Spatio-featural scale-space

Michael Felsberg

Original Publication:

Michael Felsberg, , Spatio-featural scale-space, 2009, Scale Space and Variational Methods in Computer Vision in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), volume 5567, 808-819.
http://dx.doi.org/10.1007/978-3-642-02256-2_67

Spatio-Featural Scale-Space*

Michael Felsberg

Computer Vision Laboratory, Linköping University, S-58183 Linköping, Sweden
mfe@isy.liu.se

Abstract. Linear scale-space theory is the fundamental building block for many approaches to image processing like pyramids or scale-selection. However, linear smoothing does not preserve image structures very well and thus non-linear techniques are mostly applied for image enhancement. A different perspective is given in the framework of channel-smoothing, where the feature domain is not considered as a linear space, but it is decomposed into local basis functions. One major drawback is the larger memory requirement for this type of representation, which is avoided if the channel representation is subsampled in the spatial domain. This general type of feature representation is called channel-coded feature map (CCFM) in the literature and a special case using linear channels is the SIFT descriptor. For computing CCFMs the spatial resolution and the feature resolution need to be selected.
In this paper, we focus on the spatio-featural scale-space from a scale-selection perspective. We propose a coupled scheme for selecting the spatial and the featural scales. The scheme is based on an analysis of lower bounds for the product of uncertainties, which is summarized in a theorem about a spatio-featural uncertainty relation. As a practical application of the derived theory, we reconstruct images from CCFMs with resolutions according to our theory. The results are very similar to the results of non-linear evolution schemes, but our algorithm has the fundamental advantage of being non-iterative. Any level of smoothing can be achieved with about the same computational effort.

1 Introduction

The concept of *scale* is a central ingredient to many image analysis and computer vision algorithms. Scale was first introduced systematically in terms of the concept of *linear scale-space* [1–3], establishing a 3D space of spatial coordinates and a scale coordinate. Often identified with Gaussian low-pass filtering, a rigorous analysis of underlying scale-space axioms [4] has led to the discovery of the Poisson scale-space [5] and more general α scale-spaces [6].

In practice, discrete scale-spaces are mostly sub-sampled with increasing scale parameter, leading to the concept of scale-pyramids [7, 8], multi-scale analysis

* The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 215078 (DIPLECS).

and wavelet theory [9, 10]. While pyramids and wavelets speedup the computation of linear operators and transforms, non-linear scale-space methods are widely used, e.g. for image enhancement. Non-linear scale-space is based on a non-stationary or anisotropic diffusivity function [11, 12].

More recently, non-linear methods have been introduced which are less directly connected to linear scale-space space and diffusion, but allow for faster processing and partially superior results [13, 14]. The former method is based on wavelets, whereas the latter one is based on the channel representation [15] and is called channel smoothing. Combining the channel representation with a systematic decimation of spatial resolution, similar to the pyramid approach, has been applied in blob-detection [16] and in channel-coded feature maps (CCFM) [17, 18], a density representation in spatio-featural domain, see also [19].

In this paper, we propose a new spatio-featural scale-space approach including an image reconstruction algorithm, which generates images from CCFMs. The CCFM scale-space is generated by applying the principles of linear scale-space to the spatial resolution of CCFMs and simultaneously increasing the resolution of feature space. By subsampling this space and subsequent reconstruction, image evolutions are generated which are very similar to those generated by iterative methods. We show some examples and propose a scale-selection scheme based on a new uncertainty relation: the spatio-featural uncertainty relation.

In the Section 2, we introduce lesser known relevant techniques: channel representation, channel smoothing, CCFMs. In Section 3 we propose the novel reconstruction algorithm, define the linear scale-space of CCFMs, and formulate a scale-selection scheme based on a spatio-featural uncertainty relation. In Section 4 we present experimental results and in Section 5 we give some concluding remarks.

2 Required Methods

2.1 The channel representation

Channel coding, also called population coding [20, 21], is a biologically inspired data representation, where features are represented by weights assigned to ranges of feature values [22, 15], see Fig. 1. Similar feature representations can also be found in the visual cortex of the human brain, e.g. in the cortical columns.

The closer the current feature value f to the respective feature interval center n, the higher the channel weight c_n:

$$c_n(f) = k(f - n) \ , \tag{1}$$

where $k(\cdot)$ is a suitable kernel function and where f has been scaled such that it has a suitable range (note that we chose to place the channel centers at integers). By introducing z as a continuous feature coordinate, $k_n(z) = k(z - n)$, and $\delta_f(z) = \delta(z - f)$ denoting the Dirac-delta at f, the encoding can be written as a scalar product

$$c_n(f) = \langle \delta_f | k_n \rangle = \int \delta_f(z) k_n(z) \, dz \tag{2}$$

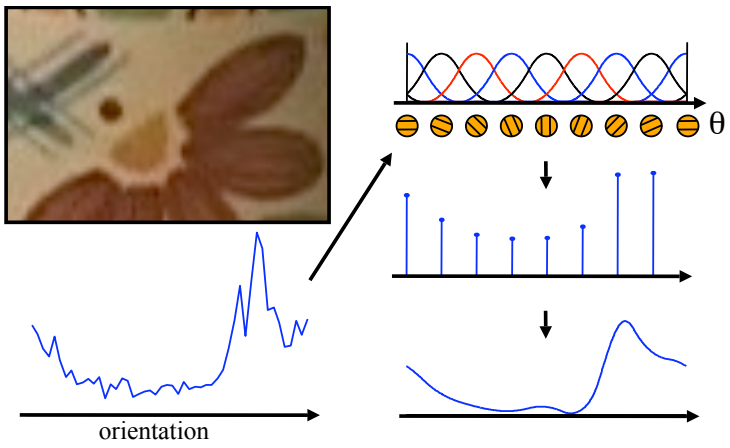

Fig. 1. Orientation distribution is encoded into channels, resulting in a (low-pass filtered) reconstruction. Figure courtesy Erik Jonsson.

or as a sampled correlation in the feature-domain:

$$c_n = (\delta_f \star k)(n) = \int \delta_f(z')k(z'-z)\,dz'\Big|_{z=n} \quad . \tag{3}$$

From the weights of all channels the feature value can be decoded unambiguously by finding the mode, where the decoding depends on the kernel function. In some theoretic considerations we will consider Gaussian functions as kernels but in the practical implementation we have been using quadratic B-splines:

$$B_2(z) = \begin{cases} (z+3/2)^2/2 & -3/2 < z \leq -1/2 \\ 3/4 - z^2 & -1/2 < z \leq 1/2 \\ (z-3/2)^2/2 & 1/2 < z < 3/2 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

The features can be scalar valued or vector valued, e.g. grey-scales, color vectors, or orientations. In the case of scalar features the decoding from quadratic B-splines has been considered in detail in [14], which we will not repeat here. For the case of non-interfering channel weights, a simplified scheme based on the quotient of linear combinations can be used:

$$M_n = c_{n-1} + c_n + c_{n+1} \qquad n_0 = \arg\max M_n \qquad \hat{f} = \frac{c_{n_0+1} - c_{n_0-1}}{M_{n_0}} + n_0 \tag{5}$$

where $\hat{f}$ is our estimate of the feature f that had been encoded in c_n.

Channel representations obviously need more memory than directly storing features, but this investment pays off in several ways which we will show in the subsequent sections.

2.2 Channel Smoothing and Channel-Coded Feature Maps

The idea of channel smoothing is based on considering the feature f in the encoding (1) as a stochastic variable. It has been shown in [14] that the distribution p_f is approximated by c_n in expectation sense:

$$E\{c_n(f)\} = (p_f \star k)(n) \tag{6}$$

such that $\hat{f}$ becomes a maximum-likelihood estimate of f.

If we assume that p_f is locally ergodic, we can estimate $\hat{f}$ from a local image region, which corresponds to a local averaging of the channel weights within a spatial neighborhood. The algorithm consisting of the three steps channel encoding, channel averaging, and channel decoding is called channel smoothing and has been shown to be superior to many other robust smoothing methods [14].

Due to the assumption of (piecewise) constant distributions, the positioning of region boundaries might violate the sampling theorem, resulting in unstable edge-pixels. To avoid this effect, a modification to the channel decoding has been proposed in [23], called α-synthesis, which creates smooth transitions between neighborhoods with different feature levels. Instead of extracting the global maximum in (5), all local maxima are extracted located at channels n_r. The decoding is then obtained according to

$$\hat{f}_r = \frac{c_{n_r+1} - c_{n_r-1}}{M_{n_r}} + n_r \qquad \hat{f} = \frac{\sum_r \hat{f}_r M^\alpha_{n_r}}{\sum_l M^\alpha_{n_l}} \ . \tag{7}$$

For the choice of α see [23]; we used $\alpha = 2$ throughout this paper.

One major drawback of channel smoothing is the extensive use of memory if many feature channels are required. A high density of channels is only reasonable if the spatial support is large, which implies that the individual feature channels are heavily low-pass filtered along the spatial dimension. Therefore, the feature channels have a lower band limit and can be subsampled in the spatial domain without losing information. If the three steps of channel encoding, channel averaging, and subsampling are integrated into a single step, channel-coded feature maps (CCFMs) are generated. The advantage of CCFMs is a much higher number of channels, e.g. by combining several features as in Fig. 2, without increasing the memory requirements significantly.

The CCFM encoding of a single feature point can be written as (cf. (1)):

$$c_{l,m,n}(f(x,y),x,y) = k_f(f(x,y)-n)k_x(x-l)k_y(y-m) \ , \tag{8}$$

where k_f, k_x, k_y are the 1D kernels in feature domain and spatial domain. Note that x and y are scaled such that they suit the integer spatial channel centers l, m. Note further, that the previous definition of CCFMs assumes separable kernels, but we could easily use non-separable kernels, e.g. in the case of orientation data. Similar to (1), the encoding (8) of a set of feature points can be written as a scalar product in 3D function space or as a 3D correlation, where we use

$$\delta_f(x,y,z) = \delta(z - f(x,y)) \tag{9}$$

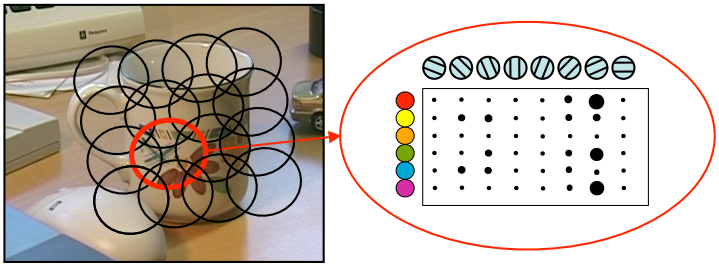

Fig. 2. Simultaneous encoding of orientation and color in a local image region. Figure taken from [17] courtesy Erik Jonsson.

and $k_{f,n}(z) = k_f(z-n)$, $k_{x,l}(x) = k_x(x-l)$, $k_{y,m}(y) = k_y(y-m)$:

$$\begin{aligned} c_{l,m,n}(f) &= \langle \delta_f | k_{f,n} k_{x,l} k_{y,m} \rangle = \iiint \delta_f(x,y,z) k_{f,n}(z) k_{x,l}(x) k_{y,m}(y)\, dz\, dy\, dx \\ &= (\delta_f \star (k_f k_x k_y))(n,m,l). \end{aligned} \tag{10}$$

The final formulation is the starting point of the CCFM scale-space.

3 The CCFM Scale-Space

In this section, we introduce the concept of CCFM scale-space. Our considerations are based on CCFMs computed from grey-scale images, i.e., we consider $f : \mathbb{R}^2 \to \mathbb{R}^+$ instead of a more general feature function.

3.1 Linear Scale-Space Theory in the Spatio-Featural Domain

The starting point is to embed the image $f(x,y)$ as a 3D surface according to (9). One might try to generate a 3D α scale-space [6] (Gaussian as a special case $\alpha = 1$ and all α-kernels are symmetric, i.e., correlation and convolution are the same):

$$F_s(x,y,z) = (k_s^{(\alpha)} \star \delta_f)(x,y,z) \tag{11}$$

However, the semi-group property of scale-space implies that all dimensions (spatial dimensions and the feature dimension) become increasingly blurred. Despite the fact that this implies a rapidly growing loss of information with increasing scale and a singular zero scale, this procedure is insensible from a statistical perspective and does not comply with the notion of scale selection [24, 25].

Since the latter argument is not straightforward, we explain our rationale in some more detail. From the requirement that the dimensionless derivative attains its maximum at a position proportional to the wavelength of the signal [24]

(section 13.1), we conclude that the scale of a structure is proportional to its spatial scale (a trivial fact) and anti-proportional to its feature scale. The latter can be shown by looking at the Taylor expansion of a harmonic oscillation $A\sin(\omega x)$ in the origin: $A\omega x$. The steepness of a sinusoid $A\omega$ in the origin grows linearly with the amplitude and the frequency, i.e., it is antiproportional to the wavelength $\lambda = \frac{2\pi}{\omega}$.

Alternatively, one can consider the energy of a harmonic oscillation. The energy is proportional to the square of the amplitude times the square of the frequency: $E \propto A^2\omega^2 \propto \frac{A^2}{\lambda^2}$. That means, if we apply a 3D lowpass filter to the spatio-featural domain, the energy decays with a power of four. Hence, scale selection would favor the highest possible frequencies in nearly all cases. If we scale the amplitude anti-proportionally to the spatial domain, the change of energy is balanced and will reflect intrinsic properties of the signal.

3.2 The Spatio-Featural Uncertainty: The Linear Case

In what follows, we analyze a linear 1D signal resulting from, e.g., the cross section of a locally planar image. Images are observations of a stochastic process, i.e., each measurement of the signal at each position follows a certain distribution in the feature domain. Furthermore, our measurements are subject to stochastic position errors and deterministic distortions (e.g. point-spread function), resulting in a distribution in the spatial domain. If we assume stationarity and independence of the two distributions, we can model the densities by a separable function that is shift-invariant in (f, x)-space, see Fig. 3, top left.

The overall distribution is obtained as the margin at angle ϕ, i.e. it is a function of $t = \cos\phi\, f - \sin\phi\, x$ obtained by integrating along $s = \cos\phi\, x + \sin\phi\, f$, see Fig. 3, top right. For the case of Gaussian distributions, this integral can be computed analytically:

$$p(t;\phi) \propto \int \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{f^2}{2\sigma_f^2}\right) ds \tag{12}$$

$$= \int \exp\left(-\frac{(s\cos\phi - t\sin\phi)^2}{2\sigma_x^2} - \frac{(s\sin\phi + t\cos\phi)^2}{2\sigma_f^2}\right) ds \tag{13}$$

$$\propto \exp\left(-\frac{t^2}{2(\sin^2\phi\,\sigma_x^2 + \cos^2\phi\,\sigma_f^2)}\right) \tag{14}$$

where σ_f^2 is the variance of the distribution in the feature domain and σ_x^2 is the spatial distribution.

In order to compute suitable kernels for a scale-space representation, the 2D distribution $p(t;\phi)$ is projected onto the spatial domain respectively the feature domain, see Fig. 3, bottom. For the case of Gaussian distributions, the

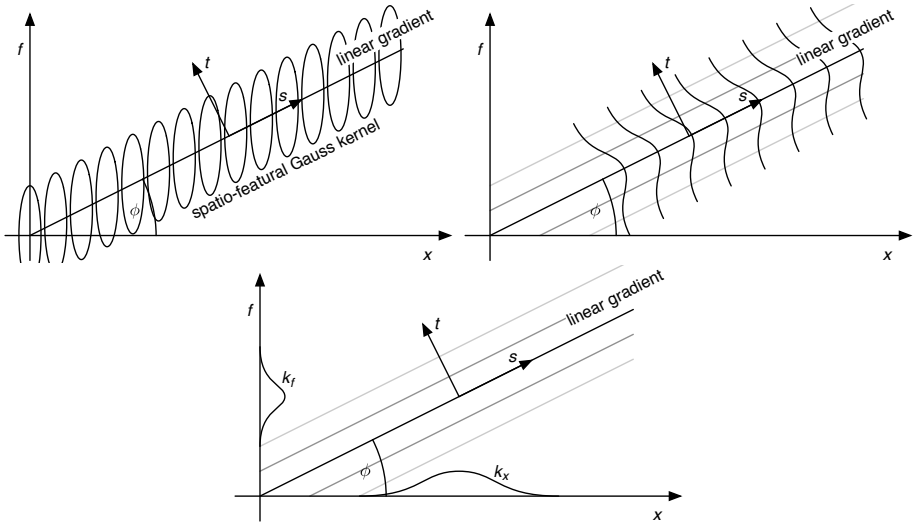

Fig. 3. Illustration for the derivation of uncertainties of a linear gradient. Top left: spatio-featural distribution moving along a linear signal. Top right: 2D distribution obtained by marginalizing along s. Bottom: projections of the 2D distribution onto the original spatio-featural coordinates.

projections can be computed analytically again:

$$p_x(x;\phi) = p(t;\phi)\big|_{f=0} \propto \exp\left(-\frac{(f\cos\phi - x\sin\phi)^2}{2(\sin^2\phi\,\sigma_x^2 + \cos^2\phi\,\sigma_f^2)}\right)\Bigg|_{f=0}$$

$$= \exp\left(-\frac{x^2}{2(\sigma_x^2 + \cot^2\phi\,\sigma_f^2)}\right) \qquad \text{and} \tag{15}$$

$$p_f(f;\phi) \propto \exp\left(-\frac{f^2}{2(\tan^2\phi\,\sigma_x^2 + \sigma_f^2)}\right) \ , \tag{16}$$

resulting in the variances

$$\sigma_{k_f}^2(\phi) = \sigma_f^2 + \tan^2\phi\,\sigma_x^2 \tag{17}$$

$$\sigma_{k_x}^2(\phi) = \sigma_x^2 + \cot^2\phi\,\sigma_f^2 \ . \tag{18}$$

Hence, we obtain the spatial uncertainty $(\Delta x)^2 = \frac{1}{2}\sigma_{k_x}^2(\phi)$ and the feature uncertainty $(\Delta f)^2 = \frac{1}{2}\sigma_{k_f}^2(\phi)$. Minimizing the product of uncertainties with respect to ϕ

$$\phi_0 = \arg\min_\phi (\Delta x)^2(\Delta f)^2 = \arg\min_\phi \frac{1}{4}\sigma_{k_x}^2(\phi)\sigma_{k_f}^2(\phi) \tag{19}$$

results in a global minimum at $\phi_0 = \tan^{-1}\frac{\sigma_f}{\sigma_x}$ giving $\sigma_{k_x}^2(\phi_0)\sigma_{k_f}^2(\phi_0) = 4\sigma_f^2\sigma_x^2$.

3.3 The Spatio-Featural Uncertainty Relation

In order to generalize the result from the previous section, we have to define the group structure of spatio-featural transformations. Readers not familiar with group theory might consider the previous example as a proof of concept and continue with the subsequent section.

We choose a methodology which is based on the isotropic model used in [26], although restricted to the 1D case. The higher-dimensional case generalizes straightforwardly. The group that we consider contains the shearing group and the translation group given as

$$x' = x + t_x \tag{20}$$

$$f' = f + \tan(\phi)x + t_f \ . \tag{21}$$

The shearing transformation corresponds to the rotation of a Euclidean space and is obtained since the f-coordinate is a null-vector [26], i.e., $f \cdot f = 0$. The parameterization is chosen such that it is consistent with the angle ϕ in the previous section and it reflects the fact that points move along the surface / curve with angle ϕ, i.e., that we cannot determine whether measurement noise comes from spatial uncertainty or feature noise. Using this definition we state the following

Theorem 1. *Let the spatio-featural domain be described by the isotropic model. The uncertainty product in the spatio-featural domain has a lower bound*

$$\exists k > 0 : \qquad (\Delta x)(\Delta f) \geq k \tag{22}$$

and the lower bound is given as

$$k = \frac{1}{2}\sigma_f \sigma_x \tag{23}$$

where σ_f^2 is the variance of the feature domain marginal distribution and σ_x^2 is the variance of the spatial domain distribution.

The proof of this theorem is given as follows. The generators of (20) and (21) are given as

$$s_x = \partial_x \qquad o_f = x\partial_f \qquad s_f = \partial_f \tag{24}$$

and the commutator of s_x and o_f is given by

$$[s_x, o_f] = s_x o_f - o_f s_x = \partial_x x \partial_f - x\partial_f \partial_x = \partial_f = s_f \ . \tag{25}$$

Hence, using the Robertson-Schrödinger relation [27] (note that the considered shearing transformations is a spinor group in the considered space), we obtain

$$E\{s_x^2\}E\{o_f^2\} \geq \frac{1}{4}E\{[s_x, o_f]\}^2 = \frac{1}{4} \ . \tag{26}$$

Taking the square root and scaling x and f by σ_x respectively σ_f, we obtain Theorem 1.

Note that (23) and the example of the previous section differ by a factor of 2. This might either mean that other types of noise distribution would lead to smaller uncertainties or that it is not possible to reach the lower bound. Despite the actual uncertainty product, Theorem 1 implies that we should not use iterative filtering with a 3D low-pass kernel as in (11). Instead, the scales in the different domains must behave reciprocal. The optimal choice of scales is the topic of the subsequent section.

3.4 Scale-Selection for CCFMs

The derivation from the previous section can be used to determine the proper change of scale when constructing a CCFM scale-space. The major trouble is, however, that we normally do not have access to the effective σ_f^2 and σ_x^2, i.e., we have to find a way to estimate these unknown parameters.

When considering the 3D embedding (9) in context of channel representations, we make the following observation. Encoding $f(x, y)$ with three arbitrarily, but finitely large channels always results in the same image after decoding (assuming infinite accuracy of real numbers):

$$(c_1(f/R) - c_{-1}(f/R))R = f \qquad 0 << R < \infty \ . \tag{27}$$

This identity is obtained directly from (1) and (5) for the case of three channels (implying $n_0 = 0$ and $M_{n_0} = 1$) or directly by (4). This means, we can always select the relative feature resolution R such that the feature resolution $\sigma_{k_f}^{-2}(\phi)$ is minimal. Furthermore, we know the original spatial resolution, which can be considered as a maximal spatial resolution [28]. These extremal resolutions can be considered as initial conditions for the CCFM scale-space.

Consider for example an image of the size $X \times X$ with values in $[-0.5, 0.5]$. We simply select $R = 1$ and obtain

$$1 = \sigma_{k_f}^2(\phi_{\min}) = \sigma_f^2 + \tan^2 \phi_{\min} \, \sigma_x^2 \ . \tag{28}$$

From the image size, we know that

$$\frac{1}{X^2} = \sigma_{k_x}^2(\phi_{\min}) = \sigma_x^2 + \cot^2 \phi_{\min} \, \sigma_f^2 \ . \tag{29}$$

This means, we get two equations and three unknowns σ_x^2, σ_f^2, and $\phi_{\min}$ and thus we cannot compute the proper scale-selection scheme from the initial image only.

However, we also have knowledge about the final state of the scale-space: There is an angle $\phi_{\max}$ for which the spatial resolution $\sigma_{k_x}^{-2}(\phi)$ becomes minimal, i.e., equal to one, which corresponds to three channels (the minimum number for quadratic B-spline channels):

$$1 = \sigma_{k_x}^2(\phi_{\max}) = \sigma_x^2 + \cot^2 \phi_{\max} \, \sigma_f^2 \ . \tag{30}$$

However, this introduces another unknown $\phi_{\max}$. This can finally be eliminated by selecting a maximal resolution in feature space. The choice of the latter is

however subject to heuristics, as there is no natural upper bound to feature resolution. As an alternative, it can be chosen according to requirements in the further processing. In any case, by selecting some constant F (we used $F = 20$), we obtain

$$\frac{1}{F^2} = \sigma_{k_f}^2(\phi_{\max}) = \sigma_f^2 + \tan^2 \phi_{\max}\, \sigma_x^2 \ . \tag{31}$$

The four equations (28–31) are solved by

$$\sigma_f^2 = \frac{X^2 - 1}{F^2 X^2 - 1} \qquad \sigma_x^2 = \frac{F^2 - 1}{F^2 X^2 - 1} \tag{32}$$

$$\phi_{\min} = \tan^{-1} X \qquad \phi_{\max} = \cot^{-1} F \ . \tag{33}$$

Inserting these parameters into (17) and (18) determines good choices of scales in spatial-featural domain and has been used in subsequent experiments.

4 Experiments

In the experiments shown in Fig. 4, we have quantized the spatio-featural resolutions resulting from (17) and (18). The image has been encoded in a CCFM with the corresponding number of channels and has been decoded afterwards. For the decoding of pixels not lying at a spatial channel center, we have linearly interpolated between the nearest channels. Obviously, the processed images maintain similar perceptual quality for a wide range of channels, before the image degrades at a resolution of 32×32. Note that any resolution can be computed in a single step, since the CCFM scale-space is linear and need not be computed iteratively.

5 Conclusion

This paper presents a new theorem for a spatio-featural uncertainty relation. The theoretic result is directly applicable in terms of scale-selection for non-linear image filtering using CCFMs. The main claim of this paper is that one should always consider the spatial domain and the feature domain in conjunction, since they are inherently connected. Still, the presented results are only a very first step and need to be considered more in detail for various applications and the theoretic results need to be generalized for non-flat manifolds (non-trivial fibre-bundles) and effects of higher dimensionality need to be considered. Classical scale-space features as preservation of the average grey-scale or the max-min principle have to be considered in future work, presumably in some modified formulation.

Acknowledgement

The authors would like to thank P.-E. Forssen for various discussions about the paper, in particular on alpha-synthesis.

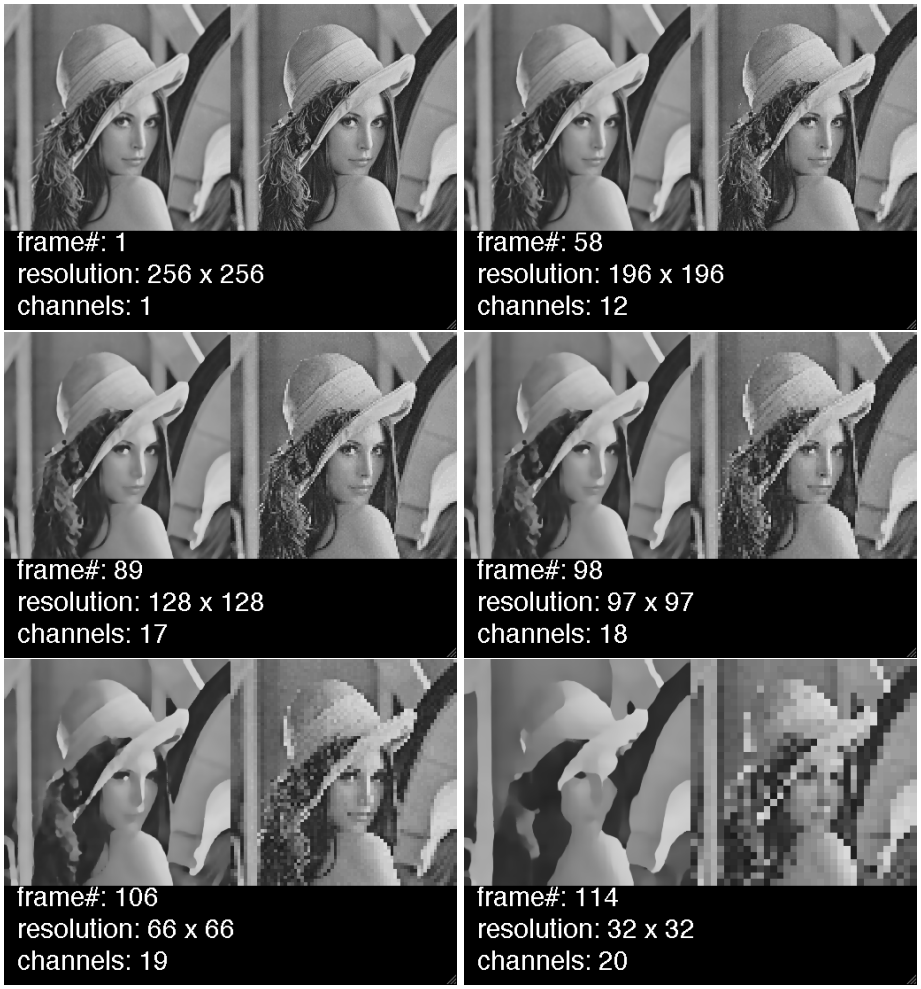

Fig. 4. Examples for CCFM-smoothing at different scales. The spatial and featural (denoted as *channels*) resolutions are given as (quantized) functions (18) and (17) where ϕ is linearly increasing with the frame number. The feature considered here is the grey-scale.

References

1. Iijima, T.: Basic theory of pattern observation. In: Papers of Technical Group on Automata and Automatic Control, IECE, Japan. (December 1959)
2. Witkin, A.P.: Scale-space filtering. In: Int. Joint Conf. Art. Intell. (1983) 1019–1022
3. Koenderink, J.J.: The structure of images. Biolog. Cybernetics **50** (1984) 363–370
4. Weickert, J., Ishikawa, S., Imiya, A.: Linear scale-space has first been proposed in Japan. Mathematical Imaging and Vision **10** (1999) 237–252
5. Felsberg, M., Sommer, G.: The monogenic scale-space: A unifying approach to phase-based image processing in scale-space. J. Math. Imag. Vis. **21** (2004) 5–26

6. Duits, R., Florack, L.M.J., de Graaf, J., ter Haar Romeny, B.M.: On the axioms of scale space theory. Journal of Mathematical Imaging and Vision **20** (2004) 267–298
7. Granlund, G.H.: In search of a general picture processing operator. Computer Graphics and Image Processing **8** (1978) 155–173
8. Burt, P.J., Adelson, E.H.: The Laplacian pyramid as a compact image code. IEEE Trans. Communications **31**(4) (1983) 532–540
9. Mallat, S.G.: A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans. Pattern Anal. Machine Intelligence **11** (1989) 674–693
10. Daubechies, I.: Orthonormal bases of compactly supported wavelets. Communications on Pure and Applied Mathematics **41**(7) (1988) 909–996
11. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. IEEE Trans. Pattern Analysis and Machine Intelligence **12**(7) (1990) 629–639
12. Weickert, J.: Theoretical foundations of anisotropic diffusion in image processing. Computing, Suppl. **11** (1996) 221–236
13. Portilla, J., Strela, V., Wainwright, J., Simoncelli, E.P.: Image denoising using scale mixtures of Gaussians in the wavelet domain. IEEE Trans. Image Processing **12**(11) (2003) 1338–1351
14. Felsberg, M., Forssén, P.E., Scharr, H.: Channel smoothing: Efficient robust smoothing of low-level signal features. IEEE Transactions on Pattern Analysis and Machine Intelligence **28**(2) (2006) 209–222
15. Granlund, G.H.: An associative perception-action structure using a localized space variant information representation. In: Algebraic Frames for the Perception-Action Cycle. LNCS 1888 (2000) 48–68
16. Forssén, P.E., Granlund, G.: Robust multi-scale extraction of blob features. In: Scandinavian Conference on Image Analysis. LNCS 2749 (2003) 11–18
17. Jonsson, E.: Channel-Coded Feature Maps for Computer Vision and Machine Learning. PhD thesis, Linköping University, Sweden (2008)
18. Jonsson, E., Felsberg, M.: Efficient computation of channel-coded feature maps through piecewise polynomials. Image and Vision Computing (in press)
19. Felsberg, M., Granlund, G.: P-channels: Robust multivariate m-estimation of large datasets. In: International Conference on Pattern Recognition, Hong Kong (2006)
20. Zemel, R.S., Dayan, P., Pouget, A.: Probabilistic interpretation of population codes. Neural Computation **10**(2) (1998) 403–430
21. Pouget, A., Dayan, P., Zemel, R.: Information processing with population codes. Nature Reviews – Neuroscience **1** (2000) 125–132
22. Howard, I.P., Rogers, B.J.: Binocular Vision and Stereopsis. Oxford University Press, Oxford, UK (1995)
23. Forssén, P.E.: Low and Medium Level Vision using Channel Representations. PhD thesis, Linköping University, Sweden (2004)
24. Lindeberg, T.: Scale-Space Theory in Computer Vision. Kluwer Academic Publishers, Boston (1994)
25. Elder, J.H., Zucker, S.W.: Local scale control for edge detection and blur estimation. IEEE Trans. Pattern Analysis and Machine Intell. **20**(7) (1998) 699–716
26. Koenderink, J.J., van Doorn, A.J.: Image processing done right. In: Proceedings European Conference on Computer Vision. (2002) 158–172
27. Santhanam, T.S.: Higher-order uncertainty relations. Journal of Physics A: Mathematical and General **33**(8) (2000) 83–85
28. Florack, L.M.J., ter Haar Romeny, B.M., Koenderink, J.J., Viergever, M.A.: Scale and the differential structure of images. Image Vision Comp. **10**(6) (1992) 376–388