Estimating Scale of a Scene from a Single Image Based on Defocus Blur and Scene Geometry

Takayuki Okatani and Koichiro Deguchi Tohoku University 6-6-01 Aramaki Aza Aoba, Aoba-ku Sendai, Japan okatani@fractal.is.tohoku.ac.jp

Abstract

Using an imaging system in which the image plane can be tilted with respect to the optical axis of the lens, the image of a large-scale scene that appears to be a miniature to human eyes can be captured. This phenomenon suggests that the image contains information regarding the scale of the scene and that human vision can extract this information and recognize the scene scale from a single image. In this study, we consider how human vision can perform this single-view scale estimation. Although it is obvious that the existence of defocus blur in the image that simulates a shallow DOF plays an essential role in the scale estimation, we propose that this alone is not sufficient to explain the estimtation mechanism. By incorporating a few assumptions, we theoretically show that scale estimation is made possible when (1) the 3D structure of the scene can be recovered from the image and furthermore, (2) the structure is combined with the defocus blur. Further, we present a simple algorithm for scale recognition and demonstrate its working using a real image.

1. Introduction

Figure 1 shows images of an actual scenery captured by a special image-capturing method. As shown here, all these images exhibit a strong defocus blur that (appears to) corresponds to their depth of field (DOF); they seem to have a shallow DOF. Probably owing to the existence of this defocus blur, these images appear to the eyes of the viewer as miniature scenes although the scenes are real. That is, the viewers (wrongly) recognize the scale of the scene to be very small.

These images were captured by an imaging system in which the image plane can be tilted. For example, by using a tilt lens, the image plane can be tilted with respect to the optical axis of the lens, as shown in Fig.2. By using this type of imaging systems for a scene that will be entirely in focus if an ordinary imaging system is used, a part of the image can be redndered out of focus. In the images Fig.1, by using this technique, the defocus blur that would have emerged if the scenes had been miniatures is artificially reproduced. There are several photographers [2, 4] who are well-known for using this technique.

In this paper, we discuss the fact that the human visual system recognizes the (small) scale of a scene based on only a single image, such as those in Fig.1. It is obvious with these images that the apparent defocus blur that simulates a shallow DOF plays a central role in this recognition process. Further, taking into consideration that the DOF is inversely proportional to the distance from the camera to the scene, it seems that the mechanism of the scale recognition is completely explained. (Since the DOF is proportional to the lens aperture as well, the knowledge of the lens aperture will also be necessary for the explanation.) However, a shallow DOF does not necessarily provide an impression of the scene scale; an example is shown in Fig.3. Unlike for the images shown in Fig.1, we do not seem to be able to recognize the scale of the scene. Moreover, the scene in the right image with a shallower DOF does not appear to be smaller than the scene in the left image. Therefore, it is evident that the DOF alone does not enable scale estimation in the human visual system. In what follows, we consider the mechanism involved in scene scale estimation.

In Section 2, we analyze the defocus blur generated by an imaging system with a tilted image plane. Based on the



Figure 2. A picture of a tilt lens attached to a camera and its optical diagram.



Figure 1. Images of real scenes that appear to be miniature scenes, which are captured by an imaging system with a tilted image plane.



Figure 3. Examples of images with shallow depth of field (DOF), which are synthetically generated using scene depth data [7]. The right image has shallower DOF than the left image. These do not give us any impression on the scene scale.

result, in Section 3, we discuss how the scene scale can be estimated and present an algorithm for scale estimation. In Section 4, we demonstrate how the algorithm actually functions using one of the images in Fig.1. Section 5 concludes this paper.

2. Imaging systems and defocus blurs

We start with modelling the defocus blur generated by a tilted image plane for an infinitely distant scene. Then, we show that the defocus blur is mostly identical to that generated by an ordinary imaging system for a planar scene that is close to the imaging system.

2.1. Defocus blur generated by a tilted image plane

Assume an imaging system as shown in Fig.4 whose focus is on a planar surface at a certain distance from the lens. Assume that the image plane is tilted about the x axis (perpendicular to the paper) by an angle ϕ while maintaining the



Figure 4. Defocus blur generated in the imaging system with a tilted image plane.

focus on the planar surface at the image center. As a result, defocus blur will emerge at off-center ($y \neq 0$) image points. We denote the size of the blur by *d* at an image height *y*. Considering the similarity relation between the two triangles sharing *P* as the common vertex (the base of one triangle is *D* and that of the other is approximately $d \cos \phi$), we have

$$\frac{d\cos\phi}{y\sin\phi} \approx \frac{D}{v'}.$$
 (1)

The approximation (\approx) holds when ϕ is small. We may assume it to be small since the tilt ϕ of the image plane is usually small. From this, we have

$$d \approx D \frac{\tan \phi}{v'} y. \tag{2}$$

Strictly, the blur at image points off the paper (i.e., the image at $x \neq 0$) will have shapes different from that of the blur at x = 0, which is considered here. Since we are interested only in the order of the blur size and the blur sizes at these



Figure 5. Defocus blur generated by an ordinary imaging system for a planar scene

points actually have the same order, we do not consider the difference in what follows.

2.2. Natural defocus blur of a planar scene

Assume that an ordinary camera with a zero-tilt image plane captures an image of an oblique planar scene that makes an angle θ with the lens (and the image plane), as shown in Fig.5. The focus on the optical axis is assumed to be on the planar surface at a distance u' along the optical axis. In other words, the resulting image of the planar scene is focused only along the x axis. For the image points off the x axis, defocus blur emerges since their focal points are off the image plane due to the tilt of the plane.

For this imaging system, the size d of the defocus blur at an image height y is given by

$$d = \frac{D|v - v'|}{v}.$$
 (3)

By denoting the focal length of the lens as f, from the Gauss rule, we have

$$\frac{1}{u} + \frac{1}{v} = \frac{1}{u'} + \frac{1}{v'} = \frac{1}{f}.$$
(4)

From the similarity relation of the two triangles sharing the optical center O as a common vertex, we have

$$\frac{u-u'}{\tan\theta} \cdot \frac{1}{u} = \frac{y}{v'}.$$
(5)

The elimination of v and v' from Eq.(3) using these equations yields

$$d = D \frac{\tan \theta}{u'} y. \tag{6}$$

2.3. Comparison between the two defocus blurs

Comparing Eq.(2), the artificial defocus blur created by the tilted image plane, with Eq.(6), which provides the natural blur of the planar scene, it can be seen that both share the property that *d* is proportional to the lens diameter *D* and the image height *y*; the only difference is in their proportion coefficients. In the former, it is $\tan \phi/v'$, i.e., the tangent of the angle of the image plane divided by the distance from the image plane to the optical center, whereas in the latter it is $\tan \theta/u'$, i.e., the tangent of the angle of the plane divided by the distance to the plane. From this, we conclude the following:

Result 1. The natural defocus blur generated by an ordinary imaging system for a tilted planar scene (that is proximal to the imaging system) can be reproduced by the defocus blur generated on an image (of an infinitely distant scene) that is captured using an imaging system with a tilted image plane.

It should be noted that in the imaging system with the tilted image plane, the natural blur associated with the DOF of the scene can coexist, depending on the setup of the imaging system. In such a case, the final blur is given by the sum (convolution) of the natural blur and the blur due to the tilt of the image plane. However, if the image has a deep DOF, for example, when the scene is sufficiently distant, then nly the blur due to the tilt of the image plane is found to exist. This is the case with regard to the images shown in Fig.1; the scenes are distant from the camera and thus their DOFs are deep. Therefore, the defocus blur emerging in these images precisely reproduces the natural blur that would have been generated for a planar scene. Since the scenes are approximately planar, the defocus blur simulates the DOF of these images. It is evident that this is one of the most important factors in the scale recognition by the human vision for the images shown in Fig.1.

It should also be noted that, other than the defocus blur, there is a difference between the images taken using a tilted image plane and those taken using an ordinary imaging system; the former will have a larger magnification ratio in the tilted direction (i.e., the direction of the y axis). That is, the image captured by a tilted image plane has a different aspect ratio from that of the image captured using an ordinary imaging system. However, as mentioned earlier, the tilt angle ϕ of the image plane is usually small; hence, we may neglect this.

3. Algorithm for estimating scale of a planar scene

In this section, based on the analysis in the last section, we discuss the mechanisms for estimating the scale of a scene from a single image.

3.1. Problem formulation

We begin with the formulation of the problem. Since the imaging system with a tilted image plane simulates the defocus blur of planar scenes, for the moment, we restrict our attention to planar scenes. As derived in Eq.(6), the defocus blur d at image height y is proportional to the distance y to

the line in focus (the *x* axis). By denoting the proportion coefficient by $\alpha \equiv d/y$, we have

$$\alpha \left(=\frac{d}{y}\right) = \frac{D}{u'} \tan \theta.$$
(7)

We introduce the following assumption.

Assumption 1. The proportion coefficient α can be estimated from the image.

We will discuss the validity of this assumption in the next section. Under this assumption, α is a known parameter in Eq.(7), and there are three unknowns on the right hand side of the equation: the lens diameter *D*, the plane angle θ , and the distance u' from the lens to (a particular point of) the scene. Since the camera distance u' is directly related to the scene scale, it is sufficient to be able to determine u'. Thus, we consider the determination of u'. However, we have one equation and three unknowns; hence, it is impossible to determine u'.

Further, we introduce the following assumptions.

Assumption 2. *The lens diameter D is known (Or at least its approximate range is known).*

Assumption 3. The angle θ of the plane can be (geometrically) recovered from the image itself (Or equivalently, some prior knowledge that makes this possible is available).

Assumption 2 might be somewhat too strong an assumption if it implies that its exact value should be known, since the lens aperture is variable in most imaging systems including our eyes; thus, we cannot know this value in advance. However, it is possible to assume a range of D values instead of a single value.

Regarding Assumption 3, a variety of methods can be used based on the target image and the prior knowledge available. We present below a simple method that is based on classical vision geometry.

3.2. Plane angle estimation: single-view 3D reconstruction

Since we do not have the camera intrinsics, we must estimate some of them from the image. Considering the nature of the problem considered here, we may assume that the focal length (precisely the distance v' from the lens optical center to the image plane) is the only unknown parameter among the camera intrinsics. In other words, we may assume that the skew is 0, the aspect ratio is 1, and the principal point coincides with the center of the image.

Further, Assumption 3 will be met, for example, when the vanishing line of the target plane can be identified on the image and v' can be estimated as follows. Let *l* be the



Figure 6. Geometry of the target plane and its vanishing line.

distance on the image plane from the image center to the vanishing line. As shown in Fig.6, the geometric relation between θ , *l*, and *v'* is given as follows:

$$\tan\left(\frac{\pi}{2} - \theta\right) = \frac{l}{v'},\tag{8}$$

from which we can determine θ , given *l* and *v'*.

By using classical geometry-based methods of computer vision, the vanishing line of a plane and v' can be estimated from a single image in several ways. The vanishing line can be determined when it directly appears on the image, when there are more than two pairs of parallel lines on the plane (the first image of Fig.1), when the structure of texture of the plane is known (the third image in Fig.1) and so on.

One of the simplest methods to estimate v' is to use a pair of parallel lines in the image if they are available. Denoting the vanishing points of the parallel lines by the homogeneous image coordinates \mathbf{p}_1 and \mathbf{p}_2 , the following holds (Eq.(7.13) in [3]):

$$\mathbf{p}_1^{\mathsf{T}} (\mathbf{K} \mathbf{K}^{\mathsf{T}})^{-1} \mathbf{p}_2 = 0, \tag{9}$$

where, in our camera model, K is given as

$$\mathbf{K} = \begin{bmatrix} v' & 0 & 0\\ 0 & v' & 0\\ 0 & 0 & 1 \end{bmatrix}.$$
 (10)

From this, we can determine v'. Even if there are no parallel lines, it could be possible to estimate v' in a few cases, for example, by using the textures of the target plane.

We may refer to these problems as the *single-view 3D reconstruction*. There are many studies in this area. Although we do not proceed further, the above discussion demonstrates that the plane angle θ can actually be determined from the single image, provided some natural prior knowledge is available.

3.3. Scale determination via fusion of defocus blur and plane angle estimation

When Assumptions 2 and 3 are met, that is, *D* is known and θ is estimated from the image, then Eq.(7) enables us to



Figure 7. The algorithm for scene scale estimation.

calculate the distance u' from the camera to the scene from the proportion coefficient α of the defocus blur. Since the distance u' is directly related to the scene scale, we have completed the estimation.

We started with the fact that the images captured using a tilted image plane provide an impression of the small scale of the scenes. Thus far, we have considered planar scenes since the imaging system with a tilted image plane precisely reproduces the defocus blur of planar scenes, as shown in the last section. However, it is not plausible that the feasibility of the scale recognition is determined by the planarity of the target scene. In the case of planar scenes, the scene scale is determined from the spatial gradient α of the size of the defocus blur as well as the plane angle θ . Therefore, by generalizing this result to nonplanar scenes, we conjecture that only when the spatial variations in both the size of the defocus blur and the scene depth are recovered, the scene scale can be determined by the combination of these two factors. Figure 7 shows the flowchart for the algorithm realizing this conjecture. We believe that this explains why Fig.3 does not provide any scale impression. That is, there is no cue for recovering the (relative) depth of the scene; there are several objects in the scene whose relative depths cannot be estimated from the image.

3.4. A note on the distinction from DFD

The method of depth from defocus (DFD) calculates the absolute distances from the lens to the scene points. Hence, some readers might assume that DFD has already explained the mechanism of the scale recognition provided that the same assumptions as those provided above are incorporated; however, this is not true.

DFD is based on the following equation, which provides the size d of the defocus blur [6, 8]:

$$d = \frac{D|uf - v'u + fv'|}{fu},$$
 (11)

where D is the lens diameter; u, the distance to the scene point corresponding to the blurred image point of interest;

v', the distance from the image plane to the optical center; f, the (true) focal length of the lens. In DFD, assuming D, v', and f to be known, u is determined using the size d of the defocus blur. Since v' and f are of the same order, the additional knowledge of f might not appear to be critical. However, in DFD, only a small difference between f and v' plays an essential role in the determination of the scene depth. Thus, DFD alone does not explain the mechanism of the scale recognition.

4. Example of the single view scale estimation

In this section, we demonstrate the scene scale estimation by the method outlined above and examine its validity. In Assumption 1, we assumed that the proportion coefficient α can be estimated from the given image alone. Thus, we start with how this is performed.

4.1. Estimation of the size of the defocus blur from a single image

The problem of estimating α from the image is, in its most general (and difficult) form, equivalent to the estimation of the size of the defocus blur at each image point. Since the scene can be arbitrary, this is obviously an illposed problem that cannot be solved if no prior knowledge is available. In fact, ordinary DFD methods use multiple images to estimate the blur. Pentland proposed a method for estimating the blur from a single image using image edges [6] (also see [5]). However, this method requires precise information regarding the pre-blurred edges, which may be too strong a requirement for our purpose.

By simple inspection, our visual system seems to be able to estimate the size of the defocus blur from a single image to some degree of accuracy, if not precisely. The problem of how human vision system can perform this estimation is another interesting problem. A possible answer to this is that the human vision system performs a computation similar to the *image hallucination*. Image hallucination recovers sharp structures of an image from a blurred image by using prior knowledge such as the appearance of human faces [1] and primary sketches (e.g., edges) [9]. Several promising results have been reported. If the pre-blurred images can be obtained by a similar method, then it is straightforward to determine the size of the blur by comparing the blurred and recovered pre-blurred images.

Although this image-hallucination-based approach seems promising, it is beyond the scope of this paper. Hence, we introduce here a somewhat strong assumption on the statistical property of the original image of the target scene and present a simple method based on this. By original image we mean a pre-blurred image of the given blurred image (hence, it is unobserved). For this original image, we assume that *the locally measured power* spectrum is identical over the entire image.

The point spread function (PSF) associated with the defocus blur behaves as a low-pass filter. Thus, the power at high spatial frequencies should have a smaller value for the blurred portions of the image than for the sharp portions. Therefore, under the above assumption that the local power spectrum is identical at every image point, by examining the power deterioration at high frequencies, we will be able to determine the size of the defocus blur.

The actual computation is performed in the following manner. For several sample heights y_1, \ldots, y_n , we calculate the powers of the corresponding local image areas at a particular frequency and then plot them to obtain a profile of the power with respect to y. Then, by comparing the profile with our theoretical model (20), which will be derived in Appendix A, we estimate α . In order to calculate the power of a local area at a particular frequency, we use the Gabor transform, being inspired by the reported fact that a similar computation is performed in biological vision systems. The details of the estimation of α are provided in Appendix A.

4.2. Demonstration of the recovery of the scene scale

We use the first image in Fig.1 to test the algorithm; its size is 910×609 pixels. As shown in Fig.7, the algorithm consists of three parts. The first is to determine the plane angle θ , the second is to determine the blur factor α , and the third is to combine their results and determine the absolute scene depth u'.

First, we determine the plane angle θ . In the image, there exist two orthogonal pairs of parallel lines. Thus, by using the lines, we compute v', identify the vanishing line, and measure the distance *l* from the principal point to the vanishing line. Subsequently, we calculate θ from Eq.(8).

As shown in Fig.8, we can identify the vanishing points of the pairs of the parallel lines as (-12.5, 1.10) and (0.44, 1.10), respectively. We assume here that the image center coincides with the principal point (0, 0) and use the image coordinates such that the vertical side of the image is length 1. (Note that we want to estimate the plane angle, which is independent of the scale of the coordinate system.) The vanishing line is obtained as the line connecting these two vanishing points. Substituting $\mathbf{p}_1 \equiv [-12.5, 1.10, 1]^{\mathsf{T}}$ and $\mathbf{p}_2 \equiv [0.44, 1.10, 1]^{\mathsf{T}}$ into Eq.(9), we immediately have $\nu' \approx 2.07$. In addition, the distance from the image center to the vanishing line is measured as l = 1.10. The substitution of these values into Eq.(8) yields $\tan \theta \approx 1.87$ ($\theta \approx 62^{\circ}$).

We then determine α , the proportion coefficient of the defocus blur at y. We split the image into multiple horizontal rectangules of 910 × 64, and then calculate the power (the total sum of Eq.(20)) for each region. The plot in Fig.9 shows the profile of the power variation with respect to y. The vertical axis represents the normalized power such that the maximum is normalized to 1, and the horizontal axis



Figure 8. Geometric recovery of v' and θ . Orthogonal pairs of two parallel lines are used.

represents the image *y* coordinates in pixels. The maximum power is found at the region in focus, or more precisely, the horizontal line around the 341-th pixel from the top. This indicates that when the image was captured, the image plane was tilted about the *x* axis. On the plot shown in Fig.9, theoretical profiles for $\alpha = 0.005$, 0.01, and 0.02 are shown. By comparing these with the measured profile, we determine α to be approximately 0.01.

Substituting the results $\tan \theta = 1.87$ and $\alpha = 0.01$ into Eq.(7), we have

$$u' \approx 190D. \tag{12}$$

This result implies that the distance from the lens to the scene (specifically, a scene point corresponding to the line in focus) is approximately 200 times the lens diameter. For example, if we assume the lens diameter to be 3 cm (a considerably large lens), the distance to the scene is 6 m. If the lens diameter is 0.5 cm, the distance is only 1 m. Considering together that the distance v' from the image plane to the optical center is given as $v' \approx 2.0 \times$ (image height), we have no choice other than to recognize the target scene to be a miniature since the scene must exist about 1 m distance from the camera.

5. Summary and discussion

Images of distant scenes that are captured using an imaging system with a tilted image plane sometimes appear to human eyes as miniature scenes. In this paper, we have discussed the mechanism of this scale recognition, as performed by the human vision system.

We have shown that the defocus blur generated by the imaging system with a tilted image plane for a distant scene coincides with the defocus blur generated by an ordinary imaging system for a planar scene. The analysis presents a relation among the spatial variation in the size of the defocus blur (i.e., α), lens diameter, angle of the target plane,



Figure 9. Estimation of α . Upper: Illustration of the power measured at a particular frequency and local image areas. Lower: Measured power profile at frequency 0.33 [1/pixel] of the image and the theoretical profiles for $\alpha = 0.005, 0.01$, and 0.02. Note that the image coordinates ($y \in [0 : 609]$) are different from those used in Fig.8. The line in focus is y = 341.

and distance to a scene point. Using this relation, the distance to a scene point can be determined if the lens diameter is known in advance and if the size of the defocus blur as well as the plane angle can be estimated from the given image. There are several studies in this area that report methods of the plane angle estimation from a single image using various prior knowledge. We have shown one of the simplest methods. Other methods could also be applied. It is plausible that the human vision system selectively uses some of these methods depending on the image. Although the estimation of the size of the defocus blur is a difficult problem to formulate, the human vision system seems to be able to perform this with some degree of accuracy. In this paper, we present a simple method based on the assumption that the power spectrum measured at local image areas is identical at all points on the image.

The main argument in this paper is that the *combination* of the shallow DOF and the single-view recovery of the 3D structure of the scene enables the estimation of the scene scale. If either of the two is lost, the estimation of the scene scale is impossible. The existence of the shallow DOF is necessary but not sufficient; there must be some cues in the image from which the 3D structure of the scene can be recovered (geometrically, in most cases).

We have also shown that for a real image captured by an imaging system with a tilted image plane, the described method can compute an approximate value for the scene scale; precisely, it computes the ratio between the distance to the scene and the lens diameter. As mentioned before, although it is difficult to assume an exact value for the lens diameter, we may assume an approximate range of values that makes it possible to determine whether the scene is very close or fairly distant from the camera.

When observing in detail the images that provide the impression of miniature scenes, it is suspected that there are several other factors that affect the scale recognition, besides the mechanism described in this paper. For example, colors and kinds of objects existing in the scenes might contribute to the scale recognition. This will be investigated in future.

References

- S. Baker and T. Kanade. Limits on super-resolution and how to break them. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 24(9):1167–1183, 2002. 5
- [2] M. Gaál. Sightseeing tour. Hatje Cantz Verlag, 2005. 1
- [3] R. Hartley and A. Zisserman. Multi-View Geometry in Computer Vision. Cambridge University Press, 2000. 4
- [4] N. Honjo. *small planet*. Sun Chiapang (Little More Co., Ltd.), 2006. 1
- [5] S. H. Lai, C. W. Fu, and S. Chang. A generalized depth estimation algorithm with a single image. *IEEE Transaction* on Pattern Analysis and Machine Intelligence, 14(4):405–411, 1992. 5
- [6] A. P. Pentland. A new sense for depth of field. *IEEE Transac*tion on Pattern Analysis and Machine Intelligence, 9(4):523– 531, 1987. 5
- [7] D. Scharstein and R. Szeliski. High-accuracy stereo depth maps using structured light. In *Proceedings of CVPR*, volume 1, pages 195–202, 2003. 2
- [8] Y. Y. Schechner and N. Kiryati. Depth from defocus vs. stereo: how different really are they? In *Proceedings of Fourth International Conference on Pattern Recognition*, volume 2, pages 1784–1786, 1998. 5
- [9] J. Sun, N. N. Zheng, H. Tao, and H. Y. Shum. Image hallucination with primal sketch priors. In *Proceedings of Computer Vision and Pattern Recognition*, volume 2, pages 729– 736, 2003. 5

A. Estimation of α

The Gabor kernel is given by

$$g(x, y; \lambda_G, \theta, \sigma_G) = \exp\left(-\frac{x'^2 + y'^2}{2\sigma_G^2}\right) \cos\left(2\pi \frac{x'}{\lambda_G}\right), \quad (13)$$

where

$$x' = x\cos\theta + y\sin\theta, \tag{14a}$$

$$y' = -x\sin\theta + y\cos\theta,$$
 (14b)

and λ_G is the frequency parameter and σ_G is the Gaussian factor, which are related by

$$\frac{\sigma_G}{\lambda_G} = \frac{1}{\pi} \sqrt{\frac{\ln 2}{2}} \cdot \frac{2^b + 1}{2^b - 1},\tag{15}$$

where *b* is the bandwidth. We choose b = 1, which gives $\lambda_G \approx 0.56\sigma_G$. As for the direction θ of the kernel, we use 12 equally spaced samples from $[0 : \pi]$. Then, we calculate the power of some local area at the frequency $1/\lambda_G$ by calculating the sum of the squares of the responses of the kernel over the local area and the direction θ .

We assume the PSF of the defocus blur to be an isotropic Gaussian function, i.e.,

$$h(x,y;\sigma) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right),\tag{16}$$

where σ expresses the size of the blur. Its Fourier transform is given by

$$H(\xi,\eta;\sigma) \propto \exp\left(-\frac{1}{2}\sigma^2(\xi^2+\eta^2)\right).$$
(17)

Assuming the power of the Gabor response at frequency $1/\lambda_G$ of the pre-blurred ($\sigma = 0$) image to be 1 (by applying normalization), the power of the Gabor response at the same frequency for the image blurred with the above PSF may be approximately given as

$$\exp\left(-\frac{1}{2}\sigma^2\left(\frac{2\pi}{\lambda_G}\right)^2\right)^2.$$
 (18)

Therefore, using the result obtained in the previous section, we may assume that

$$\sigma = \alpha y. \tag{19}$$

Then, by substituting Eq.(19) into Eq.(18), the ratio of the power of the Gabor response at image height y to that at y = 0 may be represented as

$$\exp\left(-\frac{1}{2}\left(\frac{2\pi\alpha}{\lambda_G}\right)^2 y^2\right)^2.$$
 (20)

In conclusion, under the assumption that the power spectrum is identical everywhere (independent of image height y), the squared Gabor response is given by the square of Gaussian function with respect to y. Thus, computing the squared Gabor response at several y's, we can estimate α by fitting (20) to them.