, 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 text. 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.

At this stage, we investigate the possibility of the existence of 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 the edges of text regions which we expect to be located in a specific part, or parts, of the image. The existence of bleed-through is established by comparing the Otus’s result and the 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 section, we take a similar approach to the one used in [47], except that a Gaussian smoothing filter is used to obtain a local weighted mean as the reference value for setting the threshold for each pixel. We use a rotationally symmetric Gaussian low-pass filter ($G$) of size $S$ with $\sigma$ 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:

$$T(x, y) = \sum_{i=-S}^{S} \sum_{j=-S}^{S} G(i, j) \times I(x + i, y + j) ,$$

where $I(x, y)$ is a gray-level input image. The result is 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 from 85% [47] to 95%, in order to obtain a near optimal recall value.

Some sub steps of the 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 is 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, in 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 approach is dependent 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 stroke width.

a) Document type detection: At this step, we need to determine the type of input document we are dealing with. We propose to apply the enhancement processes that are 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 machine-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. Note that even if this approach fails to accurately detect the type of document, it nevertheless produces satisfactory output.

b) Object exclusion map image ($I_{OBJ}$): We construct an object exclusion map image ($I_{OBJ}$) based on a combination of a median filter and a binary map of $I_M$ (see Algorithm 1). Any object without a reference in this binary map will be removed from the final binarization results. This approach can remove noise, local bleed-through, and interfering patterns.

It is known that a median filter can reject salt-and-pepper
noise in the presence of edges [49]. Like the method used in the previous section for the Gaussian filter, local thresholds are computed by applying an \( S \times S \) symmetric median filter for each pixel in the input image. The value for \( S \) is estimated based 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 \((M_{ed})\), 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 \( M_{ed} \), and it is set to 1 (white) otherwise. The output is called \( I_{M_{ed}} \).

4) Majority criterion: We propose a majority criterion based on a denoised image, \( I_{D} \). 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 \( I_{bwout} \), its \( 5 \times 5 \) adjacent background pixels in \( I_{D} \) are checked, and that pixel is removed from the foreground if its value in \( I_{D} \) is less than that of any of the \( 5 \) background values in \( I_{D} \). This criterion works very well on noise, unwanted lines, and strokes and sub-strokes. Algorithm 1 provides the pseudo code of the 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, \( \wedge (\lor) \) denotes pixel-wise AND(OR), respectively.

\begin{verbatim}
1: procedure PHASE-BASED BINARIZATION(I) Start
2: Calculate \( I_{D}, I_{D,bw}, I_{D,N,bw}, I_{C}, I_{DEM} \);
3: \( I_{CH} = \text{ConvexHull}(I_{D,N,bw}, I_{C}) \); 
4: \( I_{Pre} = I_{D,bw} \wedge I_{CH} \);
5: Calculate \( I_{M}, I_{L}, I_{M,F}, I_{M,F,bw} \);
6: \( I_{Temp1} = I_{Pre} \wedge I_{M,F,bw} \);
7: \( I_{Temp2} = I_{L,bw} \lor I_{bw} \);
8: \( I_{Main} = I_{Temp1} \wedge I_{Temp2} \);
9: Calculate \( I_{G}, I_{Med}, I_{M,bw}, I_{OEM} \);
10: if (global bleed-through) then
11: Apply global bleed-through removing algorithm;
12: end if
13: \( I_{bwout} = I_{Main} \wedge I_{G} \);
14: \( I_{DEM} = I_{Med} \wedge I_{M,bw} \);
15: Label each connected component in \( I_{bwout} \) \((I_{CC1..N})\) 
16: for \((i=1 \text{ to } N)\) do
17: if (\( \overline{I_{CC_i}} \in I_{DEM} \)) then
18: \( I_{bwout} = I_{bwout} - I_{CC_i} \);
19: end if
20: end for
21: if (handwritten) then
22: Apply enhancement criteria on \( I_{bwout} \);
23: end if
24: return \( I_{bwout} \)
\end{verbatim}

V. PHASEGT: AN EFFICIENT GROUND TRUTHING TOOL

We introduce a ground truthing application, called PhaseGT, in this paper, and use it to generate ground truth images. PhaseGT is mainly based on phase congruency features and a phase-denoised image. Fig. 9 shows how PhaseGT produces ground truth images in three steps:

i) The user provides \textit{a priori} information about the input document image. This information includes the type of image, e.g. machine-printed, handwritten, or a combination of the two. The user is then asked to choose degradation types from a list. If this is not possible, PhaseGT works in automatic mode to generate this intermediate output.

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

![Flowchart of the proposed ground truthing tool (PhaseGT).](image-url)
A. Document type

As mentioned above, a preprocessing step based on a priori information about the input document image is included by PhaseGT 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 paper, 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 the user 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 PhaseGT to enable it to handle these in the near future. For example, in the case of an ancient book in which the degradation types on its pages are similar, the parameters should be tuned so that PhaseGT produces the best overall output. In this way, ground truthing could be achieved more rapidly. Of course, this approach can be also extended 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 [50], [51] in documents has confirmed that manual selection of degradation types is more accurate than automatic detection of degradation. PhaseGT can deal with noise based on a priori information provided by the user, and can currently detect the following degradation types: local and global bleed-through, unwanted lines and patterns, alien ink and faded ink.

PhaseGT 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. PhaseGT in practice

In this section, the PhaseGT tool is used to develop the PHIBD’12 dataset, and is also tested on other datasets. We begin with one degraded document image from DIBCO’09 [21] 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 GT produced by PhaseGT is shown in Fig. 10(b), where the edges are shown in red and other pixels are shown in blue. The white pixels are background pixels. Based on [52], it takes about five hours to produce a ground truth for the image H03 from DIBCO’09 using the PixLabeler tool, while it takes less than a 20 minutes when image is first processed using PhaseGT and predefined patches.

Consider a machine-printed image from DIBCO’11, assuming the following a priori information: machine-printed, nebulous text. The binary image produced by PhaseGT 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, global bleed-through. The binary image produced by PhaseGT is also shown in Fig. 11.

We applied PhaseGT 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

1Available on the IAPR TC-11 website and also on the Synchromedia website: http://www.synchromedia.ca/PHIBD2012

Fig. 10. Two examples showing how PhaseGT produces a ground truth image. a) Two original document images, one from DIBCO’09 and one from PHIBD’12. b) A processed image using PhaseGT, where red pixels indicate edges. c) The image in (b) after manual correction. Also, misclassified patches and suggested patches by PhaseGT are shown, where X shows misclassified patch in the preprocessing step and ✓ shows typical selected patch by user.
could be to use different feature maps and methods based on \textit{a priori} information provided by a human. Our current version of PhaseGT 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 PhaseGT 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 organizers. Also, a ground truth image from BSU [52] and one obtained using PhaseGT are shown. It is clear that there are many differences between these three ground truth images. For example, we can see that the GT image in Fig. 12(c) appears to be a dilated image of the others. This indicates that different users will produce different GT images for a single input document image. However, user fatigue and screen resolution could also lead to different GT images being produced, even by the same user.

There are three aspects of our proposed GT creation method (PhaseGT) that are noteworthy:

i) The near-GT intermediate binary image produced based on \textit{a priori} information provided by a human. 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-GT images for a very large dataset in a short time. It should be noted that, by using \( I_L \), we have all the strokes, even very weak ones, on the intermediate binarized output (pseudo-Recall>99%).

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 in [36].)

iii) The alternative patches that we offer to the user for replacement 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 \textit{a priori} information provided. Since 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 \( L \), 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.

Fig. 12. Example showing three ground truths of a historical document image from DIBCO’09. a) The original document image. b) A GT produced by the DIBCO’09 organizers. c) A GT produced by BSU [52]. d) A GT produced using PhaseGT.

VI. EXPERIMENTAL RESULTS AND DISCUSSION

The proposed binarization method is evaluated on a number of datasets. The following datasets were used: DIBCO’09 [21], H-DIBCO’10 [22], DIBCO’11 [23], H-DIBCO’12 [32], DIBCO’13 [33], PHIBD’12 [38], and BICKLEY DIARY [53]. These datasets provide a collection of images that have suffered different types of degradation, and which give enough information and are sufficiently challenging in terms of evaluation setup to enable a meaningful examination of various algorithms.

First, we compare the subjective and objective performance of the proposed method with that of leading binarization methods in the literature. Then, we compared our 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 [23], the winning algorithm in DIBCO’09 [9], [21], and the method proposed in [10].

B. Objective evaluation

We used the well-known measures F-measure (FM), pseudo F-measure (p-FM), PSNR, distance reciprocal distortion (DRD) metric [54], the misclassification penalty metric (MPM), the negative rate metric (NRM) to evaluate various algorithms [21]–[23]. The source code of the evaluation measures used in this paper is available in [55]. The results of the proposed method are compared with state-of-the-art binarization methods. Tables I-VI show the evaluation results obtained by the proposed method and state-of-the-art binarization algorithms for the DIBCO’09 [21], H-DIBDO’10 [22], DIBCO’11 [23], H-DIBCO’12 [32], DIBCO’13 [33], PHIBD’12 [38], and BICKLEY DIARY [53] datasets respectively. PHIBD’12 is a dataset of historical Persian images consisting of 15 degraded document images. The results show that the proposed method achieved, on average, a 5% improvement over our earlier results [19]. It can be seen from these experimental results that other binarization methods produce different results for different datasets, whereas there is little difference between the results we obtained using the proposed method on different datasets, which shows the robustness of our 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 we compared. The results in Tables I-V show that the proposed algorithm shows the lowest variation among the methods in terms of FM variations.

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 outstanding level of improvement. This mask has a high recall value with an acceptable precision value. Table VII shows the improvement we achieved over other binarization methods using the proposed mask. Compared with previous works [12]–[14], which were aimed at modifying other binarization

<table>
<thead>
<tr>
<th>Proposition</th>
<th>FM±std</th>
<th>FM±</th>
<th>PSNR</th>
<th>NRM</th>
<th>MPM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed</td>
<td>92.52±2.34</td>
<td>92.58</td>
<td>19.00</td>
<td>3.39</td>
<td>0.54</td>
</tr>
<tr>
<td>Proposed (upper bound)</td>
<td>91.36±1.93</td>
<td>93.39</td>
<td>19.55</td>
<td>3.52</td>
<td>0.25</td>
</tr>
<tr>
<td>BE [9]</td>
<td>91.13±4.80</td>
<td>91.23</td>
<td>18.65</td>
<td>4.30</td>
<td>0.54</td>
</tr>
<tr>
<td>LMM [10]</td>
<td>91.06±2.93</td>
<td>91.18</td>
<td>18.50</td>
<td>6.54</td>
<td>0.45</td>
</tr>
<tr>
<td>Ms Gb Sauvola [1]</td>
<td>89.26±6.07</td>
<td>89.64</td>
<td>17.76</td>
<td>3.78</td>
<td>1.18</td>
</tr>
<tr>
<td>Sauvola Ms≥4 [40]</td>
<td>76.85±20.27</td>
<td>80.14</td>
<td>14.52</td>
<td>4.71</td>
<td>8.95</td>
</tr>
<tr>
<td>AdOtnu [3]</td>
<td>91.55±3.69</td>
<td>91.61</td>
<td>18.80</td>
<td>5.29</td>
<td>0.65</td>
</tr>
<tr>
<td>PC [19]</td>
<td>88.43±3.35</td>
<td>88.62</td>
<td>17.03</td>
<td>4.36</td>
<td>1.06</td>
</tr>
<tr>
<td>LMM+ [2]</td>
<td>93.53</td>
<td>19.65</td>
<td>3.74</td>
<td>0.43</td>
<td></td>
</tr>
<tr>
<td>1st rank of contest</td>
<td>91.13±4.80</td>
<td>91.23</td>
<td>18.65</td>
<td>4.30</td>
<td>0.54</td>
</tr>
<tr>
<td>Best results of contest</td>
<td>91.13±4.80</td>
<td>91.23</td>
<td>18.66</td>
<td>4.18</td>
<td>0.52</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Proposition</th>
<th>FM±std</th>
<th>FM±</th>
<th>PSNR</th>
<th>NRM</th>
<th>MPM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed</td>
<td>91.00±1.67</td>
<td>91.09</td>
<td>93.67</td>
<td>19.34</td>
<td>0.32</td>
</tr>
<tr>
<td>Proposed (upper bound)</td>
<td>91.18±1.54</td>
<td>91.82</td>
<td>95.08</td>
<td>19.80</td>
<td>0.20</td>
</tr>
<tr>
<td>BE [9]</td>
<td>87.10±4.99</td>
<td>87.53</td>
<td>87.87</td>
<td>18.16</td>
<td>1.04</td>
</tr>
<tr>
<td>LMM [10]</td>
<td>85.89±6.53</td>
<td>85.75</td>
<td>92.55</td>
<td>17.83</td>
<td>0.49</td>
</tr>
<tr>
<td>Ms Gb Sauvola [1]</td>
<td>85.52±3.49</td>
<td>85.89</td>
<td>90.32</td>
<td>17.62</td>
<td>0.68</td>
</tr>
<tr>
<td>Sauvola Ms≥4 [40]</td>
<td>80.03±9.26</td>
<td>82.36</td>
<td>87.07</td>
<td>16.35</td>
<td>3.41</td>
</tr>
<tr>
<td>AdOtnu [3]</td>
<td>86.01±4.81</td>
<td>86.25</td>
<td>91.63</td>
<td>18.04</td>
<td>0.20</td>
</tr>
<tr>
<td>PC [19]</td>
<td>88.52±2.40</td>
<td>88.69</td>
<td>90.37</td>
<td>18.08</td>
<td>0.03</td>
</tr>
<tr>
<td>Lelore [41]</td>
<td>91.78</td>
<td>94.44</td>
<td>19.04</td>
<td>1.33</td>
<td></td>
</tr>
<tr>
<td>LMM+ [2]</td>
<td>92.03</td>
<td>94.85</td>
<td>20.12</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>1st rank of contest</td>
<td>91.30</td>
<td>93.58</td>
<td>19.78</td>
<td>0.49</td>
<td></td>
</tr>
<tr>
<td>Best results of contest</td>
<td>89.70</td>
<td>95.15</td>
<td>19.15</td>
<td>0.28</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Proposition</th>
<th>FM±std</th>
<th>FM±</th>
<th>PSNR</th>
<th>NRM</th>
<th>MPM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed</td>
<td>91.42±2.99</td>
<td>91.56</td>
<td>18.68</td>
<td>0.91</td>
<td>2.74</td>
</tr>
<tr>
<td>Proposed (upper bound)</td>
<td>92.57±2.34</td>
<td>92.62</td>
<td>19.29</td>
<td>0.55</td>
<td>2.28</td>
</tr>
<tr>
<td>LMM [10]</td>
<td>85.56±16.16</td>
<td>86.23</td>
<td>94.45</td>
<td>16.01</td>
<td>4.05</td>
</tr>
<tr>
<td>Ms Gb Sauvola [1]</td>
<td>84.15±9.13</td>
<td>85.50</td>
<td>16.32</td>
<td>10.25</td>
<td>7.00</td>
</tr>
<tr>
<td>Sauvola Ms≥4 [40]</td>
<td>79.70±12.69</td>
<td>81.78</td>
<td>14.91</td>
<td>20.32</td>
<td>11.67</td>
</tr>
<tr>
<td>AdOtnu [3]</td>
<td>86.12±4.85</td>
<td>86.58</td>
<td>16.88</td>
<td>2.58</td>
<td>4.36</td>
</tr>
<tr>
<td>PC [19]</td>
<td>87.74±3.49</td>
<td>88.18</td>
<td>17.02</td>
<td>1.91</td>
<td>4.59</td>
</tr>
<tr>
<td>LMM+ [2]</td>
<td>89.77</td>
<td>94.57</td>
<td>5.17</td>
<td>4.48</td>
<td></td>
</tr>
<tr>
<td>Gatos et al. [5]</td>
<td>84.3</td>
<td>16.3</td>
<td>6.39</td>
<td>6.08</td>
<td></td>
</tr>
<tr>
<td>Howe (static) [15]</td>
<td>89.2</td>
<td>18.2</td>
<td>9.05</td>
<td>5.76</td>
<td></td>
</tr>
<tr>
<td>1st rank of contest</td>
<td>80.86±29.11</td>
<td>86.19</td>
<td>16.13</td>
<td>64.36</td>
<td>104.47</td>
</tr>
<tr>
<td>Best results of contest</td>
<td>88.72±7.28</td>
<td>89.42</td>
<td>17.84</td>
<td>8.61</td>
<td>5.36</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Proposition</th>
<th>FM±std</th>
<th>FM±</th>
<th>PSNR</th>
<th>NRM</th>
<th>MPM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed</td>
<td>90.41±5.36</td>
<td>90.99</td>
<td>19.44</td>
<td>2.08</td>
<td>3.47</td>
</tr>
<tr>
<td>Proposed (upper bound)</td>
<td>92.78±3.38</td>
<td>92.92</td>
<td>20.38</td>
<td>0.47</td>
<td>2.35</td>
</tr>
<tr>
<td>BE [9]</td>
<td>87.07±12.36</td>
<td>88.84</td>
<td>18.75</td>
<td>3.20</td>
<td>4.26</td>
</tr>
<tr>
<td>LMM [10]</td>
<td>84.77±14.25</td>
<td>87.07</td>
<td>18.21</td>
<td>2.00</td>
<td>4.98</td>
</tr>
<tr>
<td>Ms Gb Sauvola [1]</td>
<td>86.98±7.08</td>
<td>87.53</td>
<td>17.62</td>
<td>3.24</td>
<td>3.15</td>
</tr>
<tr>
<td>Howe (static) [15]</td>
<td>90.50±10.24</td>
<td>91.09</td>
<td>20.88</td>
<td>3.89</td>
<td>3.83</td>
</tr>
<tr>
<td>1st rank of contest</td>
<td>91.26±8.81</td>
<td>92.21</td>
<td>20.67</td>
<td>2.39</td>
<td>3.10</td>
</tr>
<tr>
<td>Best results of contest</td>
<td>91.34±12.39</td>
<td>92.69</td>
<td>21.29</td>
<td>2.39</td>
<td>3.10</td>
</tr>
</tbody>
</table>
methods, our proposed method shows even more improvement. For example, in [14], 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.

D. Time complexity

In this section, we evaluate the run time of our proposed method, performing our experiments on a Core i7 3.4 GHz CPU with 8 GB of RAM. The algorithm is implemented in MATLAB 2012a 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 the proposed algorithm would run faster and would require much less memory if the phase congruency features are calculated using the alternative monogenic filters of [43].

VII. CONCLUSION AND FUTURE PROSPECTS

In this paper, we have introduced an image binarization method that uses the phase information of the input image, and robust phase-based features extracted from that image are used to build a model for the binarization of ancient manuscripts. Phase-preserving denoising followed by morphological operations are used to preprocess the input image. Then, two phase congruency features, the maximum moment of phase congruency covariance and the locally weighted mean phase angle, are used to perform the main binarization. For post-processing, we have proposed a few steps to filter various types of degradation, in particular, a median filter has been used to reject noise, unwanted lines, and interfering patterns. Because some binarization steps work with individual objects instead of pixels, a Gaussian filter was used to further separate foreground from background objects, and to improve the final binary output. The method has been tested on various datasets covering numerous types of degradation: DIBCO’09, H-DIBCO’10, DIBCO’11, H-DIBCO’12, PHIBD’12 and BICKLEY DIARY. Our experimental results demonstrate its promising performance, and also that of the postprocessing method proposed to improve other binarization algorithms.

We have also proposed a rapid method to determine the type of document image been studied, 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 with care. Although the proposed binarization method works well 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 PhaseGT has been provided for degraded documents. This tool 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.

ACKNOWLEDGMENTS

The authors thank the NSERC of Canada for their financial support. The authors also thank the anonymous reviewers for their valuable and insightful comments and suggestions to improve the quality of the paper.

REFERENCES

