FOCUSING ON OUT-OF-FOCUS: ASSESSING DEFOCUS ESTIMATION ALGORITHMS FOR THE BENEFIT OF AUTOMATED IMAGE MASKING

Geert J. Verhoeven [a]
[a] Ludwig Boltzmann Institute for Archaeological Prospection & Virtual Archaeology, Franz-Klein-Gasse 1, 1190 Vienna, Austria
Geert.Verhoeven@archpro.lbg.ac.at

Commission II, WG II/8

KEY WORDS: Automated masking, Defocus estimation, Depth of field, Edge extraction, Image-based modelling, Out-of-focus blur

ABSTRACT:

Acquiring photographs as input for an image-based modelling pipeline is less trivial than often assumed. Photographs should be correctly exposed, cover the subject sufficiently from all possible angles, have the required spatial resolution, be devoid of any motion blur, exhibit accurate focus and feature an adequate depth of field. The last four characteristics all determine the “sharpness” of an image and the photogrammetric, computer vision and hybrid photogrammetric computer vision communities all assume that the object to be modelled is depicted “acceptably” sharp throughout the whole image collection. Although none of these three fields has ever properly quantified “acceptably sharp”, it is more or less standard practice to mask those image portions that appear to be unsharp due to the limited depth of field around the plane of focus (whether this means blurry object parts or completely out-of-focus backgrounds). This paper will assess how well- or ill-suited defocus estimating algorithms are for automatically masking a series of photographs, since this could speed up modelling pipelines with many hundreds or thousands of photographs. To that end, the paper uses five different real-world datasets and compares the output of three state-of-the-art edge-based defocus estimators. Afterwards, critical comments and plans for the future finalise this paper.

1. INTRODUCTION

Photography is a very subjective medium. Although an exciting subject, compelling illumination and attractive composition are common denominators of many excellent photographs, the rules on how to obtain these characteristics are not carved in stone. Even if there is some consensus on good practice (e.g. the “rule” of thirds in image composition), great images often result from simply ignoring or breaking these “rules”. For instance, many photographers consider the „sharpness“ of an image as its most essential characteristic since it tells a lot about the photographer’s image acquisition and post-processing skills, while it also reveals the use of high-end imaging hardware. However, adding the right amount of blur to the right portion of an image often makes for a very aesthetic outcome. Panning the camera or photographing a moving subject can result in so-called **motion blur**, while a limited depth of field or a deliberately inaccurate focusing might yield appealing **out-of-focus blur** (also known as **defocus blur**).

However, both types of blur are often unwanted in photographs that serve to construct an object’s three-dimensional (3D) surface geometry digitally. Computer vision and traditional photogrammetric workflows along with the more recent hybrid approaches to Image-Based Modelling (IBM) all assume that the surface to be digitally modelled is “acceptably” sharp depicted throughout the whole image collection. Even though none of these three fields has ever properly quantified “acceptably sharp”, it is at least more or less standard practice in most IBM pipelines to mask those portions of the image that appear “not sharp enough” for proper extraction of 3D surface data.
This paper will assess how well- or ill-suited the current pool of defocus estimating algorithms are for automatically masking a series of photographs. This masking automation could seriously speed up IBM projects with many hundreds if not thousands of photographs, for which the tedious and error-prone manual masking of out-of-focus background and unsharp object regions could quickly become very time-consuming.

2. SOME BASIC CONCEPTS

2.1 Sharpness and blur

An image that exhibits many details with distinct boundaries between them is denoted “sharp”. Although “sharpness” is a perceptual term that relates to the details seen by the human visual system, image sharpness is usually related to the concepts of spatial resolution and acutance. The spatial resolution (often shortened to just resolution) is some distance Δx that equals the minimum distance between distinguishable objects in an image. As such, the spatial resolution of an image provides a fundamental limit to the information one can extract from an image. However, to extract that information, the contrast between neighbouring objects must be high enough as well. That gradient of the tonal change between neighbouring zones is referred to as **acutance**. Images with high acutance always feature sharp transitions between their tonal boundaries. For instance, sharpening a digital image alters its acutance, but leaves its spatial resolution unaltered.

However, to generate a sharp image with lots of details, one needs an imaging system that can distinguish fine object detail. This characteristic is called resolving power, and it defines the smallest resolvable feature within the imaging system's field of view. The resolving power of an optical imaging system (and hence the sharpness of the final image) is primarily determined by the amount of – often unavoidable – blur produced by the

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLII-2-1149-2018 | © Authors 2018. CC BY 4.0 License.

imaging procedure. To understand the concept of blur, it is essential to understand that an image is a visual representation of a specific physical object. Ideally, every object point is represented by a small point in the image. In reality, the image of each object point is a small blob resulting from the accumulated blurring that occurs throughout the imaging chain.

2.1.1 Diffraction and the optical PSF

Although all imaging systems suffer from specific lens aberrations, the diffractive nature of electromagnetic radiation would still put a physical limit on the smallest resolvable object even if such aberrations would be absent. Diffraction is a phenomenon that comes into play because wavefronts of propagating electromagnetic waves bend in the neighbourhood of tiny obstacles and spread out when passing apertures. Even when imaging a distant point source of electromagnetic energy (such as a star), the resulting image is therefore never a perfect point but a diffraction pattern.

The spatial energy distribution of this image spot is called the **Point-Spread Function** (PSF) and describes in three dimensions the smear or spread in the sensor plane introduced by the optical chain (Jensen, 1968). If the imaging system is aberration-free and completely diffraction-limited, the PSF of a perfectly focused point is a so-called **Airy diffraction pattern**. This Airy pattern describes the best-focused spot of electromagnetic radiation that a perfect lens with a circular aperture can generate. In the XY plane, it looks like a bright circular central patch (the **Airy disc**) and a series of dimmer concentric rings, each ring separated by a circle of zero intensity (Figure 1).

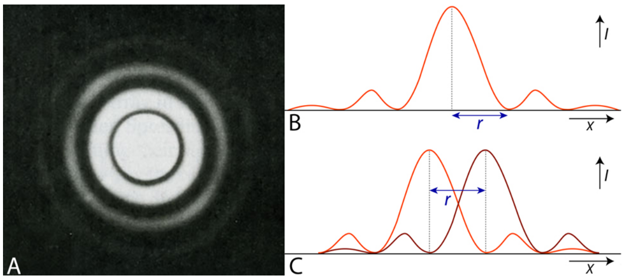

Figure 1. (A) shows an Airy diffraction pattern while (B) depicts the profile of the Airy PSF and (C) shows the Rayleigh criterion.

The radius r of this Airy disc equals $1.22\,\lambda f / D$. The spread thus depends on the wavelength of the electromagnetic radiation (λ), the lens’ focal length (f) and the diameter of the lens aperture (D) (Hecht, 2002). Note that smaller wavelengths and large lens apertures yield smaller Airy disks and that both variables can be used to minimise the size of the image spot. Obviously, smaller Airy discs yield a higher theoretical spatial resolving power of the imaging system, because the Rayleigh criterion states that two neighbouring points can be spatially resolved as long as their Airy discs do not come closer than their radii (Figure 1b-c).

In photography, diffraction has a direct effect on the spatial resolution of the final image. Even when everything is perfectly in focus, too small an aperture gives rise to bigger Airy disks and hence a lower spatial resolution. In this way, the physical nature of electromagnetic radiation (and more specifically the principle of diffraction) sets a fundamental limit by blurring the image, preventing an exact point-for-point copy of the real-world scene. In most situations, the optical PSF (PSF_{opt}) deviates from this diffraction-limited Airy pattern due to a number of reasons such as the above-mentioned lens aberrations or erroneous focusing. Both phenomena blur an image spot even more.

Finally, there is an optical low-pass filter positioned just before the imaging sensor of most cameras. By eliminating spatial frequencies above the Nyquist frequency to prevent aliasing, this filter adds a last but significant portion of optical blurring. The PSF_{opt} integrates all these blurring steps. Since a scene is a collection of countless point sources, each of these sources generates its PSF_{opt} with an amplitude proportional to the source’s radiance. Because incoming irradiance gets spatially sampled by the detector, every pixel results from the integration of many optical PSFs over the imaging sensor’s photosite.

2.1.2 The imaging system’s PSF

However, Figure 2 shows that the PSF of the optics (PSF_{opt}) is not the only contribution to the camera system’s PSF (PSF_{sys}). An additional blurring component, corresponding to the response of the sensor itself (PSF_{sen}) also contributes to the overall PSF_{sys} of the imaging system. Finally, the image motion PSF (PSF_{mot}) constitutes a third component, resulting from the fact that the imager might move during the exposure (e.g. aerial imaging but also user tremor). In this way, the imager’s PSF_{sys} can be written as a convolution of the individual components:

$$PSF_{sys}\,(x, y) = PSF_{opt} * PSF_{mot} * PSF_{sen} \tag{1}$$

in which x and y are the spatial coordinates in two-dimensional image space. These spatial coordinates indicate that the PSF_{sys} depends on the position in the imaging plane. It is very typical for a system’s PSFs to degrade and become more non-symmetrical with increasing distance from the optical axis (because lens imperfections tend to be larger there).

Equation (1) also makes clear that an entire image is composed out of different PSFs and the digital camera itself blurs the incoming analogue radiance signal. In scientific terms, one can say that the image of an object is the convolution of the object’s radiance and the spatially varying point spread function of the complete imaging system (Jensen, 1968).

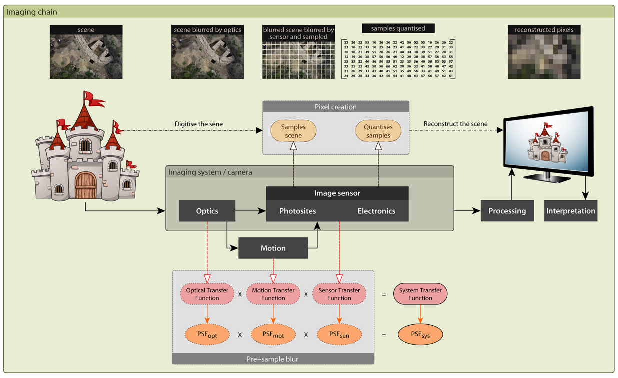

Figure 2. The complete imaging chain (including sampling and quantisation) and the individual blurring stages.

2.2 Influencing the PSF for IBM

When acquiring photographs for IBM purposes, one starts typically with a specific camera and lens. Given those and the requirements of the final output (e.g. a 1/100 map or a 3D surface model in which 1 mm details are visible), the necessary image acquisition parameters are computed and kept fixed throughout the whole imaging sequence. This means that the photographer can only influence the PSF_{sen} by opting for another imaging sensor. However, she/he can largely influence both the optical and motion PSFs. The latter can be minimised by using a tripod and remote shutter release (that is if the imaging platform itself does not vibrate) or a high shutter speed (when handholding the camera). In addition, a high shutter speed also counteracts any motion blur resulting from object motion. However, the majority of IBM projects deal with static objects, so freezing object motion should never be an issue.

Finally, the lens aperture can control the optical PSF (which varies all over the imaging sensor). As mentioned before, smaller apertures (indicated by larger *f*-numbers such as *f*/16 or *f*/22) lead to more substantial blurring effects due to diffraction. So, one might be tempted to use the lens wide open (for instance at *f*/2.8). However, too small of an aperture is contra-productive when minimising lens defects. The effects of spherical aberration, coma, astigmatism, field curvature and vignetting can be reduced to a varying extent by stopping down a lens. Since a specific aperture thus influences both positively and negatively the scene rendering, every lens has one *f*-stop or a small range of apertures at which a balance between lens aberrations and diffraction is reached. As a rule of thumb, this ideal opening is found by stopping down the lens two or three stops from wide open. This means that an *f*/2.8 lens usually delivers maximum image quality around *f*/5.6, meaning that all scene points on which the lens was focused will be rendered the sharpest possible in the resulting photograph.

However, ultimate quality and sharpness of the focused image portion is seldom the most important feature for IBM photo sets, but rather the Depth of Field (DoF). DoF is defined as the range of object distances over which objects appear acceptably sharp in the photograph (Figure 3 and the top row of Figure 5).

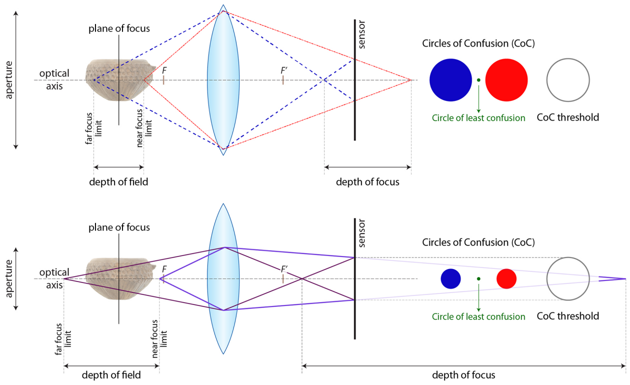

Figure 3. The influence of aperture size on the DoF, the depth of focus and the far and near focus limits.

2.3 Depth of field and defocus

Since a photographic lens is unable to render an object point as a perfect point in image space due to several lens aberrations and the effect of diffraction, the terms **blur circle** or **Circle of Confusion** (**CoC**) have been coined. As described above, the exact dimension of every CoC is given by the PSF_{opt} at that place and thus a function of diffraction, aberrations and correct focus. Usually, the PFS_{opt} is computed for a lens that is perfectly focused on an object point. In such a situation, its corresponding image point will be conjugate and located exactly on the sensor plane. Although lens aberrations and diffraction still limit the smallest size of this image spot, it is denoted the **circle of least confusion** since it is the smallest, least blurred spot reproducible by that lens (Katz, 2002). For a perfect diffraction-limited but aberration-free lens, it equals the Airy disk. Object points that are distant from the plane of focus do not come to a perfect focus on the sensor, but before or after it. In both cases, their CoCs exceed the circle of least confusion.

The upper part of Figure 3 shows that these objects spots are still perceived as sharp as long as their CoCs remain smaller than an established CoC threshold. The lower part of Figure 3 illustrates how the same object points are imaged as much smaller spots when photographed with a smaller aperture. By limiting the angle of incident radiation, all the CoCs decrease. Moreover, objects points can lie much further from the plane of focus before their CoC surpasses the CoC threshold. As a result, there is a significant increase in the distances in object space (the DoF) over which objects remain acceptably sharp in the photograph. Stopping down the lens will thus always maximise the DoF. The image–space conjugate of DoF is often termed **depth of focus**. The object distances that correspond to the object points whose CoCs are identical to the CoC threshold are denoted the **far** and **near focus limits**. Any object point that lies before the near focus limit or behind the far focus limit will be perceived as an unsharp spot. In technical terms, it means that the PSF_{opt} of those defocused points is much broader.

These far and near focus limits are not abrupt transition zones. Moreover, their exact values do not only depend on aperture, but also the field of view (given by the object distance and focal length) and the CoC threshold that is applied (Verhoeven, 2016). Often, one will find that the standard CoC threshold equals 0.03 mm, an antiquated dimension specified for particular conditions (Figure 4). First, it assumes that a human observer with normal vision looks from 25 cm at a 20 cm by 30 cm print from a 35 mm negative (i.e. full frame) (Ray, 2002). If all these conditions are satisfied, the observer should perceive any spot with a size smaller than 0.25 mm as a sharp point. This spot of 0.25 mm translates to the 0.03 mm threshold in the sensor plane. If an observer has worse than average vision, CoCs can be larger. When dealing with smaller viewing distances, the CoC threshold has to decrease. Photographs from smaller sensors also demand smaller CoC thresholds since their dimensions necessitate a larger magnification to equal a 20 cm by 30 cm print.

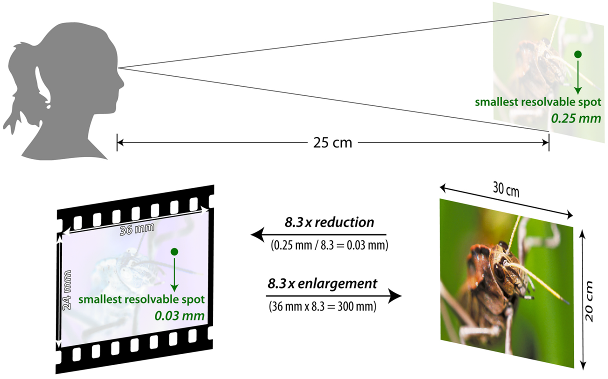

Figure 4. The commonly used (but outdated) CoC threshold of 0.03 mm is determined by human visual acuity.

https://doi.org/10.5194/isprs-archives-XLII-2-1149-2018

The consensus in IBM is that an object's surface can only be adequately constructed if all the object's dimensions fall within the depth of field of the imaging setup (Figure 3). Proper image acquisition for IBM purposes, therefore, involves the computation of an aperture for a scene-encompassing DoF. Scene portions outside the DoF are **out-of-focus/defocussed,** and the constructed surface will likely be partial and inaccurate. However, there are many problems with this assumption simply because DoF is a perceptual quantity without a direct relation to IBM. For example: there is no agreed upon value for the CoC threshold when computing the DoF for IBM purposes, let alone how the size of an object point's CoC impacts the accuracy and precision of the digitally constructed 3D surface.

Finally, it remains important to note that defocus *sensu stricto* occurs as soon as the object and image-receiving surface are not conjugate. As the distance between the object point and the plane of exact focus increases, the imaged object points become progressively more defocused. However, as noted above, defocus commonly refers to the zone of unsharpness outside the DoF. Moreover, one should be aware that the terms **focal plane**, **focus plane** and **plane of focus** have different meanings in photogrammetric computer vision, photography and optics. Here, the plane of focus is the image–space conjugate of the sensor plane, equalling an imaginary 2D plane in front of the camera at the point of focus (Figure 3). The plane is always perpendicular to the optical axis and parallel to the sensor plane (except in bellows cameras and tilt-shift lenses).

2.4 Single image defocus mapping

Estimating the local defocus blur in still images due to the spatially varying $\mathrm{PSF}_{\mathrm{sys}}$ is useful for many different applications such as depth from defocus or image deblurring. However, this paper will assess how well- or ill-suited defocus estimating algorithms are for automatically masking a series of photographs acquired for IBM purposes. The general workflow is as follows: the algorithm computes an image-specific defocus map, after which simple thresholding creates the necessary binary image mask (see Figure 7).

In the scientific literature, two broad approaches to estimate out-of-focus image regions can be distinguished: those that rely on multiple images captured with multiple camera focus settings and those that require just one image. Given the aim of this paper, only the latter methods are taken into account. Accurately estimating defocus blur from a single image is a challenging task since blur comes in so many forms and spatially varies over the whole image area. Back in 2012, Vu and his colleagues came up with a classification for single image defocus estimation algorithms (Vu et al., 2012). They discerned three major classes: 1) edge-based methods that measure the spread of edges; 2) pixel-based methods working in the spatial domain without any assumption regarding edges; 3) transform-based methods operating in the spectral domain. More recently, Karaali combined the pixel- and transform-based methods since they both explore image patches. He contrasted these patch-based approaches with edge-based methods (Karaali and Jung, 2018).

2.4.1 Edge-based methods

Many single-image defocus estimation approaches involve measuring the spread of edges as they can often be considered good blur indicators. This class of methods usually consists of two broad steps. First, a specific algorithm extracts a **sparse defocus map** by detecting edges in the original image and estimating the amount of spatially varying defocus blur at these edge locations. Such a sparse defocus map was for the first time obtained by Elder and Zucker using the first- and second-order derivatives of the input image (Elder and Zucker, 1998).

Since the pioneering work of Zhuo and Sim (2011), many edge-based methods (Tang et al., 2013; Chen et al., 2016; Liu et al., 2016; Karaali and Jung, 2018) rely on gradient magnitude ratios, obtained by reblurring the input image and computing the amount of defocus blur from the ratio between the gradients of the input and those reblurred images. This and many other sparse defocus estimators assume an isotropic 2D Gaussian PSF, although an optical or system's PSF is never truly Gaussian. A single variable such as radius then parameterises the spatially varying spread of the PSF.

Second, and using the original image as guidance, a propagation method is applied to the sparse data to yield a **complete/full defocus map** for the whole image. For the PSF-based methods, this full defocus blur map is then merely a 2D map of the PSF spread parameter, indicating for every image pixel the degree of defocus blur (Figure 5). Bae and Durand achieved the first full defocus map by propagating the Elder and Zucker blurriness measure to neighbouring pixels with a similar colour (Bae and Durand, 2007). The difficulty of this step lies in the preservation of blur discontinuities at edges while smoothly closing the gaps. Most existing techniques apply a slow Laplacian-based interpolation scheme, which makes dense defocus map extraction very time-consuming for images with large pixel counts. Faster approaches relying on superpixels (Chen et al., 2016), the fast guided filter (Andrade, 2016; Karaali and Jung, 2018) or sparse blur map downsampling (Kriener et al., 2013) have recently been introduced as well.

2.4.2 Patch-based methods

In contrast to two-stage edge-based methods, this second class estimates the full defocus map directly at all pixels, either in the spatial or frequency domain. The real pioneers of these approaches are Chakrabarti and his colleagues, who explored the convolution theorem for blur identification (Chakrabarti et al., 2010). Their approach estimates the likelihood of a small image neighbourhood being blurred by a given candidate PSF. To that end, the method involves a decomposition step to isolate localised frequency components of an image. Given that this approach can only detect optimal PSFs from a limited number of candidates, Zhu and colleagues proposed an improvement for estimating the PSF scale at every pixel using a more general local frequency component analysis in the continuous domain (Zhu et al., 2013). In addition, the method also takes smoothness and colour edge data into account. Later, machine learning approaches were introduced to infer the appropriate radius of the PSFs at every pixel (D'Andrès et al., 2016). Entirely different patch-based approaches have also been developed by Vu et al. (2012) and Yi and Eramian (2016).

3. METHODS

3.1 Defocus mapping toolbox

So far, forty different methods for mapping defocus blur have been identified in the literature. Of these forty, it was possible to obtain the MATLAB code for only seventeen algorithms, although two presented unsolvable errors. Non-MATLAB code could be retrieved for two other methods. The authors of the remaining twenty-one methods were either unable to share the code (one due to copyright issues and another one because he lost the source code) or simply unwilling (one person gave a written confirmation; the other eighteen developers simply ignored all attempts by the author to contact them).

To properly test the fifteen working approaches, a bespoke MATLAB toolbox was programmed (freely available from the author). Via the graphical user interface, one can choose any of those fifteen defocus mapping methods to compute a single image mask or a whole series of masks for an extensive image collection. Since all possible parameters of every single method are accessible, users can further fine-tune every method for their specific image(s). In addition, the toolbox often includes more variety in the variables' values than initially provided by the algorithm developers (e.g. the user can choose from eight different edge extraction methods). After (an interactive) thresholding yields the binary mask, the latter can be saved to enable its further use in many of the current IBM packages.

Since it is not feasible to report on all possible methods in the context of this paper, only three state-of-the-art but completely different edge-based methods will be assessed: Andrade (2016); Chen et al. (2016); Karaali and Jung (2018). Despite the power of some recent patch-based methods, edge-based approaches remain a very attractive choice. Moreover, many patch-based methods are computationally expensive. It would have been appropriate to include the work of Zhuo and Sim (2011) as well as the later work of Tang et al. (2013). Since both algorithms rely on a Laplacian Matting scheme, they would need more than the available 24 GB of RAM to deal with 4+-megapixel images.

3.2 Image sets

Many defocus estimation algorithms perform reasonably well on the default sets of idealistic images (commonly smaller than 0.5 megapixels and featuring clear out-of-focus regions). Since this paper wants to find out how these algorithms quantitatively (running time) and qualitatively (masking accuracy and robustness) behave on various realistic datasets, five datasets of four photographs have been created. These datasets encompass a large portion of the defocus variety one can encounter when dealing with (cultural heritage-related) IBM pipelines.

3.2.1 DoF dataset

This set consists of photographs captured with a Nikon D750, a 24-megapixel full-format camera equipped with a 60 mm lens. The images depict Datacolor's SpyderLENSCAL, an auto-focus calibration target for digital reflex cameras. Apart from the aperture, the exposure and white balance settings remained unaltered during data acquisition. More specifically, every image features a two-stop aperture difference with the next image: *f*/2, *f*/4, *f*/8 and *f*/16. The corresponding increase in DoF is easy to spot on the target's ruler. Since all images feature an identical object, illumination and post-processing, they are suited to assess the accuracy of mapping aperture-specific defocus.

3.2.2 Globus of Leonardo da Vinci

Image collection two holds four images of an ostrich egg globe attributed to Leonardo da Vinci. These images, which were acquired to digitally unfold the globe into a projected map (Verhoeven and Missinne, 2017), were also shot with a Nikon D750. In this case, a prime 105 mm lens was used, and *f*/22 dialled for all exposures. Although such a small aperture induces evident diffraction softening, it was still essential to provide sufficient DoF. This image series reflects a typical studio setup, in which an artefact is positioned in front of a neutral background. Since such backgrounds seldom feature many discernible edges and are usually thrown out-of-focus, they should be easily picked up by the defocus mapping algorithms. An element of difficulty in this series is the light-coloured egg surface which exhibits a rather low colour contrast with the background for surface portions devoid of drawings.

3.2.3 Stonehenge

The third set of images depicts some parts of the famous prehistoric monument at Stonehenge, England. Photographs were obtained using a 12-megapixel Nikon D300s using a variety of focal lengths, but keeping the lens aperture fixed at its ideal *f*/8 aperture. These photographs feature nicely textured stones against a backdrop of grass and clouds. The latter two also feature some edges, although much less pronounced than the ones of the stones. As such, the defocus mapping algorithms should at least be able to separate the stones from the sky.

3.2.4 Monastery of Saint Peter

The fourth dataset comprises a collection of Nikon D300s images depicting an old Benedictine monastery at the island of Sv. Petar, Croatia. As two sides of this building are in very close proximity to trees, the subject distance was minimal. Even though a 17 mm lens fixed at *f*/8 was used, the end of the walls was always thrown out-of-focus on the convergent images as well as the photographs of the building corners. Although the stone wall and the surrounding trees exhibit many great edges, it will be interesting to see if any algorithm can still mask those pixels that correspond to portions of the scene that were clearly outside the DoF.

3.2.5 Aerial images of Montarice

The last set of images were shot with an analogue Canon EOS 300D camera from a low-flying Cessna 172 Skyhawk aeroplane to document Montarice, a hilltop plateau in central Adriatic Italy near Porto Recanati. The aperture and focal length are undocumented. After digitising the diapositives with a Nikon SUPER COOLSCAN 5000 ED at 4000 samples per inch, they have been used for an enhanced interpretative mapping of the buried archaeological features (Verhoeven and Vermeulen, 2016). All four images display a defocused part of the aeroplane's wing strut, something that can happen when photographing from the air using a wide-angle lens. However, such obstructions of the line-of-sight from the camera to the scene are often encountered in other situations (e.g. branches of trees in front of the camera). When such obstructing objects are very close to the camera, they usually fall outside the DoF and are rendered blurry (as is the case with the wing strut).

4. RESULTS

Any evaluation of defocus estimation is somewhat challenging. To make the comparison fair, all methods were run using an edge map computed with the Canny edge detector (with a threshold of 0.1 and one standard deviation).

4.1 Accuracy

Figure 5 displays the output of the selected algorithms on the first set of images (method-specific parameters were kept at their default values). It is clear that the final defocus maps look utterly different amongst the three methods, although they started from the same set of extracted edges. However, this test with default parameter values is not to check the quality of the defocus map (which can always be optimised by fiddling around with the parameter values), but to assess whether or not the defocus estimator is accurately picking up aperture-related out-of-focus blur in images that feature clear edges. As can be seen, only the methods from Andrade and to a lesser degree Karaali and Jung perform satisfactorily.

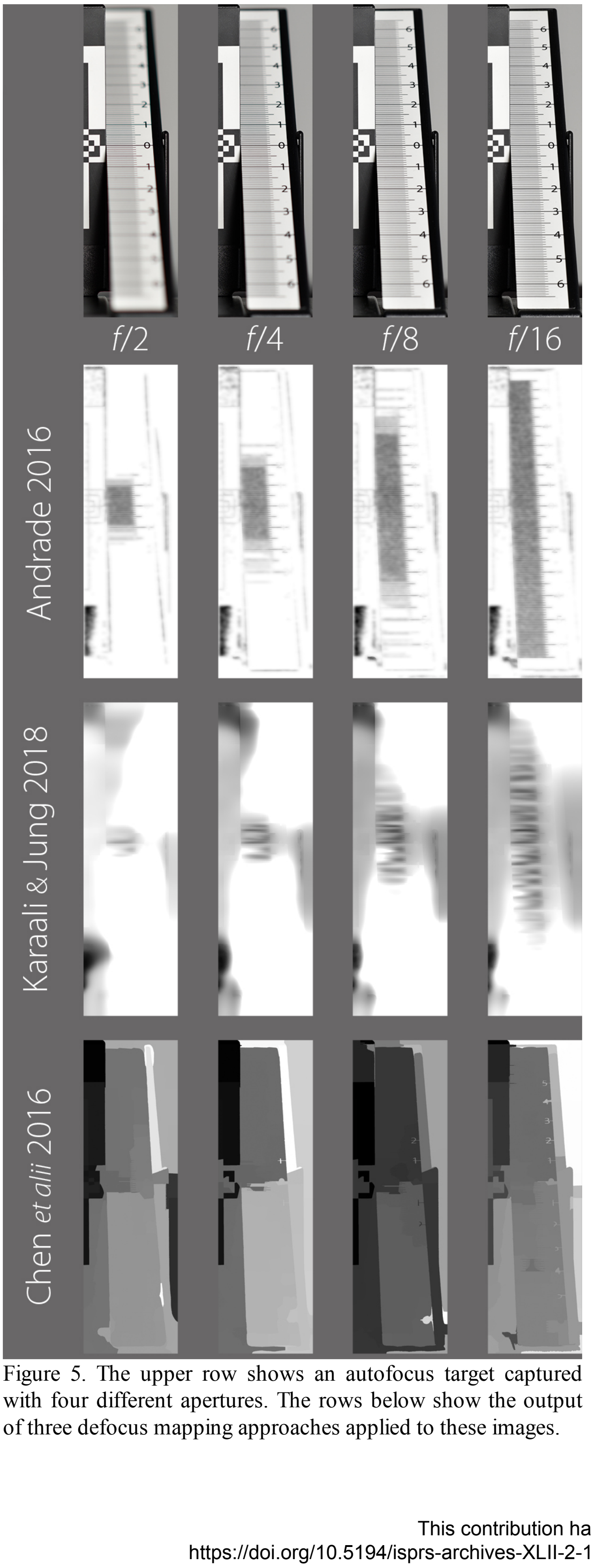

Figure 5. The upper row shows an autofocus target captured with four different apertures. The rows below show the output of three defocus mapping approaches applied to these images.

When observing Figure 6, one gets a slightly different view. Here, the first and third photograph of the remaining four image sets are displayed along with a reference binary mask. Throughout all examples, one can see that the masks from Andrade’s approach are generally slightly too small. This results from the threshold value used for the binary segmentation. Lowering this value would generate masks that were more true to the real object boundaries, but they also resulted in a few holes on the object itself. Overall, the defocus maps from this method are by far the cleanest and most accurate of all three methods. For datasets two and three, the other two methods produced results that more or less approach the reference masks, while their output was useless for dataset four and five.

4.2 Running time

Figure 6 also marks the average time it took a specific method to yield a defocus map (on a Windows 7 PC with an Intel Core i7-980X processor and running a 64-bit version of MATLAB R2017b). It is clear that the method of Chen et al. is noticeably faster than any other method, but this comes at a penalty of decreased accuracy. The algorithm of Karaali and Jung sits on the other far end of the running time spectrum. This has two implications. First, processing a series of photographs becomes impractical when the algorithm requires more than half an hour to mask one image. Second, it is almost impossible to fine-tune the parameters for this method, since one has to wait an eternity between every fine-tuning step. Finally, the method of Karaali and Jung also had the highest running time variance, both within the same dataset as well as between different pixel counts.

4.3 Robustness

Thanks to the MATLAB toolbox, it was straightforward to test the influence of most function-specific parameters, even if the original coding did usually not foresee this. Once a good set of parameter values was found for the first image of every series (which always equals the first image of the pair displayed in Figure 6), the algorithm was run on the remaining three images without altering any value. In this way, the defocus approach was tested for robustness while also assessing its invariance to changing camera viewpoints or object distances. Here, only the method of Andrade yielded satisfactory results, whereas the other two methods often failed to perform reliably with the initial settings.

However, even the settings of Andrade’s method had to be fine-tuned separately for every image collection. Although this was done on one image, a few fine-tuning iterations might easily take one hour. As mentioned above, this time aspect made it impossible to properly fine-tune the parameters values of the Karaali and Jung approach. Working on versions with a lower pixel count does not make sense, as edges change in subsampled images. Maybe an initial subsampling of the image followed by edge-aware upsampling of the defocus map – as was presented by Kriener et al. (2013) – could be a solution.

5. DISCUSSION

5.1 Critical observations

Writing this paper has been a frustrating experience because of many factors. First, some defocus mapping approaches have an enormous memory footprint. As an example: even after clearing many unnecessary temporary variables in the code, the method

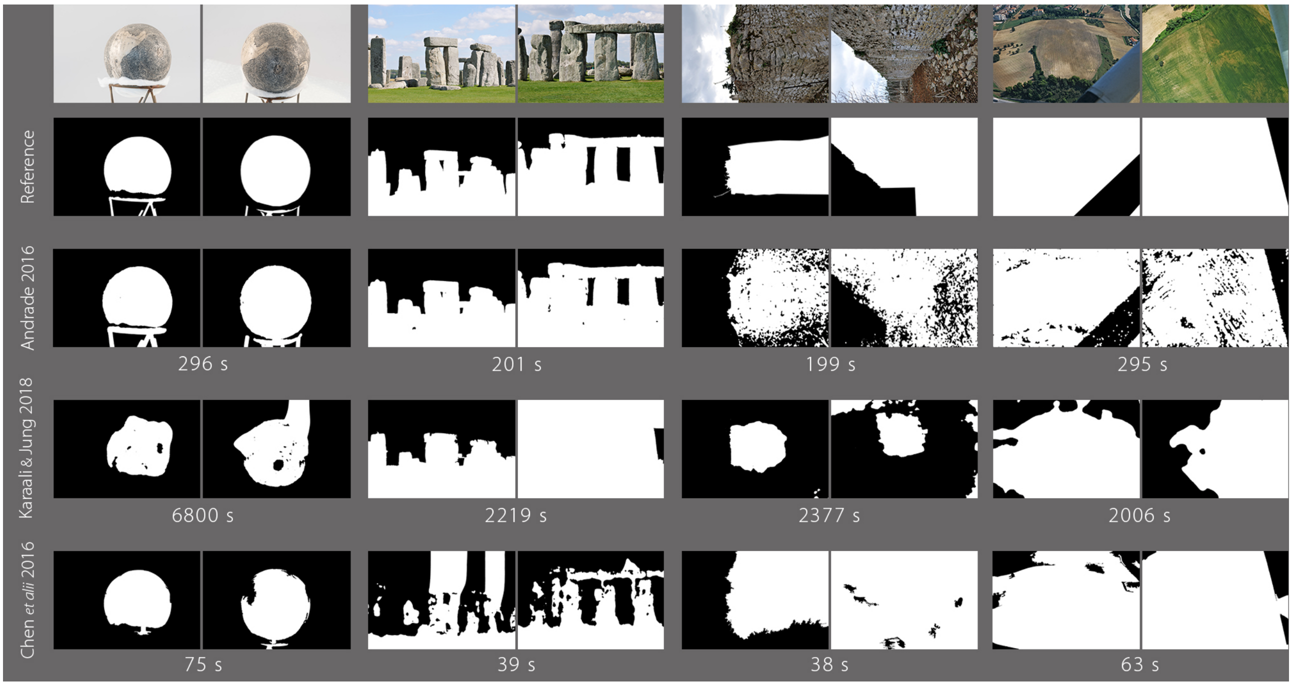

Figure 6. Comparison of the automatically generated binary image masks with a reference mask across four image sets. The masks were generated after thresholding a full defocus map that was computed by three different methods (mentioned on the left). The average running time for one type of image is displayed for every method. Finally, black means masked, white means not masked.

by Zhuo and Sim (2011) consumed all of the 24 GB of computer RAM when ran on a 4-megapixel image. Although it is fair to say that these authors mentioned this issue in their code, most developers do not. They might not even be aware, because these algorithms have never been tested on proper images. Most defocus mapping papers showcase their results on images of either 360 pixels by 360 pixels (0.13 megapixels) or 800 pixels by 600 pixels (0.48-megapixels). However, the use of such unrealistically small images is something that plagues most of the academic image processing community.

Second, it is also apparent that those developing a new image processing algorithm usually stick to images (and metrics) that support their claims. In the case of these defocus mapping algorithms, it was often observed that photographs other than the ones used in the initial paper completely undermined the claims made of the supposed superiority of that specific algorithm. Most likely, this is the reason why so many authors refuse to share their code.

Apart from those issues, the paper also highlighted what seems to be the main limitation of edge-based defocus methods. These approaches map out-of-focus areas by determining the amount of local defocus blur at edge locations and propagate these estimates to the whole image. Obviously, the accuracy of the defocus mapping output is heavily dependent on the extracted edges. To accomplish this, the image processing community offers a variety of edge detectors, each of them with one or more tuneable parameter. However, most developers of edge-based defocus mapping methods stick by default to the Canny or Sobel edge detector; empirical parameter values should then find many edges in every region of a specific image collection and produce a (visually) good result. However, even if one finds optimal parameter for one set of images, that approach might fail when the pixel count is changed (e.g. the method is applied on another photograph with much more/fewer pixels and thus better/worse defined edges). Moreover, not all extracted edges are reliable locations to estimate defocus blur. Third, edge-based methods are sensitive to image noise and edge interference.

Given this, it is striking that so little research has gone into the edge detection part of these edge-based defocus mapping algorithms? Only recently, some authors started to address the importance of estimating reliable edges and re-blurring scales (Karaali and Jung, 2018), although their method has here been shown to be a poor performer on large, real-world images. The approach taken by Andrade (2016) is also somewhat unique in the defocus mapping community. He included edge pruning to remove wide edges in blurred regions with a follow-up edge-diffusion step – based on the heat diffusion principle. Given the accuracy and consistency of his approach, it is striking that none of the other defocus mapping papers mentions his work. The reasons for this might range from never having encountered the paper (which is sloppy research), never obtained the source code (although it was received after mailing the author) or neglecting his algorithm because it usually outperforms most other methods (which is academic cheating).

5.2 Usability and future research

On a positive note, this paper has shown that some edge-based methods might be very well suited for masking homogenous, edge-free regions such as skies and studio backgrounds. To a lesser extent, they could even be applied to mask parts of an object that fell outside the camera's DoF or items obstructing the line-of-sight from the camera to the scene of interest. If developers manage to code slightly more accurate, robust and speedier defocus algorithms, it seems likely that they can in the near future assist the user in masking specific areas in extensive image collections. One could argue that these zones are also easy to mask out manually (or why one should mask them in the first place), but running such an algorithm at a time the computer sits idle might still save many hours when dealing with large sets of images. Even in suboptimal cases, the right algorithm can at least provide a base mask to work from. In all fairness, this view might be altered once a more in-depth review of all edge- and patch-based methods is undertaken. Nonetheless, Figure 7 indicates that Andrade's edge-based method is at least as accurate (if not more) than the state-of-the-art patch-based method by Golestaneh and Karam (2017).

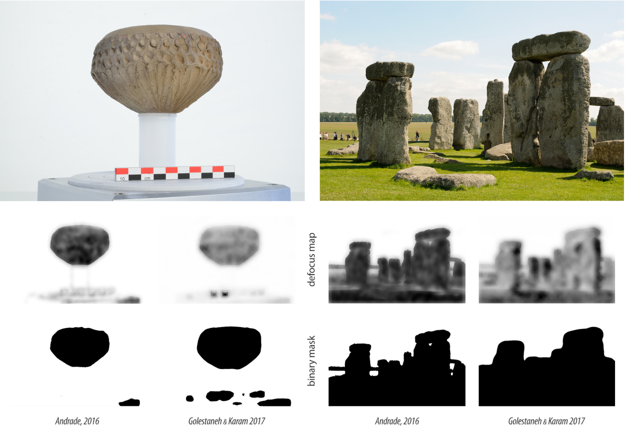

Figure 7. Defocus map estimation of two different images using an edge-based and a recent patch-based approach.

6. CONCLUSION

Many defocus blur estimation methods claim a good trade-off between accuracy and runtime. However, these claims often go unchallenged and are usually based on an unrealistic image set. In this paper, three state-of-the-art edge-based algorithms have been tested on large, real-world images for automatically masking out-of-focus areas. These tests revealed the issues that most of these algorithms face. However, at least one method showed a level of accuracy and robustness which could render it useful for future application.

REFERENCES

Andrade, J., 2016. Defocus Map Detection Using a Single Image. In: *Proceedings of the 2016 International Conference on Computational Science and Computational Intelligence (CSCI 2016)*. IEEE, Los Alamitos, pp. 777–780.

Bae, S., Durand, F., 2007. Defocus Magnification. *Computer Graphics Forum* 26 (3), pp. 571–579.

Chakrabarti, A., Zickler, T., Freeman, W.T., 2010. Analyzing spatially-varying blur. In: *Proceedings of the 23rd IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2010)*. IEEE, pp. 2512–2519.

Chen, D.-J., Chen, H.-T., Chang, L.-W., 2016. Fast defocus map estimation. In: *Proceedings of the 2016 IEEE International Conference on Image Processing*. IEEE, Piscataway, pp. 3962–3966.

D'Andrès, L., Salvador, J., Kochale, A., Süsstrunk, S.E., 2016. Non-Parametric Blur Map Regression for Depth of Field Extension. *IEEE Transactions on Image Processing* 25 (4), pp. 1660–1673.

Elder, J.H., Zucker, S.W., 1998. Local scale control for edge detection and blur estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20 (7), pp. 699–716.

Golestaneh, S.A., Karam, L.J., 2017. Spatially-Varying Blur Detection Based on Multiscale Fused and Sorted Transform Coefficients of Gradient Magnitudes. In: *Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017)*. IEEE, Los Alamitos, pp. 5800–5809.

Hecht, E., 2002. *Optics*, 4th ed. Addison Wesley, San Francisco.

Jensen, N., 1968. *Optical and Photoghraphic Reconnaissance Systems*. John Wiley & Sons, New York - London - Sydney.

Karaali, A., Jung, C.R., 2018. Edge-Based Defocus Blur Estimation With Adaptive Scale Selection. *IEEE Transactions on Image Processing* 27 (3), pp. 1126–1137.

Katz, M., 2002. *Introduction to geometrical optics*. World Scientific, River Edge.

Kriener, F., Binder, T., Wille, M., 2013. Accelerating defocus blur magnification. In: *Proceedings of IS&T/SPIE Electronic Imaging: Multimedia Content and Mobile Devices. SPIE - IS&T*, Bellingham, Springfield, 86671Q.

Liu, S., Zhou, F., Liao, Q., 2016. Defocus Map Estimation from a Single Image based on Two-parameter Defocus Model. *IEEE Transactions on Image Processing* 25 (12), 5943–5956.

Ray, S.F., 2002. *Applied photographic optics. Lenses and optical systems for photography, film, video, electronic and digital imaging*, 3rd ed. Focal Press, Oxford.

Tang, C., Hou, C., Song, Z., 2013. Defocus map estimation from a single image via spectrum contrast. *Optics Letters* 38 (10), pp. 1706–1708

Verhoeven, G., 2016. Basics of photography for cultural heritage imaging. In: Stylianidis, E., Remondino, F. (Eds.), *3D Recording, Documentation and Management of Cultural Heritage*. Whittles Publishing, Caithness, pp. 127–251.

Verhoeven, G., Missinne, S., 2017. Unfolding Leonardo da Vinci's globe (AD 1504) to reveal its historical world map. In: *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* IV-2/W2, pp. 303–310.

Verhoeven, G., Vermeulen, F., 2016. Engaging with the Canopy. Multi-Dimensional Vegetation Mark Visualisation Using Archived Aerial Images. *Remote Sensing* 8 (9), article 752.

Vu, C.T., Phan, T.D., Chandler, D.M., 2012. S3. A spectral and spatial measure of local perceived sharpness in natural images. *IEEE Transactions on Image Processing* 21 (3), pp. 934–945.

Yi, X., Eramian, M., 2016. LBP-Based Segmentation of Defocus Blur. *IEEE Transactions on Image Processing* 25 (4), pp. 1626–1638.

Zhu, X., Cohen, S., Schiller, S., Milanfar, P., 2013. Estimating spatially varying defocus blur from a single image. *IEEE Transactions on Image Processing* 22 (12), pp. 4879–4891.

Zhuo, S., Sim, T., 2011. Defocus map estimation from a single image. *Pattern Recognition* 44 (9), pp. 1852–1858.