Discrete calculation of the off-axis angular spectrum based light propagation

To cite this article: P Lobaz 2013 *J. Phys.: Conf. Ser.* **415** 012040

View the article online for updates and enhancements.

Discrete calculation of the off-axis angular spectrum based light propagation

P Lobaz

Department of Computer Science and Engineering, University of West Bohemia, Univerzitni 8, 306 14 Plzen, Czech Republic

E-mail: `lobaz@kiv.zcu.cz`

Abstract. Light propagation in a free space is a common computational task in many computer generated holography algorithms. A solution based on the angular spectrum decomposition is used frequently. However, its correct off-axis numerical implementation is not straightforward. It is shown that for long distance propagation it is necessary to use digital low-pass filtering for transfer function calculation in order to restrict source area illumination to a finite area. It is also shown that for short distance propagation it is necessary to introduce frequency bands folding in transfer function calculation in order to simulate finite source area propagation. In both cases it is necessary to define properly interpolation filters that reconstruct continuous nature of the source area out of its sampled representation. It is also necessary to zero-pad properly source area sampling in order to avoid artifacts that stem from the periodic nature of the fast Fourier transform.

1. Introduction

To calculate coherent light propagation in a free space, scalar approximation is used frequently. A common task is such a calculation where complex amplitudes of light are given in the area *source* in a plane $z = 0$ and we look for complex amplitudes in the area *target* in a plane $z = z_0$, $z_0 > 0$. A common procedure leads to the Rayleigh-Sommerfeld integral of the first kind [1], or to its mathematically equivalent form, the angular spectrum decomposition [2]. Approximations of these formulas are used frequently, namely Fresnel and Fraunhofer approximations. However, these approximations are used in paraxial regime while in computer generated holography an off-axis solution is often needed (e. g. [3]); therefore we will not discuss them.

The problem has to be solved numerically in computer generated holography. This leads to discretization of signals. Discretization of the Rayleigh-Sommerfeld solution is not straightforward [4, 5] and discretization of the angular spectrum decomposition is tricky [6, 7]. Onural [6] describes the discretization process in general and shows that it leads to formation of signal copies in spatial domain that can be filtered out. However, he does not discuss implementation issues. Matsushima [7] deals with the implementation and solves a troublesome aliasing problem by local frequency estimation; however, he does not analyse the effect of hard frequency clipping.

This article focuses on reference calculation of light propagation between parallel planes using angular spectrum decomposition. We will point out what makes the discretization difficult

and how to overcome the problems so that the method provides the same results as the Rayleigh-Sommerfeld method. We will deal with both large propagation distances discussed by Matsushima [7] and small propagation distances that were not discussed in literature yet. We will explain the meaning of "large" and "small" distance later.

Structure of the article is as follows. At first we will precisely show how to discretize the propagation calculation based on convolution (i. e. based on Rayleigh-Sommerfeld integral). We will show that it is necessary to introduce spatial limitation of a convolution kernel for successful discretization. This leads to frequency limitation of the transfer function used in the angular spectrum decomposition method. We will show that the frequency limited transfer function can be calculated using digital signal processing methods. At last we will deal with the case where this frequency limitation is actually useless due to small propagation distance. We will show that, in this case, we have to deal with exact nature of sampling and reconstruction of signals involved. We will show that the choice of reconstruction leads to introduction of artificial alias.

2. How to read the article

The mathematical explanation presented may be unpleasant to follow. Readers are therefore encouraged to go through the presentation packed as a "multimedia" attachment to this article. It shows pictures containing various problems that appear when calculating the propagation numerically. I have decided to attach these pictures as a separate media for two reasons. The first one is: the pictures show mainly problems with aliasing. It is therefore needed to control the display of these images precisely which is not possible in a PDF reader or in printed media. The second one is: separate media gives the opportunity to show much more images than any printed media allows. I should note that the presentation does not contain any information not covered by the article.

3. Convolution discretization

Let us assume that we know complex amplitudes $u(x, y, 0)$ of monochromatic coherent light of wavelength λ in the area *source* ($x_{s\min} \leq x < x_{s\max}$, $y_{s\min} \leq y < y_{s\max}$, $z = 0$) and we want to calculate complex amplitudes $u(x, y, z_0)$ in the area *target* ($x_{t\min} \leq x < x_{t\max}$, $y_{t\min} \leq y < y_{t\max}$, $z = z_0 > 0$). The solution is given by the Rayleigh-Sommerfeld integral of the first kind:

$$u(x, y, z_0) = \frac{-1}{2\pi} \iint\limits_{-\infty}^{\infty} u(\xi, \eta, 0) \frac{\partial}{\partial z} \frac{\exp(\mathrm{j}kr)}{r} \, \mathrm{d}\xi \, \mathrm{d}\eta \tag{1}$$

where $r = \sqrt{(x - \xi)^2 + (y - \eta)^2 + z_0^2}$, $k = 2\pi/\lambda$, $\mathrm{j}^2 = -1$ and $u(\xi, \eta, 0) = 0$ for $[\xi, \eta, 0] \notin source$. The second term of the multiplication inside the integral depends on $(x - \xi)$ and $(y - \eta)$ only, which means that the integral can be rewritten as the convolution with the Rayleigh-Sommerfeld kernel $h(x, y, z)$:

$$u(x, y, z_0) = u(x, y, 0) \otimes h(x, y, z_0) = \iint\limits_{-\infty}^{\infty} u(\xi, \eta, 0) h(x - \xi, y - \eta, z_0) \, \mathrm{d}\xi \, \mathrm{d}\eta \tag{2}$$

$$h(x, y, z) = \frac{-1}{2\pi} \frac{\partial}{\partial z} \frac{\exp(\mathrm{j}kr)}{r} = \frac{-z}{2\pi} \left(\mathrm{j}k - \frac{1}{r} \right) \frac{\exp(\mathrm{j}kr)}{r^2}$$

$$r = \sqrt{x^2 + y^2 + z^2}$$

The kernel $h(x, y, z)$ can be interpreted easily: the value $h(x_d, y_d, z_d)$ describes the change of the complex amplitude of light travelling from the point $[x_s, y_s, z_s]$ to the point $[x_s + x_d, y_s + y_d, z_s + z_d]$.

To calculate $u(x, y, z_0)$ in the area *target* using (1) we have to know the kernel $h(x, y, z_0)$ for $x_{t\min} - x_{s\max} \leq x < x_{t\max} - x_{s\min}$, similarly for y. The value $h(x, y, z_0)$ is not important for other x, y because in this case $u(x, y, 0) = 0$. We will use this fact in a while.

We discretize (1) easily by changing integrals to sums and differentials to differences. The sums will have finite extent of the indices thanks to (in fact) finite domain of the integration. Therefore their calculation will be easy, although the computational complexity will be high.

To reduce computational complexity, let us rewrite the equation (2):

$$u(x, y, z_0) = u(x, y, 0) \otimes h(x, y, z_0) = \mathcal{F}^{-1} \left\{ \mathcal{F}\{u(x, y, 0)\} \cdot \mathcal{F}\{h(x, y, z_0)\} \right\} \qquad (3)$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ are 2-D Fourier and inverse Fourier transform respectively. We will try to use the discrete Fourier transform (DFT) implemented as the fast Fourier transform (FFT) after discretization.

The discrete Fourier transform can be defined if the original continuous functions are periodic before discretization (sampling); then we take into account one period of functions $u_p(x, y, z_0)$, $u_p(x, y, 0)$ and $h_p(x, y, z_0)$ derived from functions $u(x, y, z_0)$, $u(x, y, 0)$ and $h(x, y, z_0)$. We can define the DFT another way if the functions to be transformed are spatially limited; in this case we can assume just one period of the functions $u_p(x, y, z_0)$, $u_p(x, y, 0)$ a $h_p(x, y, z_0)$. Both ways lead to the same results. This means that the results of the DFT can be interpreted in both ways. We will choose the way that will be more suitable in a particular situation.

Let us briefly give a hint what is the meaning of the functions $u_p(x, y, 0)$, $u_p(x, y, z_0)$ and $h_p(x, y, z_0)$ before we define them precisely. Their period in x direction is equal to the sum of widths of the *source* and the *target* (similarly in y). One period of the function $u_p(x, y, 0)$ is composed of values of $u(x, y, 0)$ in the way that one corner of the *source* is translated to the origin. The meaning of the function $u_p(x, y, z_0)$ is similar. The $h_p(x, y, z_0)$ is restricted and shifted version of the function $h(x, y, z_0)$. The meaning of the value $h_p(0, 0, z_0)$ is the change of the complex amplitude of light travelling from the point $[x_{s\min}, y_{s\min}, 0]$ to the point $[x_{t\min}, y_{t\min}, z_0]$, see fig. 1.
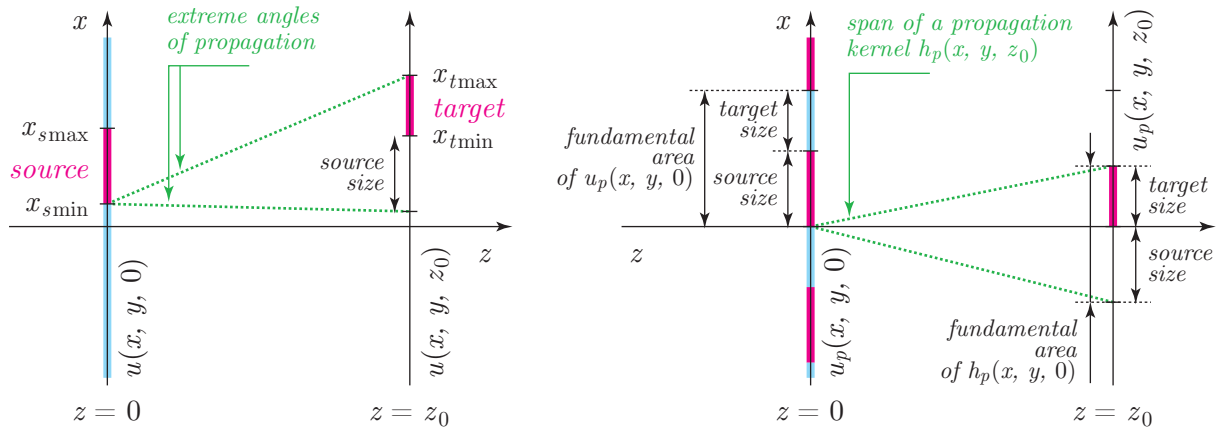

Figure 1. Geometry of the problem in xz slice. *Left image:* original setup. Red line in the source area denotes the complex amplitudes to be propagated, blue line in the same plane defines zero value. It should be clear that we need to know the values of a propagation kernel contained in a green wedge only. *Right image:* setup prepared for discretization. *Source* and *target* are shifted to the z axis, blue and red lines in the *source* plane show both zero padding and periodicity. The green wedge shows one period (fundamental area) of the function $h_p(x, y, z_0)$.

Let us define the functions precisely. The function $h_p(x, y, z_0)$ is defined in "the fundamental area" $x_{s\min} - x_{s\max} \leq x < x_{t\max} - x_{t\min}$ (similarly for y) as:

$$h_p(x, y, z_0) = h(x + x_{t\min} - x_{s\min}, y + y_{t\min} - y_{s\min}, z_0) \tag{4}$$

Definition outside the fundamental area depends on situation: it can be either zero or the function can be made periodic. However, as we will see, the values outside the fundamental area are not important, which means that the periodic nature of the function is not harmful.

Let us define the function $u_p(x, y, 0)$ in the fundamental area $0 \leq x < (x_{t\max} - x_{t\min}) + (x_{s\max} - x_{s\min})$ (similarly for y; notice that the size of the area is the same as for $h_p(x, y, z_0)$) as:

$$u_p(x, y, 0) = \begin{cases} u(x + x_{s\min}, y + y_{s\min}, 0) & \text{for} \quad 0 \leq x < x_{s\max} - x_{s\min}, \\ & \qquad 0 \leq y < y_{s\max} - y_{s\min} \\ 0 & \text{elsewhere in the fundamental area} \end{cases}$$

and again let us make it periodic.

Let us calculate $u_p(x, y, z_0) = u_p(x, y, 0) \otimes h_p(x, y, z_0)$. If we consider the function $u_p(x, y, 0)$ to be periodic and the function $h_p(x, y, z_0)$ to be zero outside the fundamental area, we can easily prove that $u_p(x, y, z_0)$ will contain correct result of the propagation from the *source* to the *target* in the area $0 \leq x < x_{t\max} - x_{t\min}$ (similarly for y); the values of $h_p(x, y, z_0)$ are meaningless for other x, y as they are damaged by periodicity of the function $u_p(x, y, 0)$. The proof easily follows from the geometry of the problem.

If we consider the function $u_p(x, y, 0)$ to be spatially limited and the function $h_p(x, y, z_0)$ to be periodic, we get the same result. If we consider both functions to be periodic, the meaning of the result remains the same; however, we are in fact in the world of discrete Fourier transform.

The form of the discrete calculation follows from the aforementioned ideas. The area *source* is discretized by $M_x \times M_y$ samples, the area *target* by $N_x \times N_y$ samples. For simplicity let us assume such parameters so that the sampling period Δ is the same in both directions x and y and in both *source* and *target* areas. Its value is then e. g. $\Delta = (x_{s\max} - x_{s\min})/M_x$.

The area *source* is discretized by samples $u_0[m, n] = u_p(m\Delta, n\Delta, 0)$, the convolution kernel is discretized by samples $h[m, n] = h_p(m\Delta, n\Delta, z_0)$ for $m \in \{0, 1, \ldots, M_x + N_x - 2\}$ (similarly n, see [4]). We consider functions $u_p(x, y, 0)$ and $h_p(x, y, z_0)$ to be periodic. Then we can calculate

$$u_{z_0}[\,] = \mathbf{IDFT} \left\{ \mathbf{DFT}\{u_0[\,]\} \cdot \mathbf{DFT}\{h[\,]\} \right\} \tag{5}$$

to get the array $u_{z_0}[\,]$ that contains complex amplitudes of the area *target* in the first $N_x \times N_y$ elements. In the equation (5), **DFT** and **IDFT** stands for forward and backward 2-D discrete Fourier transform, respectively, and $\cdot$ stands for elementwise product (Hadamard product).

We can summarize the results as follows. For the discrete calculation of light propagation, we have to use a convolution kernel spatially limited to a certain area. The *source* has to be zero-padded to span the same area. The result of their convolution contains correct values in the area of the size proportional to the size of the *target*.

4. Angular spectrum discretization for large propagation distances

The equation (1) can be written in the equivalent form called the angular spectrum decomposition [2]:

$$u(x, y, z_0) = \mathcal{F}^{-1}\{U(f_x, f_y, 0) \cdot H(f_x, f_y, z_0)\}$$

where

$$U(f_x, f_y, 0) = \mathcal{F}\{u(x, y, 0)\}$$

$$H(f_x, f_y, z_0) = \exp\left(-\mathrm{j}\, 2\pi z_0 \sqrt{\lambda^{-2} - f_x^2 - f_y^2}\right)$$

and f_x, f_y are Fourier domain variables.

It follows from (3) that

$$u(x, y, z_0) = u(x, y, 0) \otimes h(x, y, z_0) = \mathcal{F}^{-1}\left\{U(f_x, f_y, 0) \cdot \mathcal{F}\{h(x, y, z_0)\}\right\},$$

that is $H(f_x, f_y, z_0) = \mathcal{F}\{h(x, y, z_0)\}$. It seems that the propagation calculation should be advantageous using the angular spectrum decomposition compared to convolution with the Rayleigh-Sommerfeld kernel because we have one Fourier transform less – the transform of the kernel is known in the analytic form.

In the following paragraphs, we will talk about various limitations of functions in both spatial and Fourier (frequency) domain. A function $f(x, y)$ defined in spatial domain is spatially limited if it is zero outside a bounded area in the plane (x, y); it is frequency limited if $\mathcal{F}\{f(x, y)\}$ is zero outside a bounded area in the plane (f_x, f_y). Similarly, a function $F(f_x, f_y)$ defined in frequency domain is spatially limited if it is zero outside a bounded area in the plane (f_x, f_y); it is frequency limited if $\mathcal{F}^{-1}\{F(f_x, f_y)\}$ is zero outside a bounded area in the plane (x, y).

We know from the previous section that we have to introduce certain functions to discretize the calculation. We have to define the periodic function $u_p(x, y, 0)$ based on the function $u(x, y, 0)$ and to calculate its (discrete) Fourier transform; we will follow this procedure exactly. We also need to spatially restrict the function $h(x, y, z_0)$, and to make its periodic form alternatively. However, we do not know the Fourier transform of the function $h_p(x, y, z_0)$ in the analytic form.

Fortunately we can use digital signal processing tools. If the signal is spatially limited in one domain (in our case spatial domain), it is frequency limited in the other domain (in our case frequency domain). We can calculate frequency limited signal using low-pass filter $l(f_x, f_y)$. Let us assume the function $h_p(x, y, z_0)$ to be spatially limited (i. e. not periodic). For propagation calculation, we have to use the transfer function $H_p(f_x, f_y, z_0) = H(f_x, f_y, z_0) \otimes l(f_x, f_y)$.

It follows from properties of the Fourier transform that the function $H_p(f_x, f_y, z_0)$ has to be spatially unlimited. This does not matter even in numerical calculation. Since the function $u_p(x, y, 0)$ is periodic, then $U_p(f_x, f_y, 0)$ is spatially limited; and we need to calculate the product $U_p(f_x, f_y, 0)H_p(f_x, f_y, 0)$. It also follows that the value of the function $H_p(f_x, f_y, z_0)$ can be arbitrary outside the important area, and therefore we can use its periodic form to introduce discrete calculation correctly.

We cannot limit the frequency content of the function $H(f_x, f_y, z_0)$ sharply using digital low-pass filtering; the frequency limitation is approximate. In the spatial domain it means that the transition between zero and non-zero part is gradual instead of sharp. This does not matter either. All we need to do is to enlarge the fundamental area of the functions $u_p()$ and $h_p()$, i. e. zero-padding of the array $u_p[]$ will be larger than defined in section 3, so that the gradual transition will not affect the *target* area.

We will face a problem in a practical implementation. The function $H(f_x, f_y, z_0)$ is frequency unlimited – the local frequency [1, 7] in the point $[f_x, f_y]$ grows without bound as this point approaches a circle of a radius λ^{-1} centered in the origin. Moreover, if the propagation distance z_0 is large, then the local frequency will be large as well everywhere except in the origin. Therefore it is impossible to sample the function $H(f_x, f_y, z_0)$ correctly and then the low-pass filtering will not work properly.

However, we can use the same procedure as described by Matsushima [7] who realizes the aliasing problem. He evaluates local frequency when sampling the function $H(f_x, f_y, z_0)$, and if the local frequency is bigger than a half of the sampling frequency, he sets the function $H(f_x, f_y, z_0)$ to be zero. He solves the aliasing problem this way, on the other hand he introduces a spatial limitation of the function $H(f_x, f_y, z_0)$. It follows that the propagation kernel $\mathcal{F}^{-1}\{H_{\text{Matsushima}}(f_x, f_y, z_0)\}$ is then spatially unlimited which is not correct. It should be however emphasised that even though it gives remarkably good results.

The solution is easy. We can sample the function $H(f_x, f_y, z_0)$ using higher sampling frequency than desired and use Matsushima's procedure to avoid alias. Then we can filter it using $l(f_x, f_y)$ and downsample the result to get a correct sampling frequency. As the cutoff frequency of the filter $l(f_x, f_y)$ is derived from the sizes of the arrays used in the calculation, it is guaranteed that no aliasing appears when downsampling.

It follows from practical experiments that it is sufficient to sample the function $H(f_x, f_y, z_0)$ using sampling frequency twice as high as desired, and to use sinc low-pass filter with Hamming window as $l(f_x, f_y)$. The size of the filter should be chosen carefully – a long filter filters high frequencies well, but takes long to evaluate. Evaluation with a short filter is faster but slow attenuation of high frequencies has to be compensated with bigger zero-padding of the array $u_0[\,]$.

5. Discretization for short propagation distances

To calculate the propagation numerically (either using convolution approach or angular spectrum approach), we have to calculate (discrete) Fourier transform of the function $u_p(x, y, 0)$. Its result is the function $U_p(f_x, f_y, 0)$ limited to the area $A = (-1/(2\Delta), 1/(2\Delta)) \times (-1/(2\Delta), 1/(2\Delta))$, or its periodic form. We also need to calculate the function $H_p(f_x, f_y, z_0)$ in the same area, and then to calculate $\mathcal{F}^{-1}\{U_p(f_x, f_y, 0)H_p(f_x, f_y, z_0)\}$.

We can naturally ask a question: what happens if the propagation distance z_0 is so small that the low-pass filtering of the function $H(f_x, f_y, z_0)$ will not have any significant effect inside the area A? It is not easy to find the answer. Let us start with one more look into the convolution based approach described by the equation (2).

The convolution kernel $h(x, y, z_0)$ is a spatially unlimited function that is in fact frequency limited if we ignore evanescent waves. We can easily show that its local frequency grows as the point $[x, y]$ moves away from the origin. The smaler is z_0 the faster is the growth. It can therefore easily happen that the function $h_p(x, y, z_0)$ cannot be properly sampled using sampling period Δ for small propagation distances.

Physical meaning of wrong sampling is easy. Discretization is based on change of integrals to sums in the equation (1). It means that we change a continuous light field in the *source* to a number of point light sources. This replacement cannot be observed from a big distance or in on-axis case but makes a big difference close enough or in off-axis case.

We have to assume (especially in small propagation distances) that one sample of the function $u(x, y, 0)$ represents behaviour of the light field in a small neighbouring area. Let us denote the result of the sampling of the function $u(x, y, 0)$ by the function $u_s(x, y, 0) = u(x, y, 0) \operatorname{comb}(x/\Delta) \operatorname{comb}(y/\Delta)$, where $\operatorname{comb}(x)$ is the sampling function with period of samples 1 (see [1]). Then we can describe the reconstruction of the continuos form using convolution with a reconstruction kernel $r(x, y)$:

$$u(x, y, 0) \approx u_r(x, y, 0) = u_s(x, y, 0) \otimes r(x, y)$$

where $u_r(x, y, 0)$ is a continuos function reconstructed from the discrete samples. The function $u_r(x, y, 0)$ is more or less similar to the function $u(x, y, 0)$ depending on the shape of a reconstruction kernel $r(x, y)$ and size of the sampling period Δ.

We can therefore express the propagation as

$$u(x, y, z_0) \approx u_s(x, y, 0) \otimes \Big(r(x, y) \otimes h(x, y, z_0) \Big)$$

It is possible to put big parenthesses in the equation thanks to associativity of the convolution. It follows that we should not use the function $h(x, y, z_0)$ for discrete propagation calculation, but we shoud use its filtered version $r(x, y) \otimes h(x, y, z_0)$ instead.

It is most common to demand for one sample of the function $u(x, y, 0)$ to influence its close neighbourhood only. For example, the kernel $r(x, y) = \mathrm{rect}(x/\Delta)\,\mathrm{rect}(y/\Delta)$, where $\mathrm{rect}(x)$ is a rectangular pulse of unity width and height centered in the origin, implies the function $u(x, y, 0)$ to be approximated with a piecewise constant function. We can construct other kernels as well that provide piecewise bilinear approximation, piecewise bicubic approximation and so on. In either case, the kernel $r(x, y)$ acts as a low-pass filter. If we use the filtered function $h(x, y, z_0) \otimes r(x, y)$ for construction of the function $h_p(x, y, z_0)$ in the equation (4), we get the result of the propagation calculation as precise as the function $u_r(x, y, 0)$ resembles the function $u(x, y, 0)$. Practical implementation of the procedure is described in [5].

We can repeat the analysis for the angular spectrum decomposition as well. We should not use the function $H(f_x, f_y, z_0)$ for construction of the function $H_p(f_x, f_y, z_0)$; we should use the Fourier transform of the filtered propagation kernel, the function $\mathcal{F}\{h(x, y, z_0) \otimes r(x, y)\} = H(f_x, f_y, z_0)\,\mathcal{F}\{r(x, y)\}$. If we choose convenient kernel $r(x, y)$, we will know analytic form of its Fourier transform and calculation of the product will be simple. This step limits high frequencies that were caused by the discretization process.

It remains to solve the last, but important detail. In the beginning of the section we have stated that the calculations in frequency domain are done inside the area A. It is however possible that the support of the function $H_p(f_x, f_y, z_0)$ will not fit into the area A even if it was filtered with both filters $l(f_x, f_y)$ and $r(x, y)$. If we reject the values of the function $H_p(f_x, f_y, z_0)$ outside of A despite that fact, it means that the final function $u(x, y, z_0)$ was filtered by a third, still unjustified low-pass filter.

We can explain this final low-pass filter. It would be appropriate if the kernel $r(x, y)$ represents a perfect sinc low-pass filter limiting the frequency content of the function $u(x, y, z_0)$ to a range described by the area A. Then the kernel $r(x, y)$ has to be be spatially unlimited. If we do not care, the analysis is finished.

If we would rather keep the kernel $r(x, y)$ spatially limited (which is physically more natural), we have two choices to choose from. The first one is simple – we can use such a sampling period Δ/s, $s \in \mathbf{N}$ for the calculation so that the area A covers the support of $H_p(f_x, f_y, z_0)$ now. This leads to increase of time and memory demands of the calculation, of course. Moreover, if we demand sampling of the *target* to be Δ, we have to downsample the result; we have to admit that a lot of values were calculated needlessly.

The second one is a bit strange at first sight. Signal downsampling in spatial domain can be described easily in frequency domain – the frequencies f and $f + n/\Delta$ merge due to downsampling, where f is frequency f_x or f_y from the range $(-1/(2\Delta), 1/(2\Delta))$ and $n \in \mathbf{Z}$. This effect is called aliasing. We want to avoid it usually; however, if we are decided to sample the *target* with an insufficient sampling period, we have to accept aliasing. It is worth to note that aliasing needs not be harmful. If we are interested in intensities $|u(x, y, z_0)|^2$ only, aliasing in the real or imaginary part of the function $u(x, y, z_0)$ may be harmless. For example, if $u(x, y, z_0) = \cos x + \mathrm{j}\sin x$, then the intensity is always 1 regardless sampling period used, while intensity of "correctly sampled" (i. e. low-pass filtered) version can be either 1 or 0 which is not correct.

The procedure in the second case is obvious: we will perform the downsampling process in frequency domain. Let us calculate the function $H_p(f_x, f_y, z_0)$ in the area $(-s/(2\Delta), s/(2\Delta)) \times (-s/(2\Delta), s/(2\Delta))$, assume it is zero outside this area, and calculate

$$H_a(f_x, f_y, z_0) = \sum_{n_x, n_y} H_p\left(f_x + \frac{n_x}{\Delta}, f_y + \frac{n_y}{\Delta}, z_0\right)$$

for $[f_x, f_y] \in A$ and all suitable n_x, n_y. To calculate the propagation, we have to use the function $H_a(f_x, f_y, z_0)$.

6. Conclusion

The analysis shows how to numerically calculate the propagation of light using the angular spectrum decomposition method. Unlike the procedures described in the literature, it defines low-pass filters $l(f_x, f_y)$ and $r(x, y)$ that are needed to introduce for correct discretization. Then we can choose either more precise, slower calculation or faster, less precise by choosing their parameters. The analysis also shows that it is worth introducing aliasing of the transfer functions in certain situations. The results are shown in figures 2 and 3. The images may display wrong due to resampling; the reader is encouraged to look at the images in the multimedia attachment.

References

[1] Goodman J W 2004 *Introduction to Fourier Optics* 3rd ed (Roberts & Company Publishers) ISBN 0974707724
[2] Lalor E 1968 *J. Opt. Soc. Am.* **58** 1235–1237
[3] Hanák I, Janda M and Skala V 2010 *Visual Computer* **26**(2) 83–96
[4] Lobaz P 2011 *Opt. Express* **19** 32–39
[5] Lobaz P 2012 Calculation of a coherent light propagation in a free space using filtered convolution Tech. rep. University of West Bohemia URL `http://www.kiv.zcu.cz/en/research/publications/`
[6] Onural L 2007 *J. Opt. Soc. Am.* A **24** 359–367
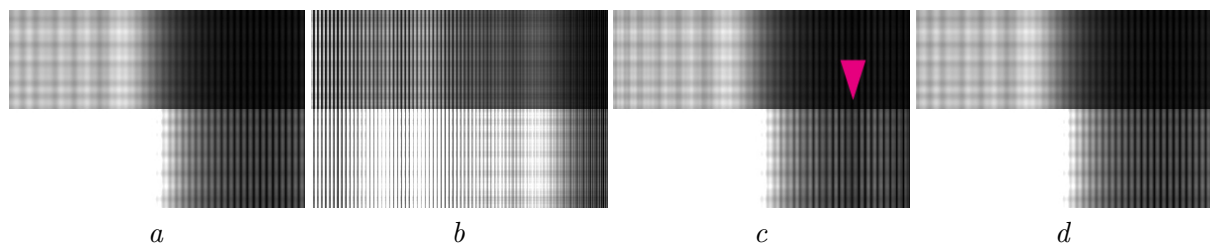[7] Matsushima K 2010 *Opt. Express* **18** 18453–18463

Figure 2. Light ($\lambda = 650$ nm) diffracted by a grating (3×3 mm^2) composed of vertical slits (slits distance 40 μm). Off-axis propagation to a distance of 300 mm, size of each image 3×3 mm^2. Lower half of each image is overexposed to show the details. *a)* Reference Rayleigh-Sommerfeld calculation. *b)* Angular spectrum based calculation without any modification. Notice that aliasing errors destroy the image completely. *c)* Angular spectrum based calculation with Matsushima's kernel filtering. Notice different brightness of fine stripes compared to the reference image. *d)* Angular spectrum based calculation with $l(f_x, f_y)$ sinc kernel of length 50.
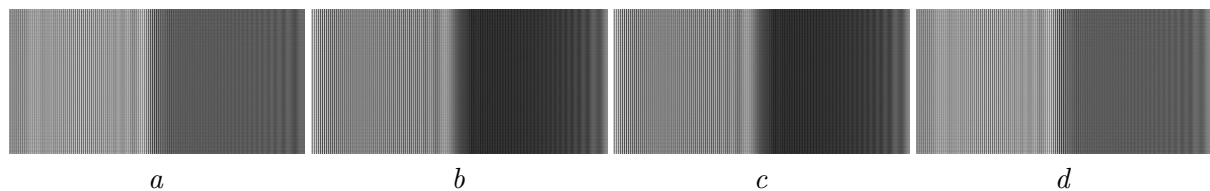

Figure 3. Light ($\lambda = 650$ nm) diffracted by a grating (3×3 mm^2) composed of vertical slits (slits distance 20 μm). Off-axis propagation to a distance of 50 mm, size of each image 3×1.5 mm^2. *a)* Reference Rayleigh-Sommerfeld calculation. *b)* Angular spectrum based calculation without any modification. Notice that the right half is much darker than in the reference image. *c)* Angular spectrum based calculation with Matsushima's kernel filtering. Notice that the image is the same as without any modification because aliasing did not occur. *d)* Angular spectrum based calculation with introduced aliasing. Notice that the image is almost the same as the reference one.