Basic morphological operations, band-limited images and sampling

Cris L. Luengo Hendriks and Lucas J. van Vliet

Pattern Recognition Group, Delft University of Technology, The Netherlands

Abstract. Morphological operations are simple mathematical constructs, which have led to effective solution for many problems in image processing and computer vision. These solutions employ discrete operators and are applied to digitized images. The mathematics behind the morphological operators also exists in the continuous domain, the domain where the images came from. We observed that the discrete operators cannot reproduce the results obtained by the continuous operators. The reason for this is that neither the operator (the structuring element) nor the result of the operation are band-limited, and thus cannot be represented by equidistant samples without loss of information. The differences between continuous-domain and discrete-domain morphology are best shown by the dependency of the discrete morphology on sub-pixel translations and rotations of the images before digitization.
This article describes an algorithm that applies continuous-domain morphology to properly sampled images. We implemented the dilation for one-dimensional images (signals), and with it constructed the erosion, the closing and the opening. We provide a discussion on a possible extension to higher-dimensional images.

1 Introduction

1.1 Band-limited signals and uniform sampling

A large class of signals can be represented by an infinite set of equidistant samples without loss of information; that is, we can reconstruct the original signal from these samples. This class is composed of all band-limited signals in which the highest frequency that is needed to construct the signal (the cut-off frequency) is less than half the sampling frequency. This condition is called after Nyquist and/or Shannon [1, 2].

A linear filter applied to a band-limited signal produces another band-limited signal with an equal or lower cut-off frequency. This implies that such an operation can be performed on the set of uniform samples, producing the sampled version of the continuous result [3]. However, almost all non-linear filters produce non-band-limited outputs, which cannot be represented correctly by equidistant samples. Therefore, the discrete implementations of these operators do not represent their continuous counterparts. This is certainly the case for morphological operations.

1.2 Sampling morphological operations

The basic morphological operations, dilation and erosion with a flat structuring element (SE), are equivalent to, respectively, a local maximum and local minimum filter (assuming the SE is a closed set, and the function it is applied to is continuous) [4, 5]. The output at each point is defined by the maximum (or minimum) over the neighborhood defined by the SE. When applied to a band-limited image, this produces an image that is not band-limited, and therefore cannot be represented correctly by equidistant samples. This can be seen by the fact that discontinuities in the first derivative are introduced.

For certain analysis operations, e.g. a granulometry, this is not important because the output of the morphological operation must be integrated to obtain a single value. That is, we are not interested in sampling the (continuous) result. When the result of a discrete operation that produces a non-band-limited result is integrated, an error is made. To reduce this error certain tricks can be used (see [6]). For example, it is possible to interpolate the input image. This causes the result of a discrete morphological operation to produce a better approximation to the sampled version of the corresponding continuous operation. Therefore a smaller error is made when integrating the result.

Because the morphological operations are local maximum or minimum filters, the result is heavily influenced by resampling the discrete image (either for interpolation, translation or rotation). This is because it is not expected that a sample exactly hits a local maximum or minimum of a function. Resampling will cause different values of the image to be sampled, thus changing the result of the local maximum or minimum filters. This dependency on the sampling grid can also be shown by translating it with respect to the original, continuous image; see Figure 1. We will call any difference between the continuous-domain and discrete-domain results *sampling error*.

In this article we propose an algorithm that implements continuous-domain morphological operators. It works for dilations as well as erosions with flat structuring elements. Openings and closings can be constructed using these two basic operations. This method is explicitly defined for 1D images (signals). It is theoretically possible to extend this method to higher dimensions, albeit with a constraint (Section 4). Some other operations, such as the watershed transform or the morphological reconstruction, might be implemented in this framework as well, but are beyond the scope of this paper.

2 Sampling-free dilations

To reduce the sampling error of morphological operations, we need a continuous representation of the signal, a function $f : \mathbb{R} \to \mathbb{R}$ defined on an interval $[x_0, x_N]$. We must be able to

- represent band-limited signals accurately,
- represent signals with discontinuities in the first and higher derivatives, and
- obtain such a representation from a set of given samples.

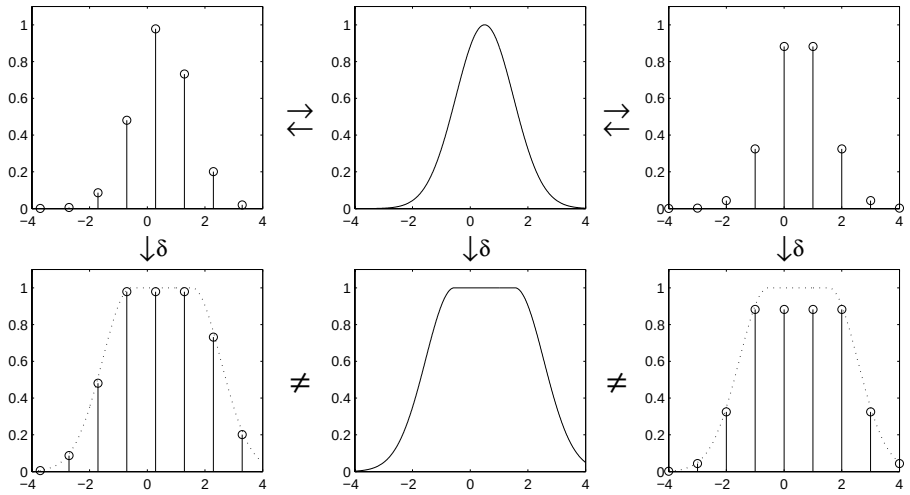

Fig. 1. The discrete dilation is not translation-invariant, as this example shows. In the middle of the top row is a continuous signal. We sample this signal twice, using uniform sampling, but with a different offset of the sampling grid. We are still able to recreate the original signal from both these instances, but the results of the dilation are different. Neither result is the same as a sampled version of the result of the continuous dilation.

We propose to use a piece-wise polynomial function, which is easy to work with. We limit ourselves to third-order polynomials, for which zero-crossings, maxima and minima can be found analytically. Also, it is possible to construct a good approximation of a band-limited function with third-order polynomials [7].

2.1 Representing a 1D signal as a piece-wise polynomial

To represent a continuous one-dimensional function as a set of third-order polynomial segments, the following information is required:

- Starting point of each polynomial (x_i)
- Polynomial coefficients (a_i, b_i, c_i, d_i)
- Length of each polynomial (l_i)

Since the function we are representing is defined everywhere in the signal domain, the end point of a polynomial is equal to the starting point of the next one. Thus the length is redundant, and we only need to store the starting points of each polynomial and the end point of the last polynomial. The function is then written as a collection of segments $S_i(x)$

$$S_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3 \quad , \tag{1}$$

$i \in [0, 1, 2, ...N - 1]$, plus a right bound x_N.

Certain operations on such a representation are trivial. For example, shifting the whole function just requires incrementing or decrementing the starting points x_i, and inverting the function is accomplished by negating all the polynomial coefficients. Other operations we apply to the polynomial function are sampling (evaluating the function at chosen locations) and integration. The integral over the function is the sum of the integral over each segment, determined by

$$\int_{x_0}^{x_N} f(x) = \sum_{i=0}^{N-1} \tfrac{1}{4} d_i x_{i+1}^4 + \tfrac{1}{3} c_i x_{i+1}^3 + \tfrac{1}{2} b_i x_{i+1}^2 + a_i x_{i+1} \quad . \tag{2}$$

2.2 Converting the sequence of samples into a piece-wise polynomial

To create the piece-wise polynomial representation $f(x)$ from the given samples $f[n]$, we require an interpolation function that has certain characteristics:

- The resulting function must have as many continuous derivatives as possible (since the original band-limited signal is infinitely differentiable). We use third-order polynomials, thus we require that the second-order derivative be continuous.
- It must be a local representation. That is, the zone of influence of a single pixel must be limited, because only a limited number of samples is available.
- It must be capable of producing a polynomial representation.

An interpolator that satisfies these constraints is the cubic spline interpolator [8–10]. It produces polynomial segments in between each two sample points. Although its impulse response decays quite quickly, it requires a filter with an infinite impulse response to determine polynomial coefficients. This filter can be implemented recursively [11]. Note that a spline of infinite order equals the ideal interpolator (the sinc function) [7]. Thus, a cubic spline is an approximation of the ideal interpolator.

In the case of noisy input samples, it might be better to use a least squares spline [8]. In this case, the reconstructed function does not need to be equal to the samples at the sample locations, and thus can be smoother. Furthermore, by computing the piece-wise polynomial in this way the number of pieces is reduced, which makes further processing faster as well.

2.3 Dilations

Examining the 1D dilation operation with a flat, compact SE B, one can readily see that the result is composed of plateaus (constant sections) as well as slopes with the exact same shape as can be found in the input signal (see Fig. 2). Let us define the set B as

$$B = \{x | x \in [-r, r]\} \quad . \tag{3}$$

The plateaus are formed when, at a point x, the maximum value over the neighborhood B comes from a local maximum (see Fig. 2a). At points near x, the

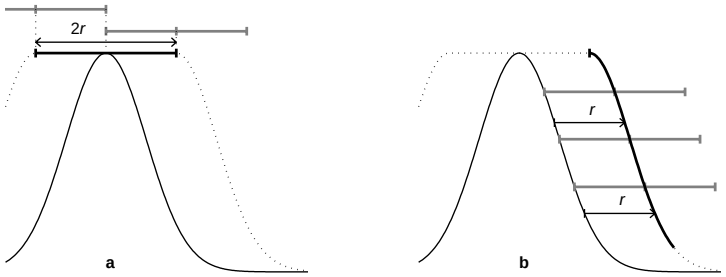

Fig. 2. Construction of the dilated function. **a:** maxima create plateaus in the output. **b:** slopes are shifted by a fixed distance, dictated by the size of the structuring element. In these graphs, the thin black line is the input signal, the dotted black line is the output signal, the thick black line shows how the output signal is constructed, and the thick grey lines give the size of the structuring element ($2r$).

maximum over the neighborhood will also come from the same local maximum, and will therefore receive the same value. These plateaus will have a width of at most $2r$, centered on the local maximum.

The sloped regions are produced when the maximum over B does not come from a local maximum. In this case, it must come from the border of the structuring element (see Fig. 2b). At nearby points, the resulting value also comes from the same edge of the neighborhood. Therefore, a slope is created that is the exact copy of a slope from the input signal, shifted by r or $-r$. This is known as the slope transform [12, 13].

Thus, for a one-dimensional signal, the output of the dilation with a flat, compact SE is the point-wise (or, in our case, the segment-wise) maximum of three functions:

- the input signal translated by r: $f(x - r)$,
- the input signal translated by $-r$: $f(x + r)$, and
- a signal composed of plateaus centered around each of the local maxima.

The above analysis is valid for flat, compact and symmetric structuring elements. Any non-symmetric flat, compact SE C, defined by

$$C = \{x | x \in [-r - d, r - d]\} = \{x | x + d \in [-r, r]\} \quad , \tag{4}$$

can be converted into a symmetric SE B by translating the input or the output signal (δ denotes the dilation operator):

$$\delta_C f(x) = \delta_B f(x - d) = [\delta_B f](x - d) \quad . \tag{5}$$

A non-compact structuring element can be constructed with the union of compact structuring elements:

$$\delta_{[\bigcup_i B_i]} f(x) = \bigvee_i \delta_{B_i} f(x) \quad . \tag{6}$$

Thus, the above analysis suffices for any flat SE.

2.4 Implementation

The Plateau Function Creating the function composed of the plateaus is not very complicated. First, all local maxima must be found. This is easily accomplished by examining the first and second order derivatives of each of the polynomials:

$$S_i'(x) = 0 \quad \wedge \quad S_i''(x) < 0 \quad \Leftrightarrow \quad x \text{ is a local maximum.} \tag{7}$$

Note that finding these derivatives is trivial, and finding the zero crossings of the first derivative is accomplished by solving a quadratic equation. Additionally, in the result of a previous morphological operation there can be maxima in the form of cusps and plateaus. These will be found only on knots (boundary points between polynomial segments), and are identified by comparing the derivatives of both polynomials at those points:

$$S_i'(x_{i+1}) \geq 0 \quad \wedge \quad S_{i+1}'(x_{i+1}) \leq 0 \quad \Leftrightarrow \quad x \text{ is a local maximum.} \tag{8}$$

These maxima must be sorted according to their value, largest first. Then the plateau image is created by adding a 0^{th} order polynomial segment, ranging from $x-r$ to $x+r$, and with value $f(x)$, for each maximum at x (see Fig. 3). Each segment added must not overlap with any of the polynomials already present in the function, so it must be cropped to the available space (actually, we are taking the maximum over these segments implicitly). At the end of this process, eventual 'holes' must be filled with segments of value $-\infty$, so that the generated function is defined everywhere in the signal domain, and can be stored in the same manner as the input signal.

Maximum over the Segment Functions The last step is to find the function that is the maximum of the three functions. This is a two-step process in which first two functions are compared, and then the result is compared to the third. To avoid complicated exceptions in the algorithm, we pad the three functions with zero-order polynomials so that they span the same interval (from $x_0 - r$ to $x_N + r$). The translated versions of the input signal are extended with the edge value (so as to keep them continuous). The function containing the plateaus is extended with $-\infty$.

This comparison is very simple, but potentially generates quite a lot of segments. For each (portion of a) segment $S_i^1(x)$ in one function that spans the same region as another (portion of a) segment $S_i^2(x)$ in the other function, the

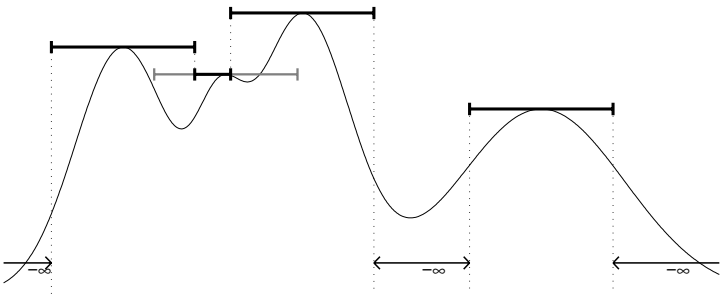

Fig. 3. Construction of the plateau function. At each local maximum a plateau (0^{th} order polynomial) with the size of the structuring element is set. In the case of overlapping plateaus, the one with the highest value is kept intact; the other one must be cropped. Empty regions are filled with segments of value $-\infty$ so that the function is defined everywhere in the signal domain.

intersection points $S_i^1(x) = S_i^2(x)$ must be found (this is a cubic equation, the solution can be found in Bronstein [9]). There are up to three intersection points, and thus up to four sub-segments. For each of these, the polynomial with the larger value is used to construct the output signal.

2.5 Erosions, Closings and Openings

Since the erosion ε is the dual operation of the dilation [14], it is implemented by inverting the signal, applying the dilation, and inverting the result again:

$$\varepsilon_B f(x) = -\delta_B[-f(x)] \quad . \tag{9}$$

As stated above, inverting the piecewise polynomial function is easily accomplished by negating all the polynomial coefficients.

The closing ϕ is created by applying an erosion to the result of the dilation,

$$\phi_B f(x) = \varepsilon_{\check{B}}[\delta_B f(x)] \quad , \tag{10}$$

and the opening γ is constructed the other way around,

$$\gamma_B f(x) = \delta_{\check{B}}[\varepsilon_B f(x)] \quad . \tag{11}$$

The algorithm as described above can be applied on its own result, so that implementing closings and openings becomes trivial.[1]

[1] The source code for the sampling-free morphology is available through the author's web site at `http://www.ph.tn.tudelft.nl/~cris/sfm.html`.

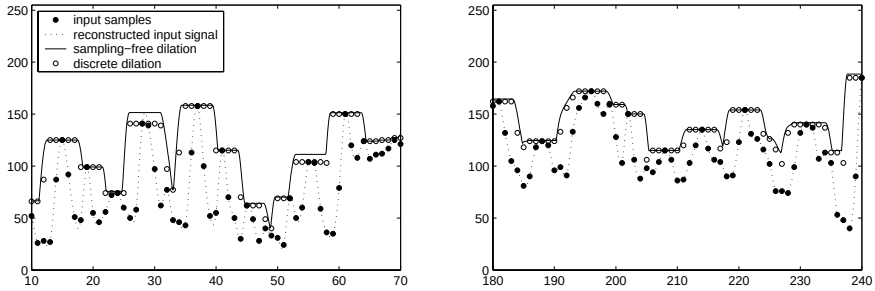

Fig. 4. Two interesting portions of a 1D signal, together with its sampling-free dilation. The open dots give the values of the discrete dilation for comparison. The SE has a length of 5 pixels.

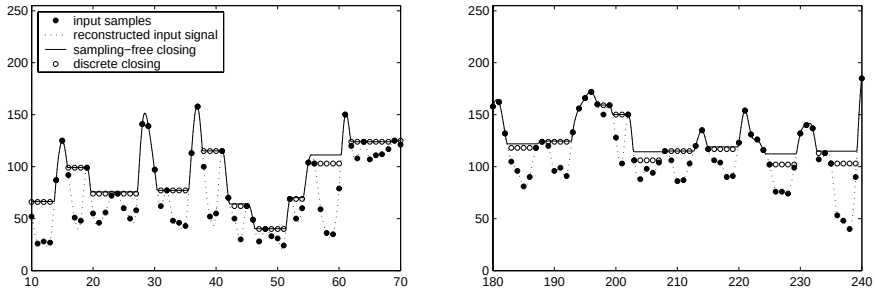

Fig. 5. Two interesting portions of a 1D signal, together with its sampling-free closing. The open dots give the values of the discrete closing for comparison. The SE has a length of 5 pixels.

3 Results

3.1 A first examination of the algorithm

We extracted a line out of an image to apply our methods to. Figure 4 shows two portions of this line, along with the reconstructed continuous function, the result of a discrete dilation (i.e. one applied to the samples directly) and that of the sampling-free dilation proposed here. Figure 5 shows the results of the discrete and sampling-free closings on the same signal.

In these figures we can see that the sampling-free dilation reaches higher values than the discrete variant at some points, especially on plateaus. The value of this signal at these points is equal to the value of the true local maximum of the input signal (or rather of the cubic spline approximation). Likewise, the closing has higher values at the plateaus (the continuous version is equal only in exceptional cases).

3.2 Granulometry

A granulometry is a multi-scale closing (or opening) [14]. The result of the operation at each scale is integrated to obtain a graph comparable to a cumulative size distribution. See Soille [5] for more information on granulometries. We normalize the measured granulometric curve so that the first value is 0 and the last value is 1.

We created a signal of which we know the function that represents the granulometric curve. To the samples of this signal we applied a granulometry with both the sampling-free and discrete closings, and compared the results to the theoretical granulometric curve.

The signal we used is a sine,

$$f(x) = sin\left(\frac{2\pi x}{T}\right) \quad , \tag{12}$$

with T the period. The sampling distance is 1, meaning that T must be larger than 2 for error-free sampling. The theoretical granulometric curve is described by

$$h(r) = \begin{cases} \frac{1}{\pi} sin\left(\frac{r\pi}{T}\right) - \frac{r}{T} cos\left(\frac{r\pi}{T}\right) & \text{for } r < T, \\ 1 & \text{for } r \geq T, \end{cases} \tag{13}$$

with r the size of the structuring element. We used two periods: $T_1 = \frac{200}{9}$ and $T_2 = \pi$. We chose these values because each period of the sine starts at a different offset with respect to the sampling grid. Both signals can be correctly sampled at a rate of 1. The first one can be interpolated very accurately using cubic splines, whereas the second will produce larger errors due to the inability of the spline to correctly reconstruct high-frequency signals (see Fig. 6). The frequency characteristic of the cardinal cubic spline can be found in Fig. 2 of [8]. The spline interpolation on the second signal produces a result that is obviously not an exact reproduction of the input signal. Therefore, the result of the granulometry is inaccurate as well. However, it lies much closer to the theoretical curve than the result of the discrete granulometry. Another obvious drawback of the discrete granulometry is the discreteness of the structuring element, which can only be constructed with integer lengths. Because of this, the granulometry with the sampling-free closing could be sampled much more densely.

We repeated the above experiments after adding noise to the input samples (see Fig. 7). The results show more or less the same characteristics, except that the granulometric curves deviate a bit more from the theoretical (noiseless) values. These results might improve when using regularized splines as mentioned above.

4 Extension to multi-dimensional images

Morphological operations can be defined for images of any dimensionality. Therefore, we would like to extend our algorithm to multi-dimensional images as well. This is, however, not an easy task.

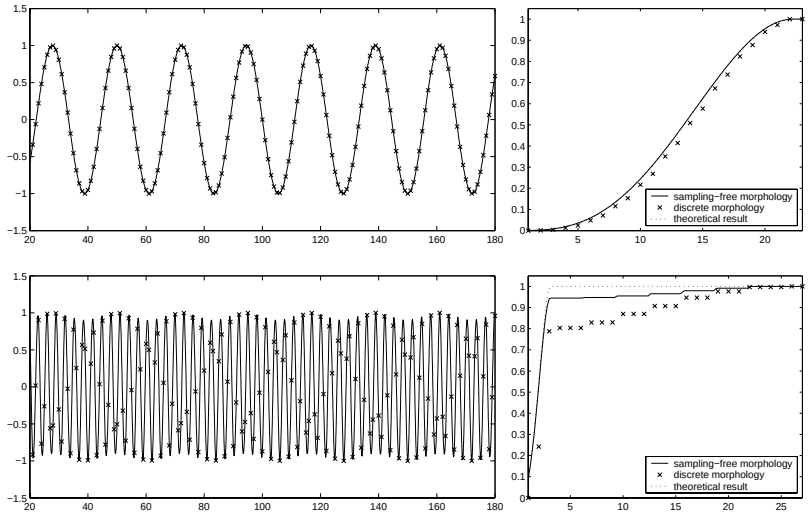

Fig. 6. Granulometry of a sine function sampled at different rates. On the left are the samples and the continuous function created with cubic splines. On the right is the result of the granulometry, computed with both discrete and continuous-domain morphology, compared to the theoretical granulometric function.

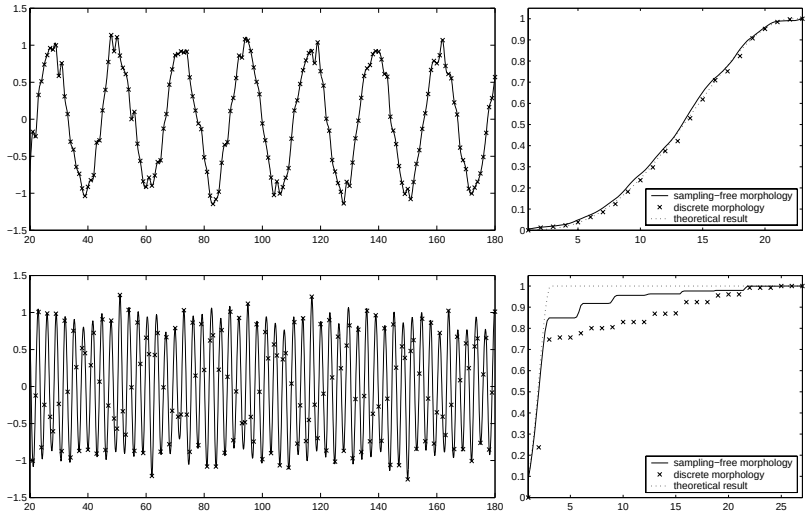

Fig. 7. Granulometry of a sine function sampled at different rates. Noise was added to the samples prior to the analysis. On the left are the samples and the continuous function created with cubic splines. On the right is the result of the granulometry, computed with both discrete and continuous-domain morphology, compared to the theoretical granulometric function of the noiseless signal.

Obviously, extension to multi-dimensional images by processing each dimension separately will not work. In this case, maxima lying in between raster lines will be missed. This shows that it is necessary to create a patch representation of the image (using multi-dimensional cubic splines), and work on that.

However, using multi-dimensional structuring elements with this representation also introduces a problem: we would need to create a translated version of the input image for each point along the boundary of this structuring element. Since there are an infinite number of these points, this is an impossible task. If we would simplify the structuring element by taking only a limited number of points along the contour, we would again miss some of the local maxima.

Thus, we should limit ourselves to one-dimensional structuring elements working on a patch representation of a multi-dimensional image. Now we can implement the same operations we suggested above: create two translated versions of the input image (one for each end of the structuring element), and an image consisting of plateaus centered around each of the local maxima in the image; then take the maximum over these three images. In this case, however, local maxima are all points for which there is a maximum in the direction of the structuring element. These points form lines (in a two dimensional image; in an N-D image this is a $(N-1)$-D plane). The plateaus we create are therefore patches with a zero-order polynomial in one direction, and some third-order polynomial in all the orthogonal directions. These polynomials are taken from the input patches along the local maxima line.

The only problem with this approach is that the patches produced by a dilation have very complex boundaries (given by third-order polynomials). This makes an implementation difficult, although the theoretical formulation is trivial.

Using linear structuring elements it is possible to create more complex multi-dimensional structuring elements such as the rectangle, the hexagon, the octagon, etc. [14, 5]. These shapes are increasingly better approximations of a disk. Thus, it is possible to create an arbitrarily accurate approximation to the isotropic SE.

5 Conclusion

In this paper we have shown that it is possible to apply continuous-domain morphology to one-dimensional images (signals). On a sampled signal, discrete morphology produces results that are not the same as the results produced by continuous morphology on the signal before sampling. By representing the continuous signal as a piece-wise polynomial function, we were able to compute dilations and erosions with flat structuring elements that produce the exact same results as their continuous-domain counterparts would produce. Some error is introduced when converting the sampled signal into a piece-wise polynomial function, but the morphological operations themselves do not introduce any further errors. It is possible to cascade dilations and erosions to obtain openings and closings and other, more complex morphological filters.

We applied a granulometry (multi-scale closings) to some sampled test signals using the sampling-free closings constructed with our algorithms. The resulting granulometric curve is compared to the theoretical result. The differences found were due to the difficulty in converting the samples of a high-frequency signal into a piece-wise polynomial function. This process is further complicated by the addition of noise to the input samples. These results should improve when using regularized splines instead of interpolating splines.

To apply these methods to higher-dimensional images, the decomposition principle must be used. Some structuring elements can be decomposed into one-dimensional operations. These can be applied to the polynomial patch representation of the multi-dimensional image. Even though it is trivial to state theoretically, the cubic boundary between these polynomial patches is complicated to use in an actual implementation.

References

1. Nyquist, H.: Certain topics in telegraph transmission theory. Transactions of the AIEE (1928) 617–644 [reprinted in: Proceedings of the IEEE **90** (2002) 280–305].
2. Shannon, C.: Communication in the presence of noise. Proceedings of the IRE **37** (1949) 10–21 [reprinted in: Proceedings of the IEEE **86** (1998) 447–457].
3. van Vliet, L.J.: Grey-Scale Measurements in Multi-Dimensional Digitized Images. PhD thesis, Pattern Recognition Group, Faculty of Applied Physics, Delft University of Technology, Delft (1993)
4. Serra, J.: Image Analysis and Mathematical Morphology. Academic Press, London (1982)
5. Soille, P.: Morphological Image Analysis. Springer-Verlag, Berlin (1999)
6. Luengo Hendriks, C.L., van Vliet, L.J., van Kempen, G.M.P., Bouwens, E.C.M.: (Using morphological sieves to detect minute differences in pore sizes) Submitted to Journal of Microscopy.
7. Unser, M., Aldroubi, A., Eden, M.: Polynomial spline signal approximations: Filter design and asymptotic equivalence with shannon's sampling theorem. IEEE Transactions on Information Theory **38** (1992) 95–103
8. Unser, M., Aldroubi, A., Eden, M.: B-spline signal processing: Part I - theory. IEEE Transactions on Signal Processing **41** (1993) 821–833
9. Bronstein, I.N., Semendjaev, K.A., Musiol, G., Mühlig, H.: Taschenbuch der Mathematik. 4th edn. Verlag Harri Deutsch, Thun, Frankfurt am Main (1999)
10. de Boor, C.: A Practical Guide to Splines. Springer, New York (2001)
11. Unser, M., Aldroubi, A., Eden, M.: Fast B-spline transforms for continuous image representation and interpolation. IEEE Transactions on Pattern Analysis and Machine Intelligence **13** (1991) 277–285
12. van den Boomgaard, R., Smeulders, A.: The morphological structure of images: The differential equations of morphological scale-space. IEEE Transactions on Pattern Analysis and Machine Intelligence **16** (1994) 1101–1113
13. Dorst, L., van den Boomgaard, R.: Morphological signal processing and the slope transform. Signal Processing **38** (1994) 79–98
14. Matheron, G.: Random Sets and Integral Geometry. Wiley, New York (1975)