Multiple Space Filter Design

Hans Knutsson   Mats Andersson   Johan Wiklund

Computer Vision Laboratory

Department of Electrical Engineering

Linköping University, S-581 83 Linköping, Sweden

Phone: +46 13 281887, Fax: +46 13 138526, email: knutte@isy.liu.se

1 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.

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. In multidimensional filtering applications the method potentially outperforms both standard convolution and FFT based approaches by 2-digit numbers.

2 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.

3 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:

$C_0 = \{1 \ldots N\};$		One set of *kernel coefficient indices*
$C_k;$	$k = 1 \ldots K$	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.

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

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:

$$\tilde{\mathbf{f}}_k \equiv \mathbf{B}_k \tilde{\mathbf{f}}_0 \quad (2)$$

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.)

3.1 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:

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

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

3.2 Minimizing the error

To find the minimum of ε it is convenient to first rewrite equation (3) making the role of the kernel coefficients, $\tilde{\mathbf{f}}_0$, explicit. The raised ^{T} denotes the conjugate transpose.

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

The error measure is quadratic and finding the minimum can be done by computing the partial derivatives of ε^2 with respect to the kernel coefficients $\tilde{\mathbf{f}}_0$ and solving equation (5). Differentiating equation (4) yields:

$$\frac{\partial \varepsilon^2}{\partial \tilde{\mathbf{f}}_0} = 2 \sum_{k=0}^{K} \mathbf{B}_k^T \mathbf{W}_k^2 (\mathbf{B}_k \tilde{\mathbf{f}}_0 - \mathbf{f}_k) = 0 \quad (5)$$

which results in:

$$\sum_{k=0}^{K} \mathbf{B}_k^T \mathbf{W}_k^2 \mathbf{B}_k \tilde{\mathbf{f}}_0 = \underbrace{\sum_{k=0}^{K} \mathbf{B}_k^T \mathbf{W}_k^2 \mathbf{f}_k}_{\mathbf{h}} \qquad (6)$$

$$\tilde{F}(\mathbf{u}) = \sum_{n=1}^{N} \tilde{f}(\mathbf{x}_n)\, e^{-i\mathbf{u}\cdot\mathbf{x}_n} \qquad (8)$$

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.

Equation (6) can be simplified to:

$$\mathbf{A}\,\tilde{\mathbf{f}}_0 = \mathbf{h} \qquad (7)$$

Solving for $\tilde{\mathbf{f}}_0$ gives the optimal kernel coefficients.

4 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.

4.1 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.

The optimized kernel is constituted by a set of optimal coefficients, $f_0(n)$; $n \in \{1 \dots N\}$. 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.)

4.2 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.

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.

4.3 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. The continuous Fourier space is defined via the Fourier transform and the spatio-temporal space:

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.

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. Equation(9) gives a class of weight functions that provides a reasonable degree of variability and has been proven useful in practice.

$$w(\mathbf{u}) = \rho^{\alpha} \prod_i \cos^{\beta}(\tfrac{u_i}{2}) + c \qquad (9)$$

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.

5 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.

$$\delta = \sqrt{\frac{\sum_{k=0}^{K} \|\mathbf{W}_k\,(\mathbf{f}_k - \tilde{\mathbf{f}}_k)\|^2}{\sum_{k=0}^{K} \|\mathbf{W}_k\,\mathbf{f}_k\|^2}} \qquad (10)$$

Distortion in space k

A somewhat more detailed information about the outcome of the optimization can be obtained by calculating distortion
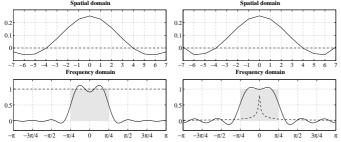
Figure 1: Left:Optimization result using a constant Fourier weighting function, i.e effectively equivalent to standard DFT. Right:Optimization result using a Fourier weighting function corresponding to monotonically decreasing signal spectrum: $w(u) \propto S(u) \propto |u|^{-1}$.
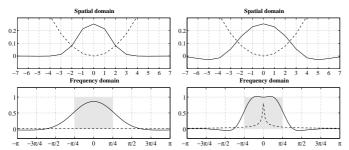

Figure 2: Left: Optimization result using a spatiotemporal weighting function favoring locally concentrated kernels: $w(x) \propto x^2$. Right: Optimization result using both Fourier and spatiotemporal weighting functions: $w(u) \propto |u|^{-1}$, $w(x) \propto x^2$.

Figure 3: 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: $\delta_2 = 0.1$.

measures separately for each representation space. In analogy with equation(10) the distortion in space k is defined:

$$\delta_k = \frac{\|\mathbf{W}_k (\mathbf{f}_k - \tilde{\mathbf{f}}_k)\|}{\|\mathbf{W}_k \mathbf{f}_k\|} \qquad (11)$$

Equation (11) 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 $\mathbf{f}_k$.
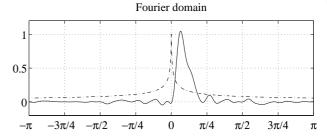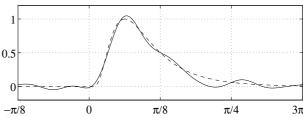
6 Single Filters

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.

Figure 1 and 2 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 shaded areas indicate the ideal Fourier function, the solid lines shows the optimization results and the dashed lines shows the weighting functions.

7 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.

7.1 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:

$$F(\mathbf{u}) \approx \tilde{F}(\mathbf{u}) = \prod_{m=1}^{M} \tilde{F}_m(\mathbf{u}) \qquad (12)$$

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 $F(\mathbf{u})$.
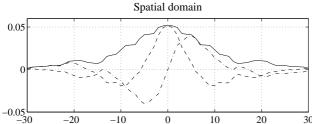
Figure 4: Spatial view of the optimization result in figure 3. Solid line shows filter magnitude. Dashed line shows real part. Dash dotted line shows imaginary part
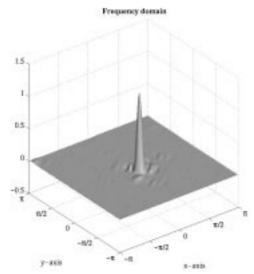

Figure 5: Fourier space view of optimization result for 2D sequential filter consisting of 12 component filters. Using this filter requires the equivalent of 24 complex multiplications per pixel. $w(\mathbf{u}) \propto \|\mathbf{u}\|^{-1}$ and $w(\mathbf{x}) = 0$.
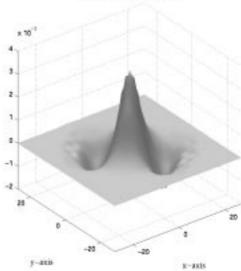
7.2 Recursive Filter Optimization

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. For a theoretical analysis of how to convert the kernel optimizer for recursive use see [7].

7.3 Optimization results

To demonstrate the potential of the sequential filtering approach the result of two optimizations are shown in figures 3 to 6. Figures 3 and 4 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 5 and 6 show an optimized 2D sequential filter consisting of 12 component filters. Each component filter has

Figure 6: Spatial view of optimization result for the filter shown in figure 5.

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.*

8 Acknowledgment

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] H. Knutsson, M. Andersson, and J Wiklund. Advanced Filter Design. In *Proceedings of the Scandinavian Conference on Image analysis*, June 1999.

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