Supplementary Information for: "Image Difference Metrics for High-Resolution Electron Microscopy"

Manuel Ederer[a], Stefan Löffler[a]

[a] *University Service Centre for Transmission Electron Microscopy, TU Wien, Wiedner Hauptstraße 8-10/E057-02, 1040 Wien, Austria*

Image difference algorithms

SSIM

In this section we present the structural similarity index measure from [1] and [2] in more detail including all parameters used in our calculations. For each pair of points $\mathbf{x}$ from image A and $\mathbf{y}$ from image B a local batch of pixels of equal size and shape around the points is taken. On this batch the luminance is compared

$$l(\mathbf{x},\mathbf{y}) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, \tag{1}$$

consisting of the local image signal mean

$$\mu_x = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i, \tag{2}$$

where $\mathbf{x}_i$ are the points of the batch around $\mathbf{x}$ and C_1 is a small constant to avoid instability when $\mu_x^2 + \mu_y^2$ is close to zero. In the next step the local mean is removed from the signal and subsequently the contrast is compared. The contrast comparison function

$$c(\mathbf{x},\mathbf{y}) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x + \sigma_y + C_2} \tag{3}$$

takes a similar form to the luminance comparison function Eq. 1 with the local signal variance

$$\sigma_x = \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{x}_i - \mu_x)^2 \tag{4}$$

and a small constant C_2. In the next step of the workflow the image signal is normalised by its own standard deviation. Lastly, the structure comparison function is defined as

$$s(\mathbf{x},\mathbf{y}) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \tag{5}$$

with the local correlation coefficient σ_{xy} between $\mathbf{x}$ and $\mathbf{y}$

$$\sigma_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{x}_i - \mu_x)(\mathbf{y}_i - \mu_y) \tag{6}$$

and a small constant C_3. Combining all three comparison functions of Eq. 1, Eq. 3 and Eq. 5 and choosing $C_3 = C_2/2$ results in the original form of the SSIM index

$$\begin{aligned} SSIM(\mathbf{x},\mathbf{y}) &= l(\mathbf{x},\mathbf{y})c(\mathbf{x},\mathbf{y})s(\mathbf{x},\mathbf{y}) = \\ &= \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x + \sigma_y + C_2)}. \end{aligned} \tag{7}$$

In order to achieve more robustness against image noise, the similarity index is extended to the complex wavelet transform domain [2]. The continuous

Email address: manuel.ederer@tuwien.ac.at (Manuel Ederer)

complex wavelet transform of a real signal $f(\mathbf{x})$ is given by

$$F(s,\mathbf{p}) = \frac{1}{|s|^{1/2}} \int_{-\infty}^{\infty} f(\mathbf{x}) \psi \left(\frac{\mathbf{x}-\mathbf{p}}{s} \right)^{*} \mathrm{d}\mathbf{x} \quad (8)$$

where $\psi(\mathbf{x})$ is a continuous, complex function called the mother wavelet from which the daughter wavelets are constructed by shifting and scaling. In our case the signal $f(\mathbf{x})$ represents the image and the 2-dimensional parameter $\mathbf{x}$ the pixel positions. The position parameter $\mathbf{p}$ is used to scan over the signal in pixel space and in the discrete case has the same range as the pixels of the image. Through the scale parameter s the wavelet transform scans the signal over multiple (spatial) frequencies. We choose s from the set $\{s|s = 1, ..., 30;\ s \in \mathbb{N}\}$. We use the complex Morlet wavelet [3] (or Gabor wavelet) defined by

$$\psi(x) = \pi^{-\frac{1}{4}} \mathrm{e}^{-\frac{1}{2}x^2} \mathrm{e}^{\mathrm{i}\omega_0 x} \quad (9)$$

where ω_0 is the center frequency of the wavelet. For each corresponding point $\mathbf{p}_0$ of the wavelet transformed images being compared, the 30 coefficients $c_i = F(s, \mathbf{p}_0)$ are used for Eq. 7. Instead of a local batch of pixels in real space, a set of coefficients of the wavelet transform is used for the calculation of mean, standard deviation and correlation coefficient. Due to the bandpass nature of the wavelet filters the coefficients have zero mean, resulting in the original form of the complex wavelet SSIM

$$\tilde{D}(\mathbf{x},\mathbf{y}) = \frac{2|\sum_{i=1}^{N} c_{\mathbf{x},i} c^{*}_{\mathbf{y},i}| + K}{\sum_{i=1}^{N} |c_{\mathbf{x},i}|^2 + \sum_{i=1}^{N} |c_{\mathbf{y},i}|^2 + K} \quad (10)$$

where $N = 30$ is the number of coefficients and K is a small, positive constant. Subsequent averaging over all points leads to a total image similarity measure, exactly as in the real space case.

SIFT

The image difference based on the scale-invariant feature transform (SIFT) [4] is calculated according to the workflow diagram in Fig. S1.

The first step consists of finding features in the reference image that are independent of global translations and rotations, noise and scaling of the image. In order to ensure the last point an image pyramid in scale space [5] is constructed. Each level, or scale, of the pyramid is an increasingly more blurred version of the original image. For a certain scale σ the scale-space representation $L(x, y, \sigma)$ of a 2-dimensional signal $I(x, y)$ is given by

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y), \quad (11)$$

where $*$ is the convolution operation in x and y, and

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (12)$$

is the Gaussian function. After an octave, i.e. the doubling of σ, the Gaussian image is resampled by a factor of 2. In each octave the difference between images of scales separated by a constant multiplicative factor k is calculated, resulting in the difference-of-Gaussian (DoG) function

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma). \quad (13)$$

Blurring an image with a Gaussian kernel suppresses only high-frequency spatial information. Thus, the DoG acts like a band-pass filter, attenuating spatial frequencies outside of the range between σ and $k\sigma$. The pyramid of Gaussians and DoG can be schematically seen in Fig. S2. In this work we have used 3 layers of DoG per octave which is also the value used in [6]. Local maxima and minima

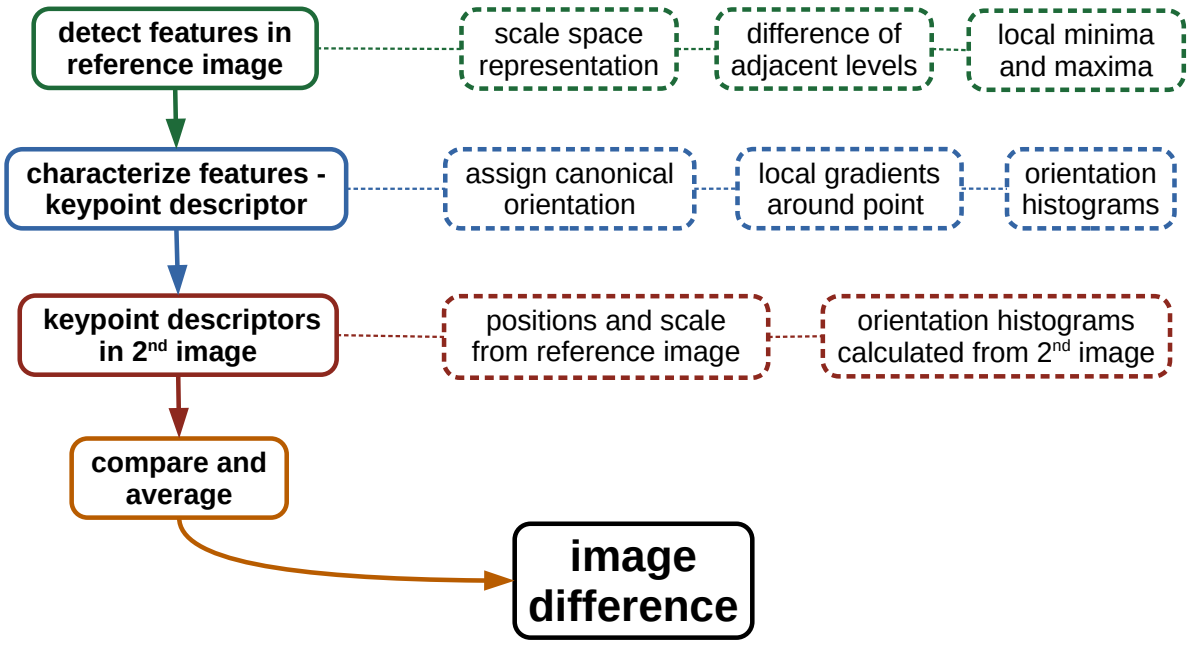

Figure S1: Workflow diagram of the SIFT measurement

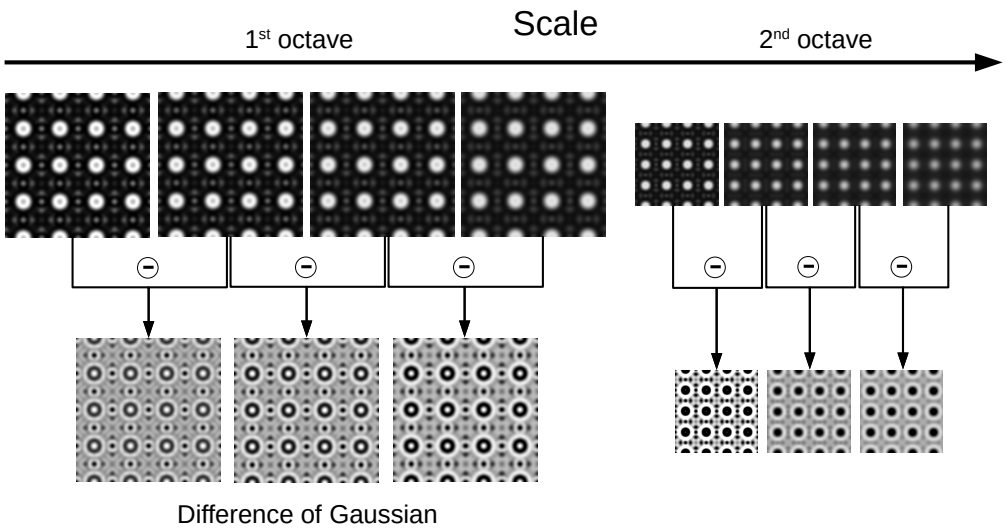

Figure S2: Pyramid of Gaussians in scale-space. Top row: each image is the result of consecutively more blurring of the original image. Neighbouring images are separated by a constant factor in scale-space. After an octave (doubling of σ) the image is resampled by a factor of two. In each octave neighbouring scales are subtracted from each other resulting in the pyramid of difference-of-Gaussian functions in the bottom row. Each pixel is then compared to its 8 neighbours at the current scale and its 9 neighbours at the neighbouring scales each in order to find local maxima and minima.

of the DoG images are detected by comparing each pixel to its 8 neighbours at the same scale and to its 9 neighbours each in the scale above and below. Afterwards, points located in areas of low contrast or along edges are discarded. The remaining keypoints are locations in scale-space marking features that are distinctive for the image.

The next step of the SIFT image difference algorithm consists of characterizing the found features, i.e. in constructing the keypoint descriptors. In order to assign a canonical orientation to the keypoint found at scale σ_0, a small circular batch of pixels around the point in the Gaussian smoothed image $L(x, y, \sigma_0)$ is taken into account. For each pixel of this batch the local gradient magnitude $m(x, y)$ and gradient orientation $\theta(x, y)$ is calculated using pixel differences:

$$\begin{aligned} m(x,y)^2 =& (L(x+1,y) - L(x-1,y))^2 + \\ &+ (L(x,y+1) - L(x,y-1))^2 \end{aligned} \tag{14}$$

$$\theta(x,y) = \arctan\left(\frac{L(x,y+1) - L(x,y-1)}{L(x+1,y) - L(x-1,y)}\right). \tag{15}$$

The local gradient angles $\theta(x, y)$ are distributed to a histogram of 36 angle bins and weighted with the respective magnitude $m(x, y)$ and by a Gaussian-weighted circular window. The highest peak of this histogram is then taken as the canonical orientation of the keypoint and all subsequent description will be relative to this orientation. For the descriptor itself a 16×16 pixel window around the keypoint of the smoothed image $L(x, y, \sigma_0)$ is used. The area is partitioned into 16 squares of 4×4 pixels and the remaining gradients are calculated, if they have not been already calculated in the previous step. In each of the 16 partitions the gradient orientation values are inserted into histograms with 8 orientation bins. In order to avoid too abrupt descriptor changes trilinear interpolation is used to distribute the values among an orientation bin and its adjacent bins. All the histograms from the partitions together then form the keypoint descriptor vector with $16 \times 8 = 128$ elements. The third step of the SIFT image difference workflow differs significantly from the original algorithm outlined in [4, 6]. In the second image we do not again detect features as described in the first step, instead we take the same positions and scale of the keypoints in the reference image. The areas around the positions are, however, taken from the (smoothed) second image and the resulting orientation histograms and descriptor vectors as well. Thus, for each feature in the reference image we arrive at a descriptor vector from the reference image and one from the second image, which are to be compared. In the final step, we take the Frobenius norm of the difference between each pair of descriptor vectors and average them, resulting in a total image difference.

Precipitate size determination

In this section we present the results of the ZrO_2 precipitate size determination using the cw-SSIM and the MSE image difference metric.

References

[1] Z. Wang, A. C. Bovik, H. R. Sheikh, E. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Transactions on Image Processing 13 (4) (2004) 600–612. doi:10.1109/TIP.2003.819861.

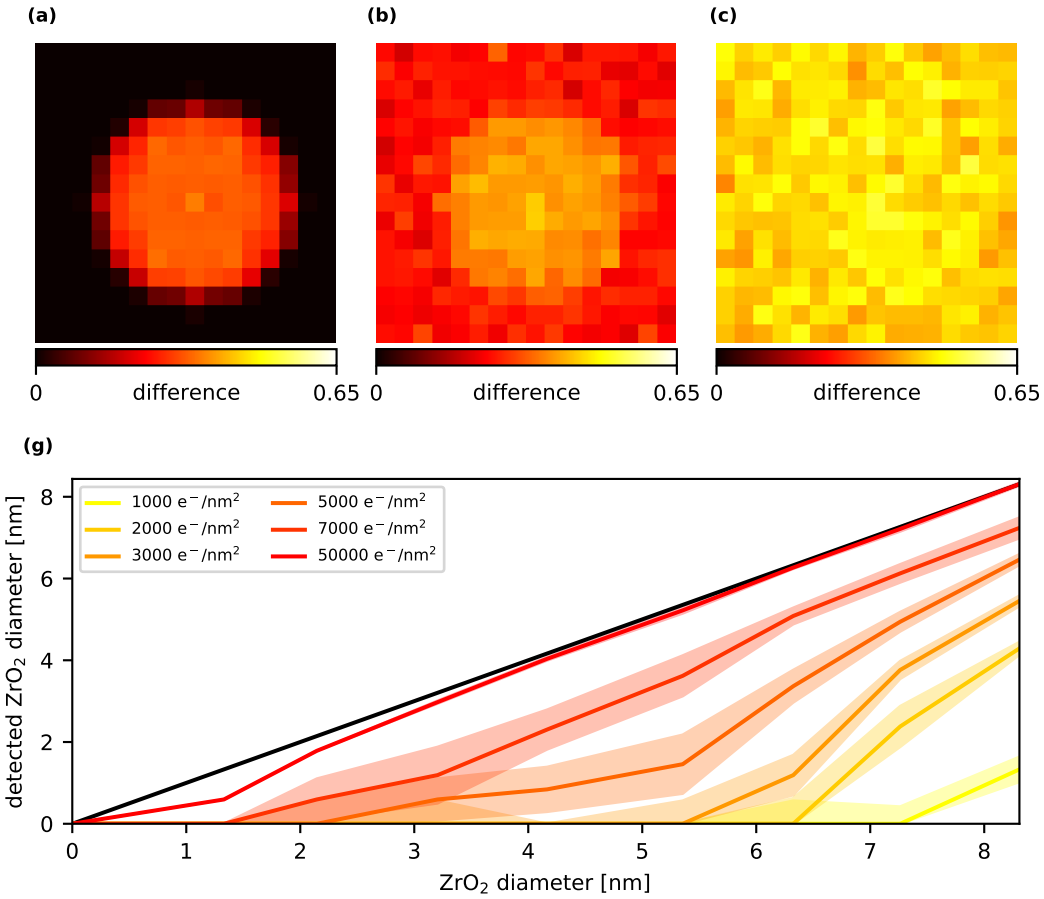

Figure S3: Same as Fig. 4 but with the cw-SSIM metric.

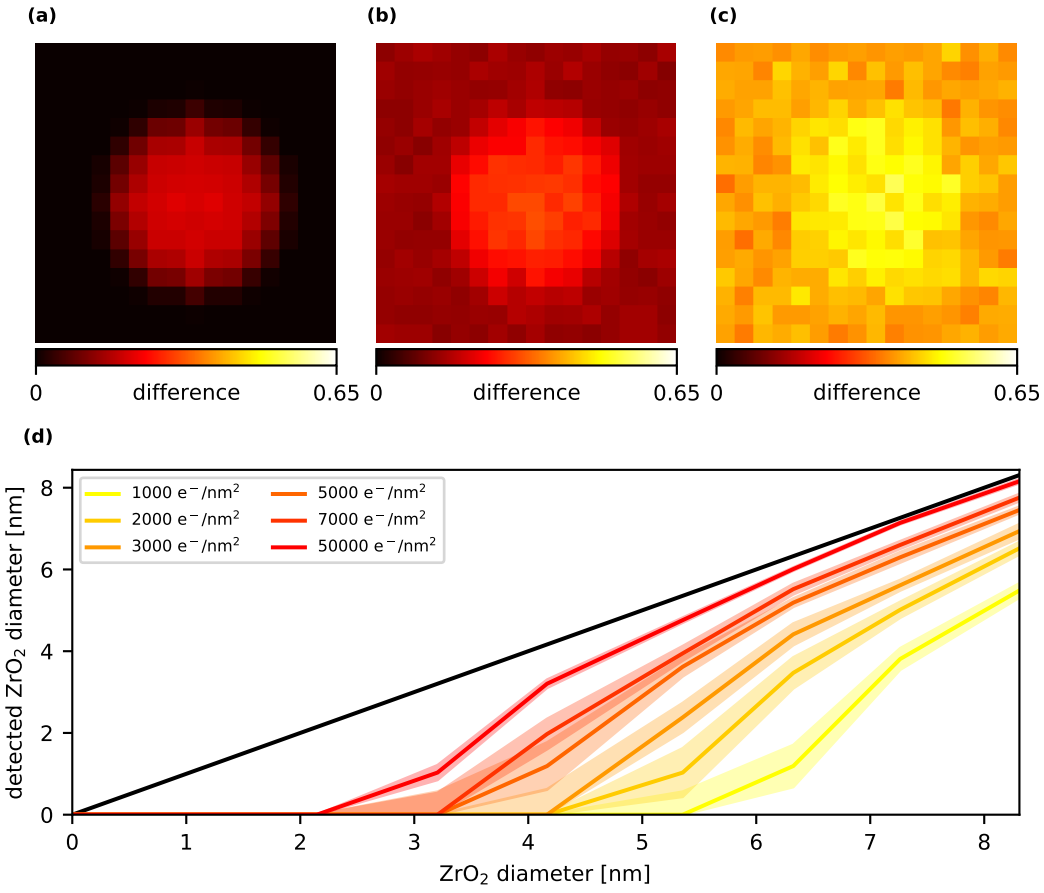

Figure S4: Same as Fig. 4 but with the MSE metric.

[2] Z. Wang, E. Simoncelli, Translation insensitive image similarity in complex wavelet domain, Acoustics, Speech, and Signal Processing, 1988. ICASSP-88., 1988 International Conference on 2 (2005) 573 – 576. doi:10.1109/ICASSP.2005.1415469.

[3] J. Ashmead, Morlet wavelets in quantum mechanics, Quanta 1 (1) (2012) 58–70. doi:10.12743/quanta.v1i1.5.
URL http://quanta.ws/ojs/index.php/quanta/article/view/5

[4] D. G. Lowe, Object recognition from local scale-invariant features, in: Proceedings of the Seventh IEEE International Conference on Computer Vision, Vol. 2, 1999, pp. 1150–1157 vol.2. doi:10.1109/ICCV.1999.790410.

[5] A. Witkin, Scale-space filtering: A new approach to multi-scale description, in: ICASSP '84. IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 9, 1984, pp. 150–153.

[6] D. G. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision 60 (Nov 2004). doi:10.1023/B:VISI.0000029664.99615.94.
URL https://doi.org/10.1023/B:VISI.0000029664.99615.94