Advanced Filter Design

Hans Knutsson *Mats Andersson* *Johan Wiklund*

Computer Vision Laboratory
Department of Electrical Engineering
Linköping University, SE-581 83 Linköping, Sweden
Phone: +46 13 281887, Fax: +46 13 138526, email: knutte@isy.liu.se

Abstract

This paper presents a general approach for obtaining optimal filters as well as filter sequences. A filter is termed optimal when it minimizes a chosen distance measure with respect to an ideal filter. The method allows specification of the metric via simultaneous weighting functions in multiple domains, e.g. the spatio-temporal space *and* the Fourier space. Metric classes suitable for optimization of localized filters for multidimensional signal processing are suggested and discussed.

It is shown how convolution kernels for efficient spatio-temporal filtering can be implemented in practical situations. The method is based on applying a set of jointly optimized filter kernels in sequence. The optimization of sequential filters is performed using a novel recursive optimization technique. A number of optimization examples are given that demonstrate the role of key parameters such as: number of kernel coefficients, number of filters in sequence, spatio-temporal and Fourier space metrics.

The sequential filtering method enables filtering using only a small fraction of the number of filter coefficients required using conventional filtering. In multidimensional filtering applications the method potentially outperforms both standard convolution and FFT based approaches by two-digit numbers.

Keywords: Filter design, multidimensional filtering, efficient filtering, filter optimization, localized kernels.

1 Introduction

This paper initially presents a formulation of the basic optimization problem and continues to discuss consequences of important constraints and choices of appropriate metrics for the optimization. As an example it is in most applications important that the filtered signal maintains a high spatio-temporal resolution. It is shown how, to this end, a spatio-temporal weighting function can be used to introduce a distance metric that favors spatio-temporally localized kernels. This 'designer metric' approach is general in that it is implementation independent, i.e. the chosen metric is meant to 'tell it all' and consequently the design is equally valid if the filters are implemented as convolutions, using an FFT based approach or any other technique. To illustrate the basic features of the optimizer and the effect of different metrics some simple one-dimensional single filter examples are given.

In multidimensional signal processing, the main part of the computational power often has to be spent on linear filtering operations. The filters do for natural reasons need to be of the same 'outer dimensionality' as the signal and for 3D and 4D data, such as image sequences and time sequences of volumes, the computational load increases dramatically and in practise limits the usefulness of the algorithm. In this paper it is shown how efficient spatio-temporal filtering can be implemented in practical situations.

A theoretical analysis shows that the effect of filters in sequence can be incorporated in the single filter optimizer. This is the basis for a novel method for optimization of general multidimensional filters where a set of sequentially applied filters can be jointly optimized by using the single filter optimizer recursively. Finally results of optimized sequential filters, using only a fraction of the number of filter coefficients required by conventional filtering methods, are presented and discussed.

2 Multiple Space Optimization

In this section the basic approach and the used notation is presented. Descriptions of desired filter features can be specified simultaneously in multiple linear transform spaces here referred to as representation spaces. The representation spaces are defined by:

$C_0 = \{1 \dots N\}$;		One set of *kernel coefficient indices*
C_k;	$k = 1 \dots K$	K sets of *representation space coordinates*

The optimization produces a set of kernel coefficients, or simply a *kernel*, $\tilde{\mathbf{f}}_0$. The optimizer solves a least squares problem and can be seen as a function, i.e.

$$\tilde{\mathbf{f}}_0 \;=\; \mathbf{g}(\;\{\mathbf{f}_k\},\{\mathbf{w}_k\},\{\mathbf{B}_k\}\;)\;;\quad k \;=\; 0\ldots K; \tag{1}$$

The intended use of the arguments is explained by the following table:

$\mathbf{f}_0 = f_0(n)$;	$n \in C_0$	*Ideal kernel function*
$\mathbf{w}_0 = w_0(n)$;	$n \in C_0$	*Kernel weighting function*
$\mathbf{f}_k = f_k(\mathbf{c}_k)$;	$\mathbf{c}_k \in C_k$	*Ideal function in representation space k*
$\mathbf{w}_k = w_k(\mathbf{c}_k)$;	$\mathbf{c}_k \in C_k$	*Weighting function in representation space k*
$\mathbf{B}_k = b_k(\mathbf{c}_k, n)$;	$\mathbf{c}_k \in C_k$ $n \in C_0$	*Basis function matrix corresponding to representation space k.*

The discrete mapping from kernel space to representation space k is defined by:

$$\tilde{\mathbf{f}}_k \;\equiv\; \mathbf{B}_k\,\tilde{\mathbf{f}}_0 \tag{2}$$

Typical representation space examples are the Fourier space, the spatio-temporal space and wavelet spaces. (Note that $\mathbf{B}_0 = \mathbf{I}$, signifying that the kernel space is mapped onto itself.)

2.1 Error measure

The optimizer finds the kernel coefficients that minimizes a weighted distance measure. The definition of the chosen measure, the error: ε, is given by:

$$\varepsilon^2 \;=\; \sum_{k=0}^{K} \|\mathbf{W}_k\,(\mathbf{f}_k - \tilde{\mathbf{f}}_k)\|^2 \tag{3}$$

where the $\mathbf{W}_k$ are diagonal weighting matrices. The diagonal terms given by the corresponding weighting function.

2.2 Minimizing the error

To find the minimum of ε it is convenient to first rewrite equation (3) making the role of the kernel coefficients, $\tilde{\mathbf{f}}_\mathbf{0}$, explicit. The raised ^{T} denotes the conjugate transpose.

$$\varepsilon^2 \;=\; \sum_{k=0}^{K} (\mathbf{f}_k - \mathbf{B}_k\tilde{\mathbf{f}}_0)^T\mathbf{W}_k^2(\mathbf{f}_k - \mathbf{B}_k\tilde{\mathbf{f}}_0) \tag{4}$$

The error measure is quadratic and finding the minimum can be done by computing the partial derivatives of ε^2 with respect to the real and imaginary parts of $\tilde{\mathbf{f}}_\mathbf{0}$ and solving equation (5).

$$\frac{\partial\varepsilon^2}{\partial\tilde{\mathbf{f}}_0} \equiv \frac{\partial\varepsilon^2}{\partial\mathrm{Re}[\tilde{\mathbf{f}}_0]} + i\,\frac{\partial\varepsilon^2}{\partial\mathrm{Im}[\tilde{\mathbf{f}}_0]} \;=\; \mathbf{0} \tag{5}$$

Differentiating equation (4) and simplifying yields:

$$\begin{cases} \frac{\partial\varepsilon^2}{\partial\mathrm{Re}[\tilde{\mathbf{f}}_0]} \;=\; 2\;\mathrm{Re}\left[\sum_{k=0}^{K}\left(\mathbf{B}_k^T\mathbf{W}_k^2\mathbf{B}_k\tilde{\mathbf{f}}_0 - \mathbf{B}_k^T\mathbf{W}_k^2\mathbf{f}_k\right)\right] \\ \frac{\partial\varepsilon^2}{\partial\mathrm{Im}[\tilde{\mathbf{f}}_0]} \;=\; 2\;\mathrm{Im}\left[\sum_{k=0}^{K}\left(\mathbf{B}_k^T\mathbf{W}_k^2\mathbf{B}_k\tilde{\mathbf{f}}_0 - \mathbf{B}_k^T\mathbf{W}_k^2\mathbf{f}_k\right)\right] \end{cases} \tag{6}$$

Combining equation 6 with equation (5) results in

$$\sum_{k=0}^{K}\mathbf{B}_k^T\mathbf{W}_k^2\mathbf{B}_k\tilde{\mathbf{f}}_0 \;=\; \sum_{k=0}^{K}\mathbf{B}_k^T\mathbf{W}_k^2\mathbf{f}_k \tag{7}$$

Equation (7) can be simplified to:

$$\mathbf{A}\,\tilde{\mathbf{f}}_0 \;=\; \mathbf{h} \tag{8}$$

where:

$$\mathbf{A} = \sum_{k=0}^{K}\mathbf{B}_k^T\mathbf{W}_k^2\mathbf{B}_k \qquad \mathbf{h} = \sum_{k=0}^{K}\mathbf{B}_k^T\mathbf{W}_k^2\mathbf{f}_k \tag{9}$$

Solving for $\tilde{\mathbf{f}}$ gives the optimal kernel coefficients. The solution can be written:

$$\tilde{\mathbf{f}}_0 \;=\; \mathbf{A}^{-1}\,\mathbf{h} \tag{10}$$

In practise the solution is obtained by solving equation (8) since $\mathbf{A}^{-1}$ is not explicitly needed.

3 Basic Representation Spaces

The kernel domain, the spatio-temporal domain and the Fourier domain are the three most common representation spaces. In cases where the input is 'raw' spatio-temporal data the kernel space coordinates are simply a sub-set of the spatio-temporal coordinates. In general, however, this is not true and it is in most cases important to consider the desired properties of the optimized kernel in these three basic spaces. The role of the different representation spaces is described and discussed below.

3.1 The kernel space 'C_0'

In most cases the kernel space naturally inherits the spatio-temporal dimensions of the input signal (the dimensions over which the convolution is performed). The kernel space may in addition have extra 'dimensions' such as scale, bandwidth and orientation.

Note that, in general, the kernel space may be quite different from the spatio-temporal space. In extreme cases the kernel space coordinates may all have just one and the same spatio-temporal component thus effectively loosing the spatiotemporal dimensions. For this reason a particular coefficient is referred to by an index, n, rather than a coordinate. The description of the kernel space can in general only be attained through a basis function matrix relating the kernel space to a 'standard' space such as the spatio-temporal or Fourier space.

The optimized kernel is constituted by a set of optimal coefficients, $f_0(n);\;\; n \in \{1 \ldots N\}$. A fundamental constraint best expressed in the kernel domain is simply that the number of coordinates, N, is limited. (In fact it is the object of the present exercise to keep this number as low as possible.)

3.2 The spatio-temporal space 'C_1'

The spatio-temporal space is in most cases given by the input signal. A fundamental constraint naturally expressed in this space is that, in all digital applications, the signals are represented by samples. As a rule the samples are distributed in a cartesian fashion uniformly in each dimension.

For clarity and to adhere to common convention the spatio-temporal coordinate, $\mathbf{c}_1$, ideal function, $f_1(\mathbf{c}_1)$, and weighting function, $w_1(\mathbf{c}_1)$, will, when appropriate, also be denoted $\mathbf{x}$, $f(\mathbf{x})$, and $w(\mathbf{x})$ respectively.

3.2.1 Spatio-temporal locality

In most applications it is important that the filtered signal maintains a high spatio-temporal resolution. The *spatio-temporal weighting function*, $w(\mathbf{x})$, can be used to introduce a distance metric that favors/forces spatio-temporally localized kernels. An example of a suitable weighting function is given by: $w(\mathbf{x}) = |\mathbf{x}|^\gamma;\; \gamma > 0$.

The pro-locality feature is of particular interest in FFT based implementation where control of spatio-temporal locality is completely lost using standard approaches. In convolution based implementation the feature can be used to introduce a continuous metric as opposed to the 'hard metric' introduced by limited spatio-temporal kernel size.

3.3 The Fourier space 'C_2'

For the same reasons as in the spatio-temporal case the Fourier coordinate, $\mathbf{c}_2$, ideal function, $f_2(\mathbf{c}_2)$, and weighting function, $w_2(\mathbf{c}_2)$, will, when appropriate, also be denoted $\mathbf{u}$, $F(\mathbf{u})$, and $w(\mathbf{u})$ respectively.

In the present work the discrete Fourier space will be constructed by an appropriate sampling of the continuous Fourier space. The sampling topics are discussed in section 3.3.2. First, however, some important features of the Fourier space that are independent of the sampling will be discussed. The continuous Fourier space is defined via the Fourier transform and the spatio-temporal space:

$$\tilde{F}(\mathbf{u}) = \sum_{n=1}^{N} \tilde{f}(\mathbf{x}_n)\, \mathrm{e}^{-i\mathbf{u}\cdot\mathbf{x}_n} \tag{11}$$

Any given distribution of spatio-temporal coordinates, $\mathbf{x}_n$, can be interpreted as fundamental constraints on the Fourier space representation of attainable filter functions.

Smoothness constraint

The limited kernel size is equivalent to a smoothness constraint on the realizable filter functions. Taking the symmetry of the Fourier transform into account the effect of a limited spatio-temporal kernel size can be seen as a low pass filter acting in the frequency domain. Approximation of an ideal filter function containing rapid changes will, for this reason, always require a kernel of large spatio-temporal extent.

Repetetivity constraint

If the spatio-temporal space is sampled in a regular cartesian fashion and the inter sample distance in dimension i is given by Δ_i, then:

$$\tilde{F}(\mathbf{u}) \;=\; \tilde{F}(\,\mathbf{u} + \sum_i \frac{2\pi k_i\, \hat{\mathbf{u}}_i}{\Delta_i}\,) \tag{12}$$

where $\hat{\mathbf{u}}_i$ is a unit vector in the direction of the i-th dimension and the k_i:s are integers.

Equation (12) implies that the Fourier transform of the kernels will be repetitive in all dimensions. If the sampling distance in all dimensions is normalized to be unity the repetition period is 2π.

3.3.1 Fourier space metric

The *Fourier weighting function* $w(\mathbf{u})$ can be used to produce an appropriate Fourier space metric. The metric will determine the importance of a close fit for different spatial frequencies. In general the weighting function should be chosen based on all a-priori information available about the situations in which the optimized filters are to be used. (Note that optimizing without using a weighting function is equivalent to setting $w(\mathbf{u}) = 1$.)

The suggestion made here is that one major factor determining the importance of a close fit is the expected spectrum of the signal to be filtered. Below some general arguments concerning spectra of typical images are considered.

Expected spectra

For purely spatial signals there is, in general, no reason to expect a non-isotropic spectrum, i.e. the expected spectrum will only depend on the Fourier domain radius, $\rho = \|\mathbf{u}\|$. To find a suitable form for the expected spectrum two observations can be made. First, there does not seem to be a large difference in terms of spectrum when imaging the real world at very different scales, say a microscope image vs a satellite image. Secondly, for normal images the energy is usually concentrated around the origin and decreases as ρ increases. It can, based on these observations, be argued that a reasonable assumption is that the expected spectrum of spatial 2D or 3D data should be of the following form:

$$S(\rho) \propto S(a\rho); \qquad \forall a \tag{13}$$

This means that a scale change does not change the shape of the spectrum but is reflected as only a change of magnitude. A class of functions that exhibit both of the above properties is given by the following equation.

$$S(\rho) \propto \rho^{\alpha}; \qquad \alpha < 0 \tag{14}$$

Two examples of 2D images having this type of spectrum are given by:

1. A large number of random lines yields: $S(\rho) \propto \rho^{-0.5}$
2. A large number of random edges yields: $S(\rho) \propto \rho^{-1.5}$

Another fact about digital images is that they have been obtained through a cartesian sampling procedure. Three side effects influencing the spectrum may be caused by this procedure:

a) The data may have been band-limited, b) There may be aliased signal contributions and b) Noise may have been added,

In most cases it will not be possible to define an 'optimal' weighting function but the above discussion provides a useful background for making an appropriate choice. Equation(15) gives a class of weight functions that provides a reasonable degree of variability and has been proven useful in practice. The "cosine-terms" relate to the degree of band limiting of the signal. (The separability of this factor corresponds to the standard case of cartesian separable band limiting.) The constant c relates to the expected level of broadband noise and aliasing.

$$w(\mathbf{u}) \;=\; \rho^{\alpha} \prod_i \cos^{\beta}\!\left(\tfrac{u_i}{2}\right) \;+\; c \tag{15}$$

Note that $w(\mathbf{u}) \to \infty$ when $\rho \to 0$, this is equivalent to the constraint that the sum of the kernel coefficients should be exactly equal to the value given by the ideal function at $\rho = 0$. In most cases this is a desired feature, in particular for bandpass or highpass type filters where ideally $\tilde{F}(0) = 0$ and even small deviations from zero can cause major problems.

3.3.2 Fourier space sampling

Up til now the Fourier space has been treated as continuous. However, to find an optimal filter in practice implies sampling the Fourier space. In the present work the Fourier space sampling is performed in a regular cartesian manner.

In principle the higher the sampling density the 'closer' the sampled case solution will be to the continuous case. In practise using 2-3 times as many points, for each dimension, as the spatial size in pixels (voxels etc.) has proven to be adequate. Note that the number of samples does not change the size of the basic problem, i.e. the size of the matrix $\mathbf{A}$ in equation(8). However, further increasing the sample density will, as a rule, have an insignificant effect on the solution.

As the Fourier space representation will be repetitive samples will only be needed in the interval $-\pi/\Delta_i < u_i \leq \pi/\Delta_i$. Note that specifying the ideal Fourier function only in this interval may 'hide' discontinuities that are present at the interval boarders.

It is important that the sampling is regular in a circular fashion, i.e. it must hold that the distances between the samples in Fourier space are $2\pi/(J_i\Delta_i)$ where J_i is the number of samples in dimension i. Violation of this rule will cause the metric to deviate from what is specified using the weighting functions.

Since it is preferable to have one sample point at $\mathbf{u} = \mathbf{0}$, this Fourier coordinate often having particular significance, the sampling pattern is completely given by the number of samples in each dimension, J_i.

4 Distortion Measures

Assessing the quality of an optimized kernel is difficult using only the distance measure since it is an absolute measure and not directly related to the kernel quality. For a quick quality assessment it is helpful to calculate a few distortion measures.

Normalized error

An overall distortion measure is obtained simply by normalizing the error measure, i.e.

$$\delta \;=\; \sqrt{\frac{\sum_{k=0}^{K} \|\mathbf{W}_k\,(\mathbf{f}_k - \tilde{\mathbf{f}}_k)\|^2}{\sum_{k=0}^{K} \|\mathbf{W}_k\,\mathbf{f}_k\|^2}} \tag{16}$$

Distortion in space k

A somewhat more detailed information about the outcome of the optimization can be obtained by calculating distortion measures separately for each representation space. In analogy with equation(16) the distortion in space k is defined:

$$\delta_k \;=\; \frac{\|\mathbf{W}_k\,(\mathbf{f}_k - \tilde{\mathbf{f}}_k)\|}{\|\mathbf{W}_k\,\mathbf{f}_k\|} \tag{17}$$

Equation (17) gives the distortion in space k in the metric given by the weighting function, $\mathbf{w}_k$. The distortion gives the ratio between the RMS error and the RMS value of the ideal function, $\mathbf{f}_k$, and is invariant to scaling of $\mathbf{w}_k$ and/or $\mathbf{f}_k$.

Shape distortion

A measure that in addition to the invariances of equation (17) also is invariant to scaling of $\tilde{\mathbf{f}}$ is given by:

$$\sigma_k \;=\; \sin(\varphi_k) \;=\; \sqrt{1 - \frac{\mathrm{Re}[\,\tilde{\mathbf{f}}_k^T \mathbf{W}_k^2 \mathbf{f}_k\,]^2}{\tilde{\mathbf{f}}_k^T \mathbf{W}_k^2 \tilde{\mathbf{f}}_k \quad \mathbf{f}_k^T \mathbf{W}_k^2 \mathbf{f}_k}} \tag{18}$$

This measure only depends on the angle, φ_k, between the 1-D subspaces spanned by $\mathbf{f}_k$ and $\tilde{\mathbf{f}}$ and can be thought of as a shape distortion measure.

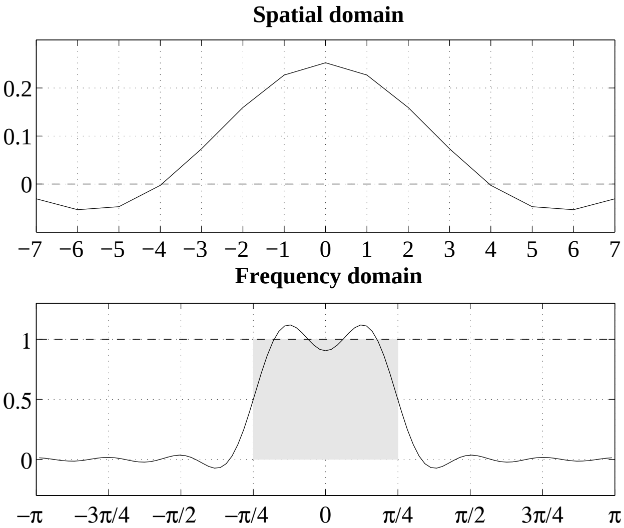

Figure 1: Optimization result using only a constant Fourier weighting function, i.e effectively equivalent to standard DFT.

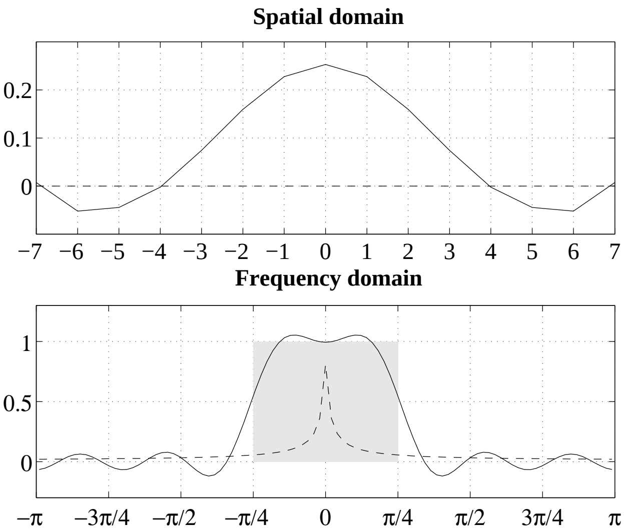

Figure 2: Optimization result using a Fourier weighting function corresponding to monotonically decreasing signal spectrum: $w(u) \propto S(u) \propto |u|^{-1}$.

5 Single Filters

In this section a few examples of optimization of simple single filters are given. The examples illustrates the basic features of the optimizer and the effect of different metrics. To simplify visualization of the results all optimized kernels are one-dimensional. Note, however, that the method is completely 'invariant' to spatio-temporal dimensionality.

The ideal Fourier filter used in all four examples is a 'box' given by: $F(u) = 1$ if $|u| < \pi/4$ and $F(u) = 0$ otherwise. Figures 1 to 4 shows how changing the spatio-temporal and Fourier weighting functions can be used to design a suitable metric and obtain a filter having the desired features. The

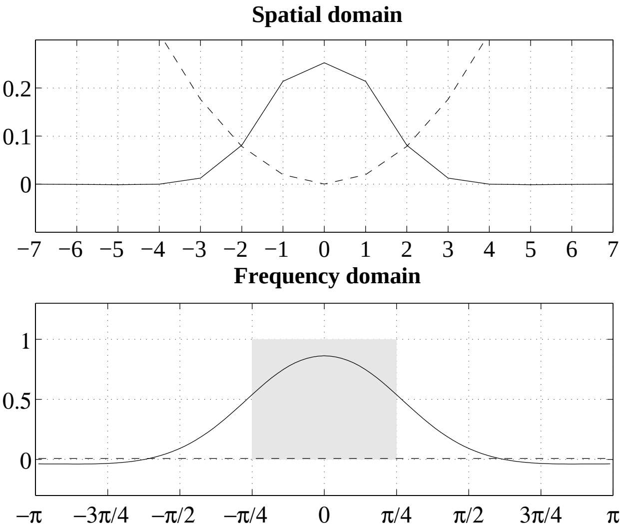

Figure 3: Optimization result using a spatiotemporal weighting function favoring locally concentrated kernels: $w(x) \propto x^2$.

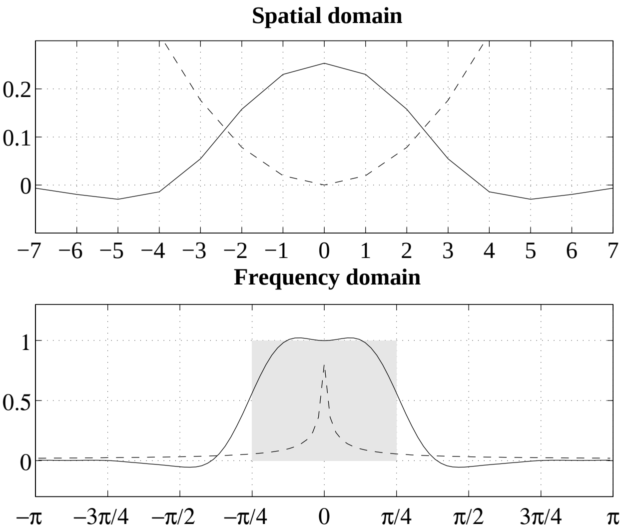

Figure 4: Optimization result using both Fourier and spatiotemporal weighting functions: $w(u) \propto |u|^{-1}$, $w(x) \propto x^2$.

shaded areas indicate the ideal Fourier function, the solid lines shows the optimization results and the dashed lines shows the weighting functions.

6 Sequential Filters

Using N coefficients it is always possible to find a best approximation to a given filter using all coefficients at once for a single filter. In many cases, however, a far more efficient way of attaining essentially the same filter is to distribute coefficients over a number of filters. These filters are then applied in sequence to obtain the final filter response.

Using this technique it is in many cases possible to attain equally good filter approximations using only a fraction of the the number of coefficients required for the single filter approach. In many situations it is in this way possible to reduce computational load by 2-digit numbers.

The proposed method involves the following four steps:

1. Chose the number, M, of filters in sequence that is likely to be appropriate in the present situation.
2. Chose the number of coefficients, N_m, to be used by each filter in the sequence, $m = 1..M$.
3. Chose the spatio-temporal coordinate for each coefficient in the M filters.
4. Optimize the values of the $N = \sum_m N_m$ coefficients (distributed over the M filters) so that the combined effect of the filter sequence approximates the ideal filter as closely as possible.

Coefficient distribution

Ideally it would be desirable to optimize all four steps jointly. This is, however, an extremely complex problem and a method to for finding an overall optimal solution has not been found. (and it's doubtful if it ever will be). For this reason the choices in steps 1-3 have to be made based on experience.

A general approach for distributing the coefficients is to let the component filters operate on different 'scales'. The coefficients can, for example, be spread equidistantly within each component filter using different distances for different filters, e.g. Δ for the first component filter, 2Δ for the next etc.

6.1 Products in Fourier space

The final step in the sequential filter optimization procedure is to find the values of all N coefficients of the sequential filter such that the difference between the reference function, $F(\mathbf{u})$, and $\tilde{F}(\mathbf{u})$ is minimized according to the distance measure.

The required analysis is best carried out in the Fourier domain as the effect of sequentially applied filters is obtained by simple multiplication of individual filter responses.

A sequential filter, $\tilde{F}(\mathbf{u})$ is to approximate an ideal filter, $F(\mathbf{u})$, by M sequential filter components, $\tilde{F}_m(\mathbf{u})$. In the Fourier domain this is expressed as:

$$F(\mathbf{u}) \approx \tilde{F}(\mathbf{u}) = \prod_{m=1}^{M} \tilde{F}_m(\mathbf{u}) \tag{19}$$

Again, the motivation for this operation is that the filters $F_m(\mathbf{u})$ can potentially be implemented using a considerable smaller number of kernel coefficients than a direct ($M = 1$) implementation of $F(\mathbf{u})$. In the spatial domain eq. (19) corresponds to:

$$\tilde{f}(\mathbf{x}) = \tilde{f}_1(\mathbf{x}) * \tilde{f}_2(\mathbf{x}) * \ldots * \tilde{f}_M(\mathbf{x}) \tag{20}$$

6.2 Recursive Filter Optimization

Even if the number and spatio-temporal positions of all coefficients are considered given optimizing a sequential filter is no longer a quadratic problem and finding the optimum must be done using an iterative search.

A method that has proven to work well in practice is to optimize a single filter component with respect to the present value of the other $M-1$ components. In this way a 'rotating' recursion of the single filter optimization method can be used to rapidly find a set of close to optimal component filters. Convergence of the algorithm is fast and typically less than 10 iterations are needed.

6.2.1 Modified weighting function view

To convert the kernel optimizer for recursive use requires some consideration as the ideal filter function and weight function for a certain filter component depends on the current value of the other $M-1$ filter components as well as the reference function $F(\mathbf{u})$. The case where the Fourier weighting function is the only non-zero weighting function will be considered first. The distance measure is then given by:

$$\varepsilon^2 \ = \ \|\mathbf{W}_2\,(\mathbf{f}_2 - \tilde{\mathbf{f}}_2)\|^2 \tag{21}$$

or more explicitly,

$$\varepsilon^2 \ = \ \sum_{C_2} \|w(\mathbf{u})(F(\mathbf{u}) - \tilde{F}(\mathbf{u}))\|^2 \tag{22}$$

In the ideal case, i.e. $\varepsilon = 0$, the following equation holds:

$$F(\mathbf{u}) = \prod_{l=1}^{M} \tilde{F}_l(\mathbf{u}) \tag{23}$$

Consider the optimization of filter component $m \in [1, M]$. If all other filter components are considered given then the ideal filter function for the m:th filter component can be obtained using equation(23).

$$F_m(\mathbf{u}) = \frac{F(\mathbf{u})}{\prod_{l \neq m} \tilde{F}_l(\mathbf{u})} \tag{24}$$

For the weight function $w(\mathbf{u})$ the relation is not quite as obvious as for the ideal function. A suitable weight function can, however, be found if the reasoning for obtaining a weight function in the single filter case is applied. The signal spectrum will, for a component filter, effectively be multiplied with the filtering effect of all other filters in the sequence.

(If, for example, the other filters cause a zero for some frequency it is obvious that trying to change this fact in the optimization of the present component filter would not be a good idea.) A weight function consistent with this reasoning is for filter component m given by:

$$w_m(\mathbf{u}) = w(\mathbf{u}) \prod_{l \neq m} \tilde{F}_l(\mathbf{u}) \tag{25}$$

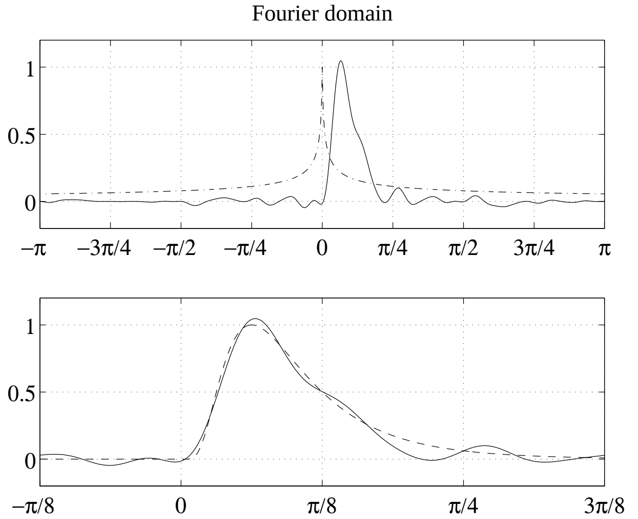

Figure 5: Optimization result for sequential filter consisting of four component filters. The implementation of the filter requires the equivalent of 18 complex coefficients. The result is shown at two different scales. The solid lines shows the resulting filter in Fourier space. The dashed line shows the ideal function (lognormal). The dash dotted line shows the weighting function $w(u) \propto |u|^{-0.5}$. Fourier space distortion: $\delta_2 = 0.1$.

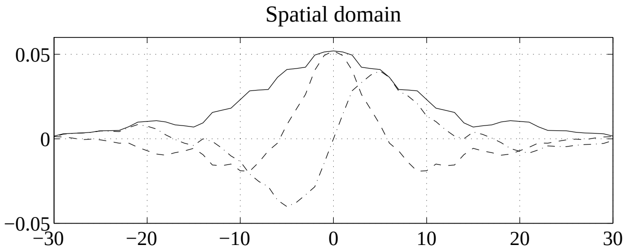

Figure 6: Spatial view of the optimization result in figure 5. Solid line shows filter magnitude. Dashed line shows real part. Dash dotted line shows imaginary part

If the ideal component filter, equation(24), and weight function, equation(25), are inserted in eq. (22) the resulting distance measure for filter component $\tilde{f}_m(\mathbf{u})$ results in:

$$\varepsilon^2 = \sum_{C_2} \|w(\mathbf{u})(\prod_{l \neq m} \tilde{F}_l(\mathbf{u}))\Big[\frac{F(\mathbf{u})}{\prod_{l \neq m} \tilde{F}_l(\mathbf{u})} - \tilde{F}_m(\mathbf{u})\Big]\|^2 \quad (26)$$

Simplifying eq. (26) yields:

$$\varepsilon^2 = \sum_{C_2} \|w(\mathbf{u})\,[F(\mathbf{u}) - \prod_{l=1}^{M} \tilde{F}_l(\mathbf{u})]\|^2 \quad (27)$$

showing that the recursive optimization procedure will tend to minimize the distance between the combined effect of the component filters, $\tilde{F}_l(\mathbf{u})$, and the ideal filter $F(\mathbf{u})$.

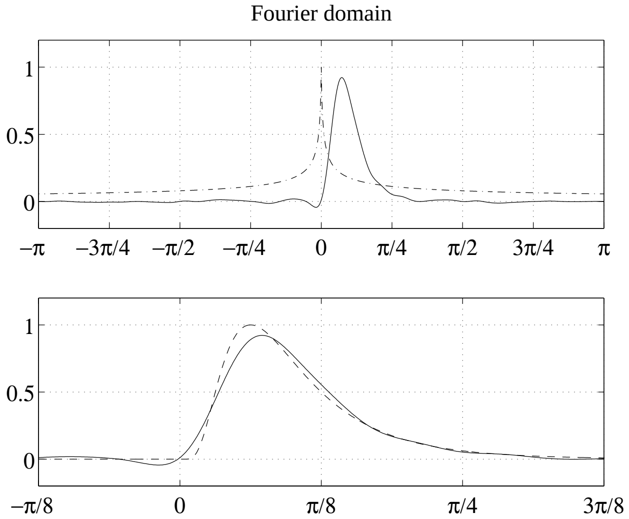

Figure 7: Optimization result for same sequential filter setup as in figure 5 but including spatial weighting function $w(x) \propto x^2$

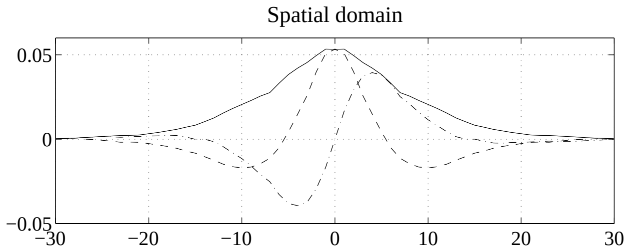

Figure 8: Spatial view of the optimization result in figure 7. Spatial weighting function $w(x) \propto x^2$

6.2.2 Modified basis view

Another way of viewing the situation emerges if $\tilde{F}_m(\mathbf{u})$ is expressed using a more explicit version of equation(2):

$$\tilde{F}_m(\mathbf{u}) \equiv \sum_n b_2(\mathbf{u}, n)\,\tilde{f}_{0m}(n) \quad (28)$$

The filter product can then be written:

$$\prod_{l=1}^{M} \tilde{F}_l(\mathbf{u}) = \prod_{l \neq m}^{M} \tilde{F}_l(\mathbf{u}) \sum_n b_2(\mathbf{u}, n)\,\tilde{f}_{0m}(n) \quad (29)$$

Moving the summation over n outside the product yields:

$$\prod_{l=1}^{M} \tilde{F}_l(\mathbf{u}) = \sum_n \prod_{l \neq m}^{M} \tilde{F}_l(\mathbf{u}) b_2(\mathbf{u}, n)\,\tilde{f}_{0m}(n) \quad (30)$$

Finally, using equations (27), (28), (30) and defining:

$$b_{2m}(\mathbf{u}, n) \equiv \prod_{l \neq m}^{M} \tilde{F}_l(\mathbf{u}) b_2(\mathbf{u}, n) \quad (31)$$

the error can be written as:

$$\varepsilon^2 = \sum_{C_2} \|w(\mathbf{u})\,[F(\mathbf{u}) - \sum_n b_{2m}(\mathbf{u}, n)\,\tilde{f}_{0m}(n)]\|^2 \quad (32)$$

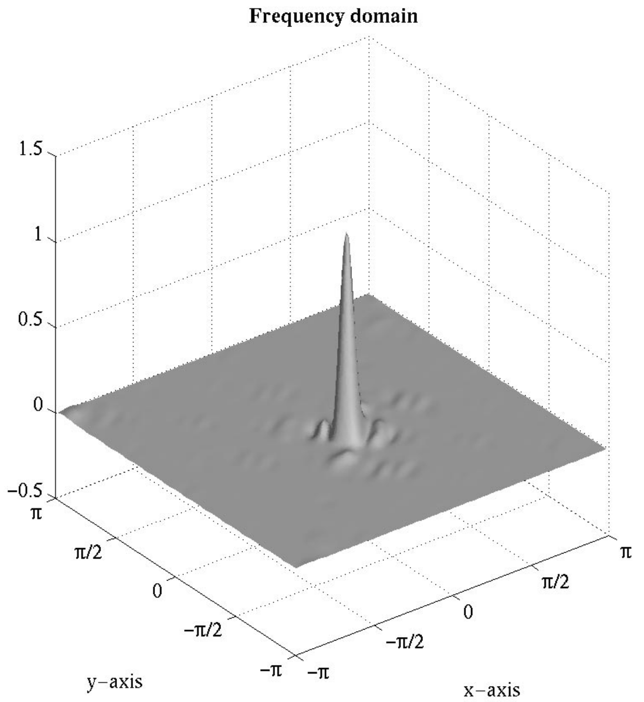

Figure 9: Fourier space view of optimization result for 2D sequential filter consisting of 12 component filters. Using this filter requires the equivalent of 24 complex multiplications per pixel. $w(\mathbf{u}) \propto \|\mathbf{u}\|^{-1}$ and $w(\mathbf{x}) = 0$.

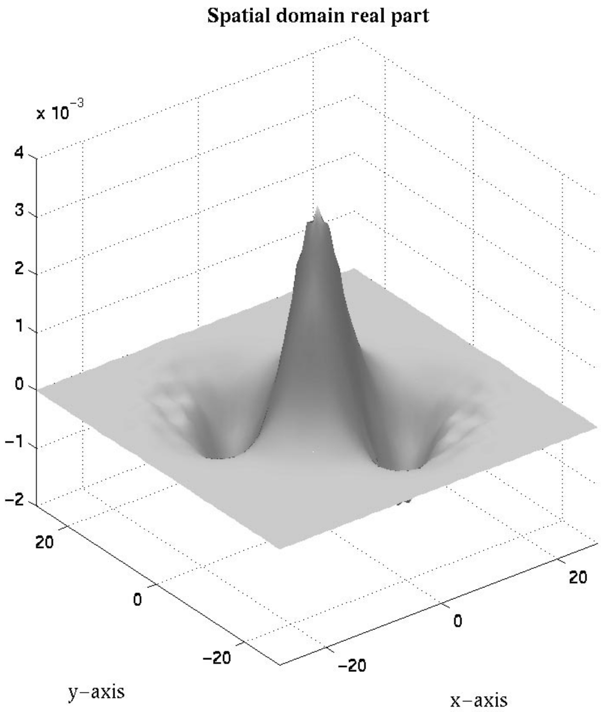

Figure 10: Spatial view of optimization result for the filter shown in figure 9.

or equivalently using matrix notation:

$$\varepsilon^2 = \|\mathbf{W}_2(\mathbf{f}_2 - \mathbf{B}_{2m}\tilde{\mathbf{f}}_{0m})\|^2 \tag{33}$$

where $\mathbf{B}_{2m}$ is a matrix with elements given by $b_{2m}(\mathbf{u}, n)$.

This shows that the effect of the other filters in the sequence can also be seen as modifiers of the basis function matrix where the modified matrix is given by equation(31). This formulation avoids the use of a division, (equation(26)), and was used for the implementation of the optimizer.

6.2.3 Adding spatially specified metric

Adding a spatial weighting function can now be done by simply transforming the component filter $\mathbf{B}_{2m}\tilde{\mathbf{f}}_{0m}$ (eq: 33) to the spatial domain. Denoting the matrix corresponding to this transformation, $\mathbf{B}_{12}$ yields:

$$\varepsilon^2 = \|\mathbf{W}_2(\mathbf{f}_2 - \mathbf{B}_{2m}\tilde{\mathbf{f}}_{0m})\|^2 + \|\mathbf{W}_1(\mathbf{f}_1 - \underbrace{\mathbf{B}_{12}\mathbf{B}_{2m}}_{\mathbf{B}_{1m}}\tilde{\mathbf{f}}_{0m})\|^2 \tag{34}$$

Thus, in general, the error measure for a component filter can be written:

$$\varepsilon^2 = \sum_{k=0}^{K} \|\mathbf{W}_k(\mathbf{f}_k - \mathbf{B}_{km}\tilde{\mathbf{f}}_{0m})\|^2 \tag{35}$$

where $\mathbf{B}_{km}$ is the matrix transforming from kernel space to space k for component filter m.

This result makes it possible to benefit from the advantages of multiple domain weighting functions also in the design of sequential filters.

6.3 Optimization results

To demonstrate the potential of the sequential filtering approach the result of two optimizations are shown in figures 5 to 10. Figures 5 to 8 show an optimized 1D sequential filter consisting of four component filters. Each component filter has five coefficients, the coefficients being spread by 1, 2, 4 and 8 spatio-temporal sample distance units respectively. Using the filter requires the equivalent of 18 complex multiplications per output value. This implies that, for a 20 second signal sampled at 50kHz, this filter implementation reduces the number of multiplications compared to standard convolution **and** FFT based approaches by factors 4 and 5 respectively.

Figures 7 and 8 demonstrate the the effect of using a spatial weighting function to improve spatial locality for the sequential filter.

Figures 9 and 10 show an optimized 2D sequential filter consisting of 12 component filters. Each component filter has five coefficients. The coefficients are placed along lines at 0, 45, 90 and 135 degrees each direction spread by 1, 2 and 4 spatio-temporal sample distance units respectively. Most of the coefficients can be real-valued and using the filter only requires the equivalent of 24 complex multiplications per output pixel. In other words, for a 512×512 image, the implementation of the filter *outperforms standard convolution and FFT by factors exceeding **30** and **10** respectively.*

7 Conclusion

A novel method for optimizing efficient filters and filter sequences has been presented. The method allows specification of the metric via simultaneous weighting functions in multiple domains, e.g. the spatio-temporal space *and* the Fourier space. This 'designer metric' approach is general in that it is implementation independent, i.e. the chosen metric 'tells it all' and consequently the design is equally valid if the filters are implemented as convolutions, using an FFT based approach or any other technique. The methods capability to simultaneously improve spatio-temporal and Fourier domain filter properties has been demonstrated in a number of examples.

It was shown how convolution kernels for efficient spatio-temporal filtering can be implemented in practical situations. Combining the sequential filtering method and spatially sparse kernels efficient filtering can be performed. It was demonstrated that the method is capable of outperforming both standard convolution and FFT based approaches by two-digit numbers.

The presented method is equally applicable to filters of any dimensionality. Although no 3D or 4D examples has been presented here, initial results indicate that computational gain will increase even further for higher filter dimensionality.

8 Acknowledgment

The support from the Swedish National Board for Technical Development and WITAS, The Wallenberg Laboratory for Technical Development and Autonomous Systems is gratefully acknowledged. The authors also wish to thank the members of our computer vision group for many inspiring discussions.

References

[1] M. Andersson. *Controllable Multidimensional Filters in Low Level Computer Vision*. PhD thesis, Linköping University, Sweden, SE-581 83 Linköping, Sweden, September 1992. Dissertation No 282, ISBN 91-7870-981-4.

[2] R. Bracewell. *The Fourier Transform and its Applications*. McGraw-Hill, 2nd edition, 1986.

[3] D. E. Dudgeon and R. M. Mersereau. *Multidimensional Digital Signal Processing*. Prentice-Hall signal processing series. Prentice-Hall, 1984. ISBN 0-13-604959-1.

[4] G. H. Granlund and H. Knutsson. *Signal Processing for Computer Vision*. Kluwer Academic Publishers, 1995. ISBN 0-7923-9530-1.

[5] H. Knutsson. *Filtering and Reconstruction in Image Processing*. PhD thesis, Linköping University, Sweden, 1982. Diss. No. 88.

[6] H. Knutsson and M. Andersson. Optimization of Sequential Filters. In *Proceedings of the SSAB Symposium on Image Analysis*, pages 87–90, Linköping, Sweden, March 1995. SSAB. LiTH-ISY-R-1797. URL: http://www.isy.liu.se/cvl/ScOut/TechRep/TechRep.html.

[7] J. Wiklund and H. Knutsson. A Generalized Convolver. In *Proceedings of the 9th Scandinavian Conference on Image Analysis*, Uppsala, Sweden, June 1995. SCIA.