ON THE INVARIANCE OF DIFFERENTIAL SCALE INVARIANTS

Andreas Siebert

Department of Computer Science
The University of British Columbia
siebert@cs.ubc.ca

ABSTRACT

Many geometric invariants have been reported in the literature. For two specific differential scale invariants we discuss the parameters involved and analyze the accuracy of their computation on real camera data. Our experimental data show significant errors in the observed values. Those errors seem to be inherent in the image formation and scaling process. They indicate the limitations of local differential computations and, more generally, the problems of synthetically scaled images as models for zoomed images.

1. INTRODUCTION

In an earlier work [5], we have proposed differential invariants under similarity transformations, plus linear brightness change. While invariance under translation, rotation, and brightness change is straightforward to achieve, invariance under scaling, where scaling refers to the zooming process of a camera, poses some fundamental problems for sampled images.

- The standard sampling theory assumes *point* sampling. However, a camera is an *area* based sampler, integrating over the necessarily finite size of the sensor elements [1].
- It is generally assumed that the signal is bandlimited and sampled above the Nyquist rate, but with real world scenes, this usually doesn't hold.
- The concept of derivatives is ill-defined on a matrix of sample points. We have to assume an underlying continuous image function. However, there are infinitely many ways to fit a continuous function to given data points. If Gaussian kernels are used to compute the derivatives, the value of the differential invariant depends on the σ of the Gaussian (or, equivalently, the window size) chosen for the computation.

On top of this, we have to consider the usual error sources that are part of the image formation process, like noise and lense distortions, or even JPEG compression. Given the long list of potential error sources, the question is whether local scale invariants can be computed robustly.

2. THE INVARIANTS

The two differential scale invariants proposed in [5], based on the idea to form ratios of derivatives such that the transformation parameters cancel out, are defined for a 1-d function $f(x)$ as follows.

$$\Theta_{m12}(x) = \begin{cases} 0 & \text{if } f'^2 = 0 \ \wedge \ f'' = 0 \\ f'(x)^2/f''(x) & \text{if } |f'^2| < |f''| \\ f''(x)/f'(x)^2 & \text{else} \end{cases} \quad (1)$$

$$\Theta_{m123}(x) = \begin{cases} 0 & \text{if c1} \\ (f'(x)f'''(x))/f''(x)^2 & \text{if c2} \\ f''(x)^2/(f'(x)f'''(x)) & \text{else} \end{cases} \quad (2)$$

where c1 is the condition $f''(x) = 0 \wedge f'(x)f'''(x) = 0$, and c2 specifies $|f'(x)f'''(x)| < |f''(x)^2|$. Note that -1$\leq \Theta_{m12} \leq$1, and 0$\leq \Theta_{m123} \leq$1.

The derivative operators are implemented with Gaussian kernels [3, 5, 7]. They are computed over a finite window size, and, critically, the σ of the Gaussian is a parameter of the computation. In 2-d, the derivatives in eqs. (1) and (2) are replaced by rotationally symmetric derivative operators. That is, the first derivative operator becomes the gradient magnitude, the second derivative operator becomes the Laplacian, and the third derivative operator becomes the cubic variation. Note that Θ_{m12} is actually invariant under similarity transformations (translation, rotation, scale), while Θ_{m123} is also invariant under linear brightness changes. It is assumed that the scene object of interest is always in focus, i.e. there are no significant changes in depth along the object of interest.

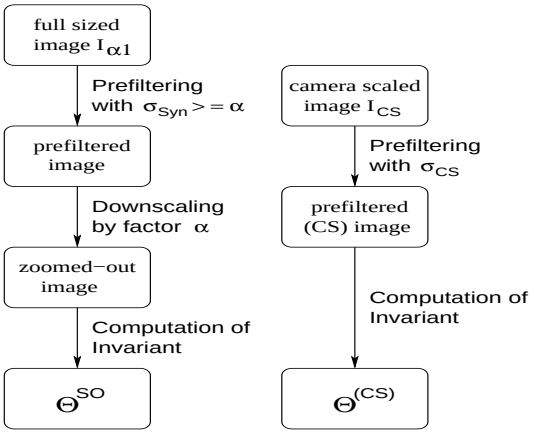

Figure 1: Computation of scale invariants on images of different size.

3. THE SCALING PROCESSES

In order to analyze the accuracy and variability of the invariants, we compare the invariants computed at each pixel of an image $I_{\alpha 1}$ to the ones of another image I_{CS} of the same scene, but reduced in size. That is, we take an image of an object, called the *alpha-1* image, move the camera away from the object, and take another image, called the *camera scaled* (CS) image. In the CS image, the object of interest is mapped onto fewer pixels, and the decrease in size defines the scaling factor α.

For the purpose of comparison, we downsample $I_{\alpha 1}$ to the same size as I_{CS}, i.e. by a factor α. This downsampling is done by Gaussian prefiltering with parameter σ_{Syn} and subsequent resampling using cubic spline interpolation. We assume a Gaussian model for the camera scaling process, i.e. the camera's point spread function is assumed to have approximately the shape of the Gaussian.[1]

The two processes to arrive at the invariants are depicted in fig. 1. The simulated process, called scaling by simulated optical zooming (SO), is shown on the left, producing Θ^{SO}, and the hardware based camera scaling process on the right producing $\Theta^{(CS)}$, where Θ is either Θ_{m12} or Θ_{m123}.

The SO process downsamples the alpha-1 image to the same size as the CS image. So we can expect identical invariants Θ^{SO} and $\Theta^{(CS)}$ if the underlying images, before the computation of the invariants, are identical. Therefore, we first compute the difference between the CS image and the downsampled alpha-1 image and minimize it by adequate prefiltering.

Let $I_\alpha(\sigma_{Syn})$ be the synthetically downsampled image prefiltered with σ_{Syn}, and $I_{CS}(\sigma_{CS})$ the camera scaled image prefiltered with σ_{CS}. Then the absolute difference Δ_G at each pixel location is

$$\Delta_G(x, y; \sigma_{CS}, \sigma_{Syn}) = |I_{CS}(x, y; \sigma_{CS}) - I_\alpha(x, y; \sigma_{Syn})| \tag{3}$$

the maximum absolute difference is

$$\widehat{\Delta}_G(\sigma_{CS}, \sigma_{Syn}) = \max_{x,y} \Delta_G(x, y; \sigma_{CS}, \sigma_{Syn}) \tag{4}$$

and if we regard the absolute difference as noise, we can define the signal to noise ratio

$$\mathrm{SNR_G}(\sigma_{\mathrm{CS}}, \sigma_{\mathrm{Syn}}) = 10 \log_{10}(\mathrm{V_{CS}} \,/\, \mathrm{V_{\Delta_G}}) \tag{5}$$

where V_{CS} is the variance of $I_{CS}(x, y; \sigma_{CS})$ and V_{Δ_G} is the variance of $\Delta_G(x, y; \sigma_{CS}, \sigma_{Syn})$.

Let's assume that σ_{Syn} has been set to a certain value, preferably $\sigma_{Syn} \geq \alpha$ for anti-aliasing purposes. For the camera scaled image, the corresponding value of the prefiltering parameter σ_{CS} depends on both σ_{Syn} and α. Since the cascade of two Gaussians of size σ_1 and σ_2, respectively, is equivalent to one Gaussian of size $\sqrt{\sigma_1^2 + \sigma_2^2}$, we have $\sigma_{Syn}^2 = \alpha^2 + (\alpha\,\sigma_{CS})^2$ so that the theoretical size of the prefilter for the zoomed image is

$$\widehat{\sigma}_{CS} = \frac{1}{\alpha}\sqrt{\sigma_{Syn}^2 - \alpha^2} \tag{6}$$

σ_{CS}	σ_{Syn}	$\widehat{\Delta}_G$	$\mathrm{SNR_G}$
0	0	127.46	11.443
0	1.35	113.80	10.546
1.42	2.35	27.96	20.410
1.72	2.35	21.24	22.191
2.27	3.35	16.60	22.470
2.57	3.35	14.76	22.857

Table 1: Error measures for I_{CS} vs. I_α for different prefilter settings.

For the example image `Building` in fig. 2, table 1 shows the error measures computed according to eqs. (4) and (5) for different settings of σ_{CS} and σ_{Syn}. If σ_{Syn}=0, aliasing occurs, and the maximum absolute difference $\widehat{\Delta}_G$ is large. If σ_{Syn}=α=1.35, but σ_{CS}=0, $\widehat{\Delta}_G$ is still large because now $I_\alpha(\sigma_{Syn} = 1.35)$ is a smoothed image compared to the unfiltered image $I_{CS}(\sigma_{CS} = 0)$.

[1] Our camera, a Sony DXC 950, shows a certain contrast enhancing effect in response to a sharp edge, but the PSF is still close to the Gaussian.

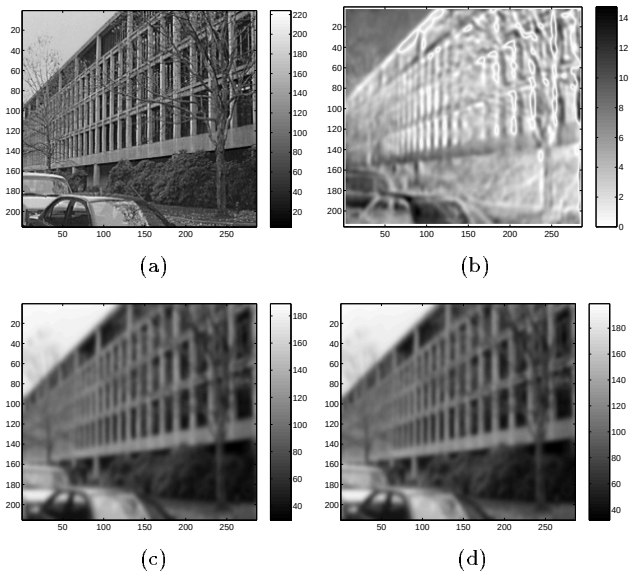

Figure 2: Minimizing I_{CS} vs. I_α, α=1.35. (a) $I_{CS}(\sigma_{CS}=0)$; (b) $\Delta_G(\sigma_{CS}=2.57, \sigma_{Syn}=3.35)$; (c) $I_{CS}(\sigma_{CS}=2.57)$; (d) $I_\alpha(\sigma_{Syn}=3.35)$.

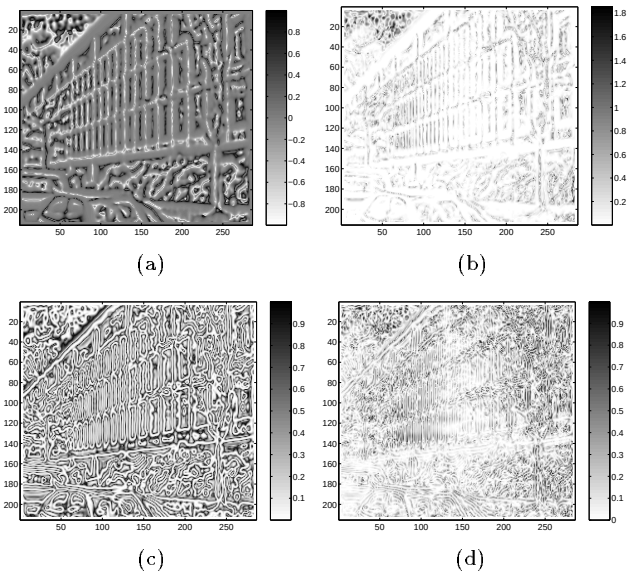

Figure 3: $\Theta^{(CS)}$ vs. Θ^{SO}, α=1.35. (a) $\Theta^{(CS)}_{m12}(\sigma_{CS}=2.28)$; (b) $\Delta_{\Theta_{m12}}$; (c) $\Theta^{(CS)}_{m123}(\sigma_{CS}=2.28)$; (d) $\Delta_{\Theta_{m12}}$.

Therefore, both I_α and I_{CS} have to be smoothed. If we set σ_{Syn}=α+1.0=2.35 and σ_{CS}=1.42=$\widehat{\sigma}_{CS}$ according to eq. (6), then the SNR increases significantly while $\widehat{\Delta}_G$ decreases sharply. Slight improvements in the SNR can be achieved by filtering with a larger σ_{Syn}, but also by filtering with a $\sigma_{CS} > \widehat{\sigma}_{CS}$. For our camera and our image data, the empirical data suggests $\sigma_{CS} = \widehat{\sigma}_{CS} + 0.3$ for maximizing the SNR. The general insight is that the minimization of the difference between scaled images not only depends on the scaling factor but also on the interdependent prefiltering parameters for both images.

We proceed with the invariants Θ_{m12} and Θ_{m123} for I_{CS} and I_α. We compute the same error measures as in eqs. (3)-(5), where now the intensity values have been replaced by the invariant values. The results in table 2 show that the maximum absolute differences are large, and the SNR has dropped significantly, especially for Θ_{m123} which depends on third order derivatives. Fig. 3 shows Θ_{m12} and Θ_{m123} computed for I_{CS}, and their respective differences when computed for I_α. Not surprisingly, image areas of constant intensity pose a problem. But for Θ_{m123}, the data is so noisy almost everywhere that the computation of the invariant becomes meaningless.

One reason for the low SNR has to do with the finite area of physical sensors. They are aperture samplers which integrate over a certain area of the imaged object. Both the object area that is mapped onto a specific pixel and the relative phase between object and sensor elements change as the scale changes. These changes can produce significant imaging artifacts, as demonstrated in fig. 4. The image taken of a chessboard pattern, at a certain distance, shows gray lines that are not part of the chessboard pattern. The emergence of these gray lines can be simulated by integrating over the chessboard with the appropriate ground area, as indicated by the overlayed dashed grid where the length of the integration area is seven eighths of the length of a black or white square. The resulting pattern is shown on the right. Clearly, such artifacts that may appear and disappear any time as the distance between camera and objects changes limit the robustness of any local computation.

α	$\widehat{\Delta}_{\Theta_{m12}}$	$\mathrm{SNR}_{\Theta_{m12}}$	$\widehat{\Delta}_{\Theta_{m123}}$	$\mathrm{SNR}_{\Theta_{m123}}$
1.09	1.961	6.444	0.995	5.288
1.18	1.860	6.837	0.992	6.016
1.27	1.936	4.837	0.998	3.619
1.35	1.854	5.121	0.998	4.111

Table 2: Error measures for $\Theta^{(CS)}$ vs. Θ^{SO} for different values of α, image `Building`, for Θ_{m12} and Θ_{m123}. Prefiltering with σ_{Syn}=3.0, $\sigma_{CS} = \widehat{\sigma}_{CS}$+0.3.

A similar problem exists even in 2-d Gaussian scale space, as pointed out in [2] and illustrated in fig. 5. Although the Gaussian has been shown [7] to have the

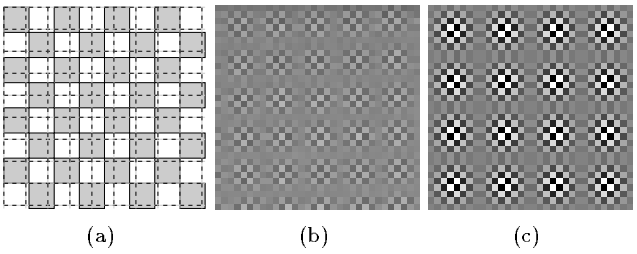

Figure 4: Area based sampling of a chessboard pattern. (a) synthetic chessboard pattern with overlayed grid; (b) image of chessboard pattern taken by a camera; (c) simulated aperture scanning.

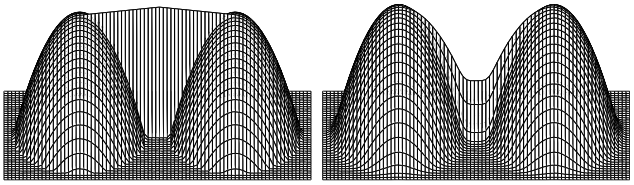

Figure 5: Creation of new extrema in 2-d: (left) before (right) after Gaussian smoothing.

causality property of not adding new intensity levels when filtering an image with increasing σ, it is still possible that new extrema emerge. Fig. 5 shows two cones, connected by a narrow ridge, with a maximum in the center. Smoothing removes that maximum and creates two new maxima at the peaks of the cones. The previous global maximum has been turned into a saddle point. Again, such events distort the derivatives.

Problems with the robustness of the computation of local features when the image undergoes a scale change have been reported in a related context. In [4], Schmid et al. try to identify *key points*. They apply various key point detectors to an image, record the key points, change the scale of the image by changing the focal length of the camera, and apply the same key point detectors again. Then they measure the so-called repeatability rate, i.e. the percentage of keypoints found in both images within a distance of ϵ from each other. They report repeatability rates of less than 5% at α=2.0 for the best key detectors for ϵ=0.5. When ϵ=1.5, the repeatability rates are higher, but they drop quickly to below 40% at α=2.0. This again indicates that the robust computation of local features under scale change is challenging.

4. CONCLUSION

We have compared images at different scales, both in terms of their intensities and in terms of differential scale invariants computed on them. The scaling process not only involves the scaling factor α, but also prefiltering of *both* images involved. For a Gaussian model, we showed how the prefiltering parameters are related.

Even with an optimized parameter setting, the differences between a camera scaled image and an image of same size downsampled from an image of larger size remains remarkably large. Taking ratios of derivatives aggravates the problem. The poor results question the usefulness of differential invariants under scaling.

Some reasons for these negative results have been pointed out. Most notable, area based sampling causing unwanted artifacts is inherent in the nature of the imaging process and plays a major role as the scale changes.

However, it should not be concluded that differential invariants are generally of limited usefulness. For example, in [6] we have shown how differential invariants can be employed to increase retrieval performance if the transformation of interest is gamma correction.

5. ACKNOWLEDGEMENTS

The author would like to thank Bob Woodham for his support and feedback.

6. REFERENCES

[1] G. Holst, "Sampling, Aliasing, and Data Fidelity", JCD Publishing & SPIE Press, 1998.

[2] L. Lifshitz, S. Pizer, "A Multiresolution Hierarchical Approach to Image Segmentation Based on Intensity Extrema", *IEEE PAMI*, Vol.12, No.6, pp.529-540, June 1990.

[3] C. Schmid, R. Mohr, "Local grayvalue invariants for Image Retrieval", *IEEE PAMI*, Vol.19, No.5, pp.530-535, May 1997.

[4] C. Schmid, R. Mohr, C. Bauckhage, "Comparing and Evaluating Interest Points", *ICCV*, 1998.

[5] A. Siebert, "A Differential Invariant for Zooming", *ICIP*, Kobe 1999.

[6] A. Siebert, "Differential Invariants under Gamma Correction", *Vision Interface*, Montreal 2000.

[7] B. ter Haar Romeny (ed.), "Geometry-Driven Diffusion in Computer Vision", Kluwer 1994.