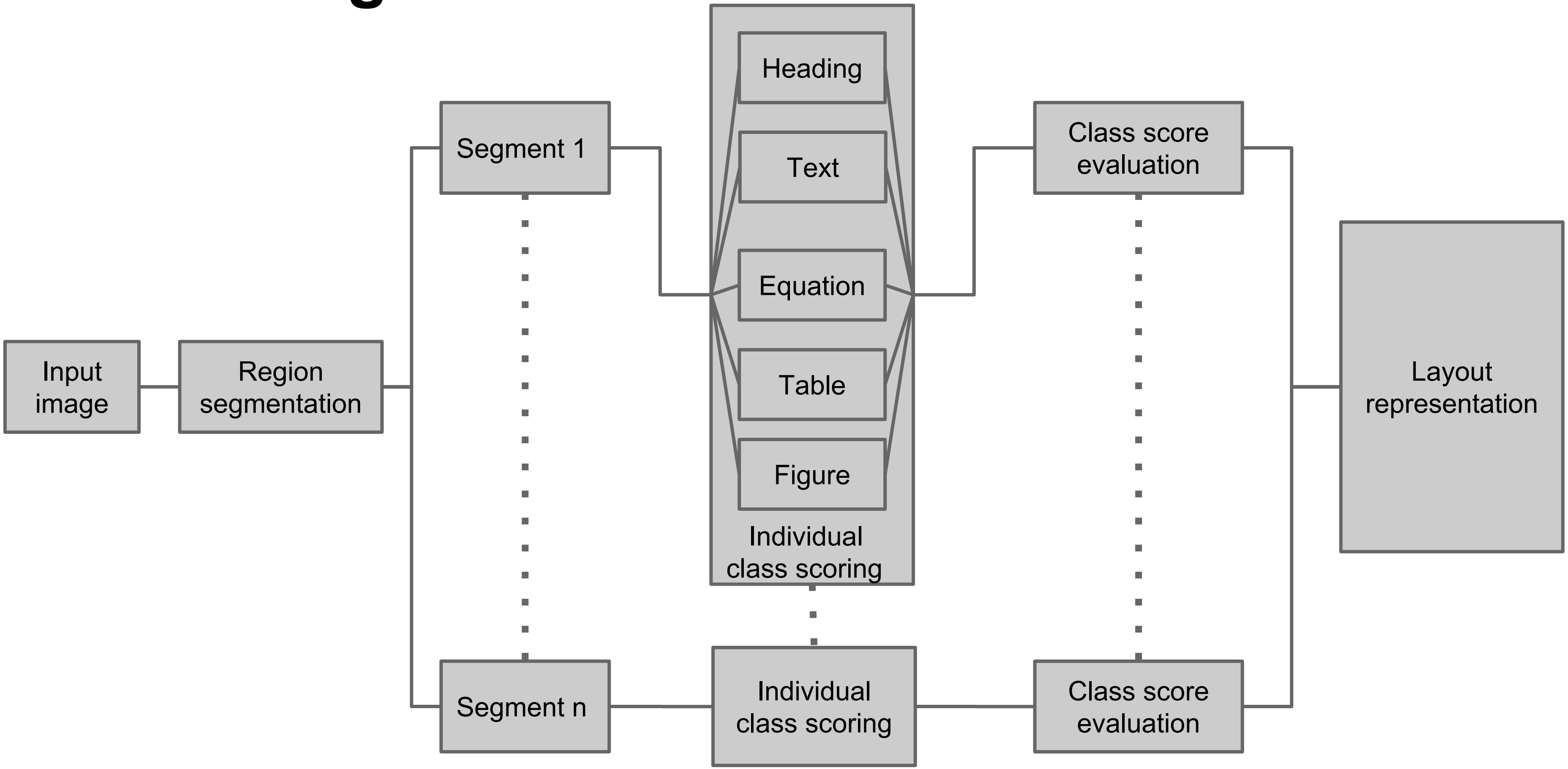
Document layout analysis

Mallikarjun B R

Problem statement:

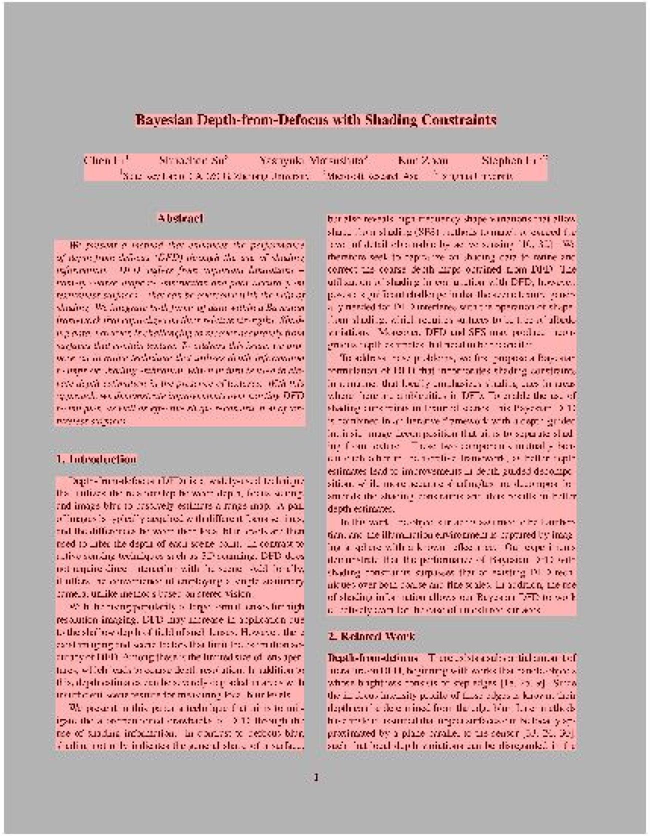
- Aim of the project is to recognize the layout of technical documents.
- Layout is defined on the basis of location and class of segments within the document.
- Region segmentation is done using morphological processes.
- Classes considered are Headings, Text, Tables and Figures.
    - Headings are scored on the basis of location.
    - Text is scored on the basis of statistical analysis of connected components.
    - Tables are scored on the basis of number of parallel lines.
    - After thresholding, rest are considered as figures.
- Based on scores, classes are assigned.

Block diagram:



Segmentation

- Identification of columns using vertical SE.
- Identification of text lines and calculation of average space between lines.
- Division into sub-regions using SE of height equal to calculated line spacing.
- Identifying multi-column titles and figures.
- Improved segmentation of text blocks based on empty space at the end of paragraph.


Bayesian Depth-from-Defocus with Shading Constraints

Abstract

1. Introduction

2. Related Work

Segmentation

where $P(\mathbf{d})$ is the prior distribution of depth map $\mathbf{d}$, $P(\mathbf{I}^{(1)}, \mathbf{I}^{(2)}|\mathbf{d})$ is the likelihood of observations $\mathbf{I}^{(1)}, \mathbf{I}^{(2)}$, and L is the log likelihood of P, i.e. $L = -\log P$.

The likelihood term can be modeled as the basic DFD energy from Eq. (8), and the prior term as depth smoothness along the links [33]:

$$L(\mathbf{I}^{(1)}, \mathbf{I}^{(2)}|\mathbf{d}) = \sum_{x \in \mathbf{d}} (I_x^{(1)} * h(\sigma, d) - I_x^{(2)})^2, \quad (10)$$

$$L(\mathbf{d}) = \lambda \sum_{(s,t)} (d_s - d_t)^2. \quad (11)$$

Hereafter, this particular formulation will be referred to as *standard DFD*.

To optimize the MRF model of Eqs. (10)-(11), we use the max product variant of the belief propagation algorithm [27], with a message update schedule that propagates messages in one direction and updates each node immediately.

3.3. Shading-based prior term

The smoothness prior of Eq. (11) can reduce noise in the reconstructed depth, but does not provide any additional knowledge about the scene. We propose to use a more informative prior based on the shading observed in the DFD image pair, which is helpful both for reconstructing surfaces with little texture content and for incorporating the fine-scale shape details that shading exhibits. In this section, we consider the case of uniform-albedo surfaces, for which shading can be easily measured. The more complicated case of textured surfaces will be addressed in Sections 3.4-3.5.

Lambertian shading can be modeled as a quadratic function of the surface normal [23, 10]:

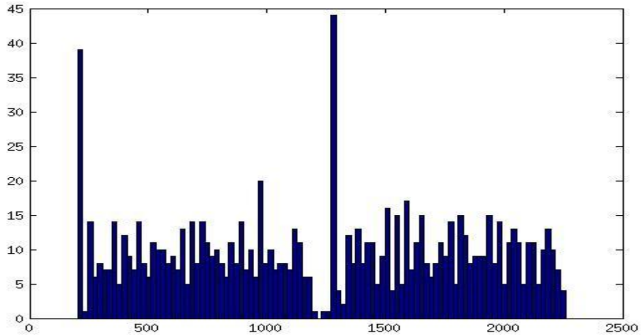
$$s(\mathbf{n}) = \mathbf{n}^T M \mathbf{n}, \quad (12)$$

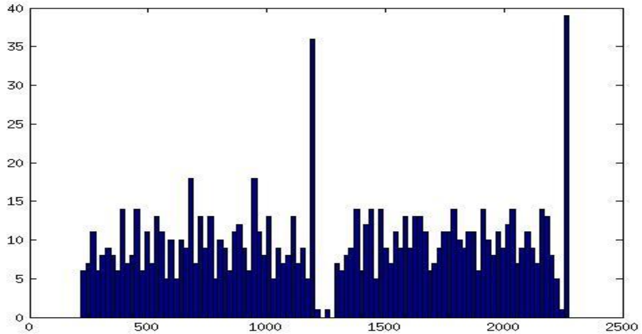
where $\mathbf{n}^T = (n_x, n_y, n_z, 1)$ for surface normal $\mathbf{n}$, and M is a symmetric 4×4 matrix that depends on the lighting environment. With this shading model, we solve for the surface normal of each pixel using the method in [10]. We also obtain the 3D coordinates for each point by re-projecting each pixel into the scene according to its image coordinates (x, y), depth value d from DFD, and the perspective projection model:

$$\left(\left(x - \frac{w}{2}\right) u d, \left(y - \frac{h}{2}\right) u d, d\right),$$

where $w \times h$ is the resolution of the image, and u is the pixel size.

For each pair of linked pixels s, t in the MRF, we now have their depths d_s, d_t, 3D positions p_s, p_t, and normals n_s, n_t. Since the vector $\overrightarrow{p_s p_t}$ should be perpendicular to





but also reveals high frequency shape variations that allow shape-from-shading (SFS) methods to match or exceed the level of detail obtainable by active sensing [10, 32]. We therefore seek to capitalize on shading data to refine and correct the coarse depth maps obtained from DFD. The utilization of shading in conjunction with DFD, however, poses a significant challenge in that the scene texture generally needed for DFD interferes with the operation of shape-from-shading, which requires surfaces to be free of albedo variations. Moreover, DFD and SFS may produce incongruous depth estimates that need to be reconciled.

To address these problems, we first propose a Bayesian formulation of DFD that incorporates shading constraints in a manner that locally emphasizes shading cues in areas where there are ambiguities in DFD. To enable the use of shading constraints in textured scenes, this Bayesian DFD is combined in an iterative framework with a depth-guided intrinsic image decomposition that aims to separate shading from texture. These two components mutually benefit each other in the iterative framework, as better depth estimates lead to improvements in depth-guided decomposition, while more accurate shading/texture decomposition amends the shading constraints and thus results in better depth estimates.

In this work, the object surface is assumed to be Lambertian, and the illumination environment is captured by imaging a sphere with a known reflectance. Our experiments demonstrate that the performance of Bayesian DFD with shading constraints surpasses that of existing DFD techniques over both coarse and fine scales. In addition, the use of shading information allows our Bayesian DFD to work effectively even for the case of untextured surfaces.

Table Feature Identification

Tables can be identified using the following properties:

- Number of parallel lines
- Angles of lines (only horizontal or vertical)
- Start and end of parallel lines are aligned

Work Done:

- Extracted edges using Canny Edge Detection
- Removed text edges using pre-processing (to avoid noise in line detection)
- Identified lines and their angles using Hough Transform
- Accept only if lines are horizontal or vertical

Text & Heading Feature Identification

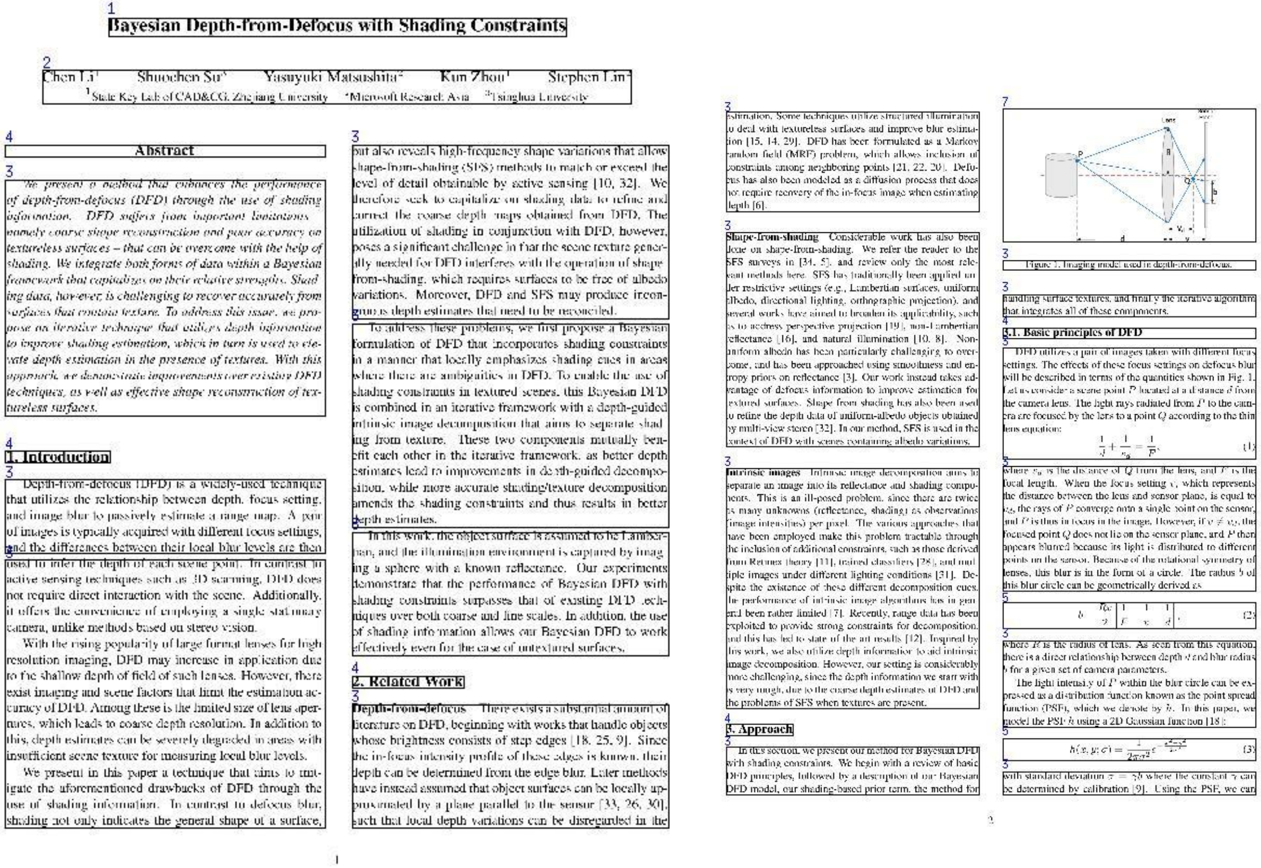
Text can be identified using average and variance of area of connected components.

Work Done:

- Pre-processing to remove noise and improve connectivity within letters.
- Calculation of area of each of the connected components.
- Experimentation has been done based on the mean and variance of area of connected components.
- Headings and Subheadings are identified based on connected component area and position.
- Equations are identified based on statistical data of connected component area.

Results:

Class	Label
Heading	1
Authors	2
Text	3
Sub Heading	4
Equations	5
Tables	6
Figures	7



1
Bayesian Depth-from-Defocus with Shading Constraints

2
Chen Li[1] Shuochen Su[3] Yasuyuki Matsushita[2] Kun Zhou[1] Stephen Lin[2]
[1]State Key Lab of CAD&CG, Zhejiang University [2]Microsoft Research Asia [3]Tsinghua University

4
Abstract

3
We present a method that enhances the performance of depth-from-defocus (DFD) through the use of shading information. DFD suffers from important limitations, namely coarse shape reconstruction and poor accuracy on textureless surfaces – that can be overcome with the help of shading. We integrate both forms of data within a Bayesian framework that capitalizes on their relative strengths. Shading data, however, is challenging to recover accurately from surfaces that contain texture. To address this issue, we propose an iterative technique that utilizes depth information to improve shading estimation, which in turn is used to elevate depth estimation in the presence of textures. With this approach, we demonstrate improvements over existing DFD techniques, as well as effective shape reconstruction of textureless surfaces.

4
1. Introduction

3
Depth-from-defocus (DFD) is a widely-used technique that utilizes the relationship between depth, focus setting, and image blur to passively estimate a range map. A pair of images is typically acquired with different focus settings, and the differences between their local blur levels are then

3
used to infer the depth of each scene point. In contrast to active sensing techniques such as 3D scanning, DFD does not require direct interaction with the scene. Additionally, it offers the convenience of employing a single stationary camera, unlike methods based on stereo vision.

With the rising popularity of large format lenses for high resolution imaging, DFD may increase in application due to the shallow depth of field of such lenses. However, there exist imaging and scene factors that limit the estimation accuracy of DFD. Among these is the limited size of lens apertures, which leads to coarse depth resolution. In addition to this, depth estimates can be severely degraded in areas with insufficient scene texture for measuring local blur levels.

We present in this paper a technique that aims to mitigate the aforementioned drawbacks of DFD through the use of shading information. In contrast to defocus blur, shading not only indicates the general shape of a surface,

3
but also reveals high-frequency shape variations that allow shape-from-shading (SFS) methods to match or exceed the level of detail obtainable by active sensing [10, 32]. We therefore seek to capitalize on shading data to refine and correct the coarse depth maps obtained from DFD. The utilization of shading in conjunction with DFD, however, poses a significant challenge in that the scene texture generally needed for DFD interferes with the operation of shape-from-shading, which requires surfaces to be free of albedo variations. Moreover, DFD and SFS may produce inconsistent depth estimates that need to be reconciled.

3
To address these problems, we first propose a Bayesian formulation of DFD that incorporates shading constraints in a manner that locally emphasizes shading cues in areas where there are ambiguities in DFD. To enable the use of shading constraints in textured scenes, this Bayesian DFD is combined in an iterative framework with a depth-guided intrinsic image decomposition that aims to separate shading from texture. These two components mutually benefit each other in the iterative framework, as better depth estimates lead to improvements in depth-guided decomposition, while more accurate shading/texture decomposition amends the shading constraints and thus results in better depth estimates.

3
In this work, the object surface is assumed to be Lambertian, and the illumination environment is captured by imaging a sphere with a known reflectance. Our experiments demonstrate that the performance of Bayesian DFD with shading constraints surpasses that of existing DFD techniques over both coarse and fine scales. In addition, the use of shading information allows our Bayesian DFD to work effectively even for the case of untextured surfaces.

4
2. Related Work

3
Depth-from-defocus There exists a substantial amount of literature on DFD, beginning with works that handle objects whose brightness consists of step edges [18, 25, 9]. Since the in-focus intensity profile of these edges is known, their depth can be determined from the edge blur. Later methods have instead assumed that object surfaces can be locally approximated by a plane parallel to the sensor [33, 26, 30], such that local depth variations can be disregarded in the

1

3
estimation. Some techniques utilize structured illumination to deal with textureless surfaces and improve blur estimation [15, 14, 29]. DFD has been formulated as a Markov random field (MRF) problem, which allows inclusion of constraints among neighboring points [21, 22, 20]. Defocus has also been modeled as a diffusion process that does not require recovery of the in-focus image when estimating depth [6].

3
Shape-from-shading Considerable work has also been done on shape-from-shading. We refer the reader to the SFS surveys in [34, 5], and review only the most relevant methods here. SFS has traditionally been applied under restrictive settings (e.g., Lambertian surfaces, uniform albedo, directional lighting, orthographic projection), and several works have aimed to broaden its applicability, such as to address perspective projection [19], non-Lambertian reflectance [16], and natural illumination [10, 8]. Non-uniform albedo has been particularly challenging to overcome, and has been approached using smoothness and entropy priors on reflectance [3]. Our work instead takes advantage of defocus information to improve estimation for textured surfaces. Shape from shading has also been used to refine the depth data of uniform-albedo objects obtained by multi-view stereo [32]. In our method, SFS is used in the context of DFD with scenes containing albedo variations.

3
Intrinsic images Intrinsic image decomposition aims to separate an image into its reflectance and shading components. This is an ill-posed problem, since there are twice as many unknowns (reflectance, shading) as observations (image intensities) per pixel. The various approaches that have been employed make this problem tractable through the inclusion of additional constraints, such as those derived from Retinex theory [11], trained classifiers [28], and multiple images under different lighting conditions [31]. Despite the existence of these different decomposition cues, the performance of intrinsic image algorithms has in general been rather limited [7]. Recently, range data has been exploited to provide strong constraints for decomposition, and this has led to state of the art results [12]. Inspired by this work, we also utilize depth information to aid intrinsic image decomposition. However, our setting is considerably more challenging, since the depth information we start with is very rough, due to the coarse depth estimates of DFD and the problems of SFS when textures are present.

4
3. Approach

3
In this section, we present our method for Bayesian DFD with shading constraints. We begin with a review of basic DFD principles, followed by a description of our Bayesian DFD model, our shading-based prior term, the method for

7

3
Figure 1. Imaging model used in depth-from-defocus.

3
handling surface textures, and finally the iterative algorithm that integrates all of these components.

4
3.1. Basic principles of DFD

5
DFD utilizes a pair of images taken with different focus settings. The effects of these focus settings on defocus blur will be described in terms of the quantities shown in Fig. 1. Let us consider a scene point P located at a distance d from the camera lens. The light rays radiated from P to the camera lens are focused by the lens to a point Q according to the thin lens equation:

$$\frac{1}{d} + \frac{1}{v_d} = \frac{1}{F}, \quad (1)$$

3
where v_d is the distance of Q from the lens, and F is the focal length. When the focus setting v, which represents the distance between the lens and sensor plane, is equal to v_d, the rays of P converge onto a single point on the sensor, and P is thus in focus in the image. However, if $v \neq v_d$, the focused point Q does not lie on the sensor plane, and P then appears blurred because its light is distributed to different points on the sensor. Because of the rotational symmetry of lenses, this blur is in the form of a circle. The radius b of this blur circle can be geometrically derived as

5
$$b = \frac{Rv}{2}\left|\frac{1}{F} - \frac{1}{v} - \frac{1}{d}\right|, \quad (2)$$

3
where R is the radius of lens. As seen from this equation, there is a direct relationship between depth d and blur radius b for a given set of camera parameters.

The light intensity of P within the blur circle can be expressed as a distribution function known as the point spread function (PSF), which we denote by h. In this paper, we model the PSF h using a 2D Gaussian function [18]:

5
$$h(x, y; \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (3)$$

3
with standard deviation $\sigma = \gamma b$ where the constant γ can be determined by calibration [9]. Using the PSF, we can

2

Conclusion on results:

- Approximately 80% accuracy in segmentation and classification.
- Errors include,
    - Equations get segmented as text because of similar line spacings.
    - Incorrect paragraph segmentation due to inconsistent line spacing. Solved in most of the cases.
    - Incorrect subheadings classification as text because of similar statistics.