Exploiting Uncertainty Propagation in Gradient-based Image Registration

Kevin Köser and Reinhard Koch
Institute of Computer Science
Christian-Albrechts-University Kiel, Germany
{koeser,rk}@mip.informatik.uni-kiel.de

Abstract

Parametric, gradient-based image alignment is used nowadays in many applications such as object tracking, image registration or camera calibration. In such processes intensity differences between a template and a warped image are minimised based on Newton-like optimisation algorithms. It has been known for a long time that pre-filtering the images under inspection and the use of coarse-to-fine strategies can somehow increase the convergence radius, but a general derivation is missing. We present a generic framework which relates parameter uncertainty with positions in the image's scale space instead of heuristic isotropic smoothing to improve the convergence radius. Specifying parameter uncertainty is often more intuitive than selecting a good pyramid level and improves convergence particularly in settings where a single parameter's influence (e.g. a rotation angle) varies largely across a patch. We show that the classical application of image pyramids in displacement estimation embeds into the novel formulation and demonstrate our approach on refinement of robust feature correspondences and homography estimation.

1 Introduction

During the last years automatic image matching based on invariant or robust features such as SIFT[3] has shown tremendous progress: Multi-scale or affine covariant region detectors find interest regions (cf. to [8]) which, together with local orientation[3], define a local coordinate system. An approximation of the local image warp between a feature correspondence is then readily available from the relative parameters of the detected regions. However, the accuracy of relative scale, orientation, position or affine parameters from the detector is typically only very rough and the question arises whether the local image-to-image transformation can be optimised using standard methods [2, 9] based on the grey values and gradients of the feature regions to allow exploitation of these correspondences, e.g. for homography[12] or pose[11] estimation. If gradient-based image alignment is directly applied to the full-resolution images, e.g. based upon affine warps, one notices that some pixels of a patch can provide contradictory information. They disturb the optimisation, because they are outside the valid linear environment of the true correspondence (compare fig. 1) and such outliers have to be avoided. For instance, if the initial parameters have a rotational error of 10° (which is the orientation histogram quantisation proposed in SIFT[3]), this results in a position error of 3 pixels for the corners of

BMVC 2008 doi:10.5244/C.22.9

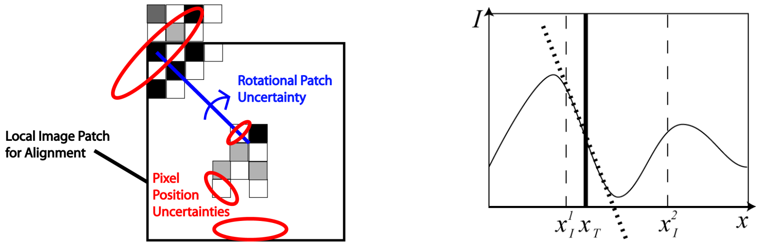
Rotational Patch Uncertainty
Local Image Patch for Alignment
Pixel Position Uncertainties
I
x_I^1
x_T
x_I^2
x

Figure 1: Given rotational, similarity or affine parameter uncertainty, the 2D pixel positions within a warped patch are unequally certain (left image). The ellipses indicate the 2D position uncertainties in a patch induced by significant rotational and minor scale uncertainty. Outer regions (far from the warp center) are typically more uncertain. For positions near a patch border the assumption of locally linear intensities in gradient based alignment quickly gets violated with such uncertainties. The linearity assumption can be seen in the 1D intensity profile of the right image, where the intensity near x_T is approximated by the tangent (dotted line). For a small displacement, e.g. x_I^1, the linearity is quite correct, but for a larger, e.g. x_I^2, the linear extrapolation is far away from the real data.

a square patch with half window size 12. In presence of fine detail, this violates the grey value linearisation assumption made in alignment. Typically, image pyramids are used in this case or heuristic smoothing is applied, but if we reduce the resolution, we may run into a high-dimensional version of the aperture problem, where we do not have enough data to estimate the parameters: The feature has a relatively good localization at its intrinsic scale in scale-space (cf. to [1]), but does not necessarily provide much structure at significantly coarser levels. Therefore, a goal of this contribution is to use as much of the data as possible and filter away only disturbing information. Also, when optimising parameterised warps based on multiple parameters, some may allow for a better prediction than others and some may have a stronger influence on the convex environment and validity of the local linearisation. Finally, not all of the pixels in a patch have to be sensitive to an incorrect start value in the same way (compare center and border in fig. 1). This contribution addresses all these issues and incorporates uncertainty in a unified way, thereby embedding the classical image pyramid from displacement estimation. Consequently, the novel approach will not improve convergence in simple displacement scenarios but its goal is to automatically exploit a heterogeneous structure of more complex warps better than constant isotropic smoothing. The paper is structured in the following way: Section 2 relates the contribution to previous work while section 3 presents the gradient-based alignment problem and shows how parameter uncertainty can be incorporated. The convergence is demonstrated on synthetic and on real images using patch-based local alignment and homography estimation in section 4.

2 Related Work

One of the first publications on gradient based image alignment is the work[2] by Lucas and Kanade in 1981. In a stereo setting they stated that under the *Image Brightness Constancy Assumption* the correspondence problem can be formulated as that of minimising the grey value difference using Newton's method, provided the prediction is close to the true value. They also state that smoothing the image can increase the convergence radius. Since then, a vast quantity of articles has been published on extensions, improvements, accelerations and applications of this topic. We refer the interested reader to the work of Baker and Matthews [9], which provides an excellent overview and comparison. An approach different from the low-parametric global model is often taken in optic flow estimation [10], where a 2D displacement is estimated for each pixel leading to a huge number of parameters, which are only important locally. Additional regularisation terms are applied to overcome the local aperture problem. In our contribution, we concentrate on the case of estimating the parameters of one model warp typically with high redundancy (a large number of intensity measurements but few global warp parameters), where we inspect the influence of the uncertainty of the global parameters.

Since the work of [2], alignment and tracking was performed on image pyramids in coarse-to-fine strategies, although this was handled rather as an implementation detail. For example, [7] mentions that parameters are propagated from one pyramid level to the next. Christmas [5] investigated the relation between smoothing and optical flow estimation in more detail, however he provided a specialised filter analysis for pure displacement only. Later, Molton et al. [6] examined parametric image warps in a probability-theoretic framework. However, they were focused on formalising and characterising all sources of noise and to incorporate priors on the warp parameters. Although they already give the intuition that "smoothing should be done over a range similar to the expected change of pixel position" they do not conclude that different pixels in a patch should be subject to different amounts of smoothing or that this smoothing could be anisotropic. Uncertainty was also handled in other works [13, 14], however, not incorporated into the minimisation but viewed as an outcome. To our knowledge, so far nobody considered the influence of parameter uncertainty within the grey value difference minimisation. Therefore, in contrast to previous work we propose to propagate parameter uncertainty to pixel position uncertainty, which helps in selecting a good filter scheme. We then give an implementation exploiting the image's scale space to obtain local convexity with high probability.

3 Parametric Image Alignment with Uncertainty

The image brightness constancy assumption states that corresponding points in two images have the same grey value, when the images are related by some warp W. According to Baker and Matthews [9] we refer to the first image as the template T and the second as the image I, where the warp depends on some parameters p:

$$I(W(x,p)) = T(x) \tag{1}$$

If a parameter prediction $\tilde{p}$ is given, which is sufficiently close to the true value $\check{p}$, we may use Newton's method to find the $\hat{p}$, which minimises the squared sum of intensity differences at position x in the patch P. We use the term *patch* here for intuition, in fact x may be from a set P of arbitrarily distributed sample points in an image. For example, for refinement of robust image features, we use a fixed grid attached to the local feature, such that the absolute number of samples does not depend on the size of the feature, or to obtain an infinite homography for a purely rotated camera, we may select a number of samples uniformly distributed across the image. Although our contribution is not restricted to a particular alignment method, we use what Baker and Matthews call the inverse compositional approach (see [9] for details), which exploits a prediction $\tilde{p}$ to obtain an inverse compositional update $\Delta\hat{p}$ in each iteration

$$\Delta\hat{p} = argmin_{\Delta p} \sum_x (T(W(x,\Delta p) - I(W(x,\tilde{p})))^2 \tag{2}$$

which is composed into $\tilde{p}$ for the next iteration. The equation system is based solely upon the gradients in T to estimate the missing transformation which is close to the identity transform. If $\tilde{p}$ is very close to $\check{p}$, this means that Δp is nearly zero, we are in the convex surrounding of the minimum of the error function and we can linearise the above sum

$$\sum_x (T(W(x,0)) + \nabla T \frac{\partial W}{\partial p} \Delta p - I(W(x,\tilde{p}))^2 \tag{3}$$

which (assuming $W(x,0) = x$) leads to the solution

$$\Delta p = H^{-1} \sum_x \left[\nabla T \frac{\partial W}{\partial p}\right]^T [I(W(x,\tilde{p})) - T(x)], \qquad H = \sum_x \left[\nabla T \frac{\partial W}{\partial p}\right]^T \left[\nabla T \frac{\partial W}{\partial p}\right] \tag{4}$$

The term in (3) is only a valid approximation of the term in (2) as long as $\tilde{p}$ is quite correct. It states that near the position x the template has the grey value $T(x) + \nabla T \frac{\partial W}{\partial p}$ which is only valid in a very small neighbourhood. E.g. if p parameterises translation and $\tilde{p}$ is 10 pixels away from the true optimum, in presence of fine detail there may be multiple local extrema in between, which are not represented by the linear approximation.

Incorporating Uncertainty

In previous algorithms the images had to be smoothed *sufficiently* to allow convergence, which is a rather unintuitive requirement. While we keep the idea, that the image brightness constancy assumption is also valid at coarser scales, we compute an appropriate scale now automatically on a per-sample basis *given an initial parameter uncertainty*: For robust feature refinement such uncertainty estimates can be obtained e.g. from an empirical feature detector evaluation [8] or from noise models [13]. We assume that the uncertainty of the parameter vector p is unimodal and characterised well by the first two moments of its distribution, mean $\tilde{p}$ and covariance Σ_{pp}. Since the normal distribution has the maximum entropy of all distributions for a given mean and covariance, we assume p being normal-distributed in the following. However, qualitatively the derivation also applies to the uniform distribution or other unimodal distributions. Now, let the warp W map coordinates of T to I. We now investigate how much the coordinates change, when we change the parameters p. Under the assumption that W is locally approximated well by its first order Taylor approximation linear error propagation yields:

$$\Sigma_{x_I x_I} \approx \frac{\partial W}{\partial p} \Sigma_{pp} \frac{\partial W}{\partial p}^T \tag{5}$$

If W is actually linear, then x_I is normal distributed with covariance $\Sigma_{x_I x_I}$. Now we select the iso-density curve at 2σ which comprises nearly 90% of probability inside and call this the *target region*. In the following it is assumed that almost always the true correspondence $\check{x}_I$ is somewhere in the target region and that we therefore require linear intensity within this region. The shape and the size depend on the projection of the parameter uncertainty Σ_{pp} into the image. We first consider the simple case that $\Sigma_{x_I x_I}$ has two equal eigenvalues. This means that x_I's distribution is an isotropic Gaussian with circular isodensity curves and that a point x_T is mapped to a disc around x_I, whose radius l is

$$l = 2\sqrt{0.5\, trace(\Sigma_{x_I x_I})} \tag{6}$$

Then, we select an appropriate scale in Gaussian *scale space* (cf. to [1]) such that structures of smaller size are suppressed to a large extent and our region can be considered approximately linear. This can be reached by convolution of the image with an isotropic Gaussian having standard deviation l. The grey value is then computed at this scale.

If on the other hand $\Sigma_{x_I x_I}$ has two different eigenvalues, this means that x_I's position is more uncertain in some direction. In this case imagine that we normalise the image size, such that in the transformed image the uncertainty becomes isotropic again. Then we can apply the method of above. These operations can efficiently be combined by smoothing the image with a Gaussian filter with covariance $4\Sigma_{x_I x_I}$. However, in both the isotropic and the anisotropic case we are interested only in a single grey value, so basically the image convolution boils down to a single weighted sum of intensities in the target region.

Consequently, we also have to compute the region in the template, where an image sample at x_I is backward-mapped given the parameter prediction and its distribution. Requiring the warp to be invertible is no restriction, since inverse compositional alignment assumes the warp to be invertible anyway.

$$\Sigma_{x_T x_T} \approx \frac{\partial W^{-1}}{\partial p} \Sigma_{pp} \frac{\partial W^{-1}}{\partial p}^T \tag{7}$$

This represents the region around x_T where the warp prediction maps an image position x_I into the template. Since linearity is desired within this region, we proceed in the same way as with the image. Now, we also calculate the gradient at the obtained scale.

To summarise, we propose that each grey value is obtained using an individual level of smoothing such that it is linear within the predicted parameter uncertainty. Since each pixel can be chosen from the best resolution available, it is less likely that one runs into the aperture problem, which happens often when the whole patch is lifted to a very coarse level, because then more information than necessary is suppressed. In case the warp uncertainty leads to an anisotropic position distribution anisotropic smoothing should be applied at this position, e.g. for small purely rotational uncertainty smoothing is only required tangential to the warp. The scale and the shape of the smoothing will in general vary from pixel to pixel.

In early works (e.g. [2]), where only 1D or 2D displacement was estimated, isotropic image smoothing or the use of image pyramids was suggested. This embeds perfectly into our framework, because in the case of pure displacement estimation, isotropic 2D parameter uncertainty leads to a constant and isotropic pixel position uncertainty ($\Sigma_{x_I} = \Sigma_{pp}$) for all positions in the patch. This results from the fact that the Jacobian of the warp with respect to the parameters (the displacement) does not depend on the pixel position. Therefore in our novel method all intensities would be picked from the same level in scale space or the same pyramid level, which is exactly what was proposed in earlier works. In the case of more complicated warps however, the more differentiated scheme of above is the consequent generalisation. As a side note: When prior knowledge about the

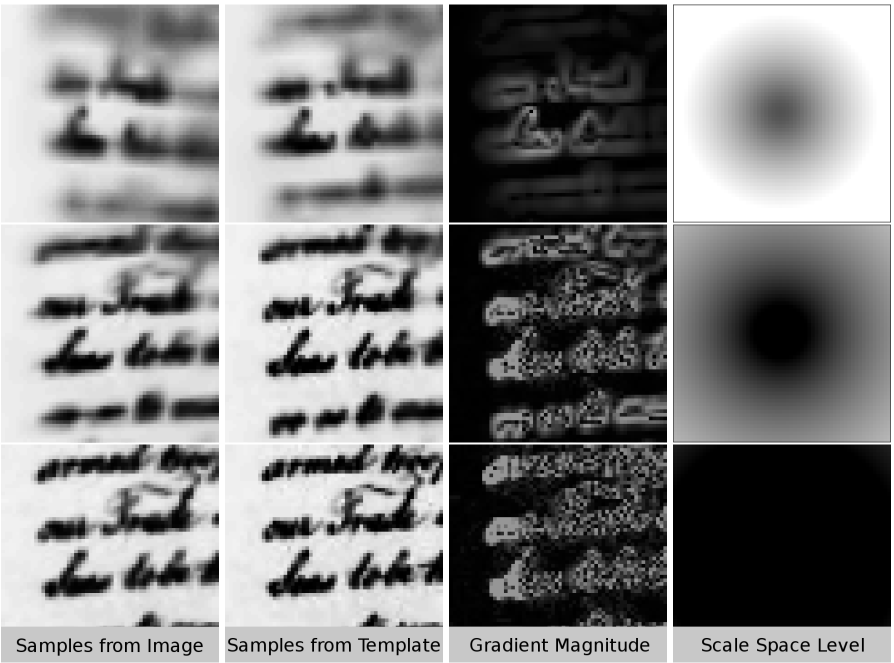

Figure 2: A patch containing hand-written text is aligned using the novel *scale* method. The initial affine warp to be estimated between the image in the first and the template in the second column contains a 15° rotation, a scale of 10% and a position offset of 1 pixel. The gradients used for estimation and the scale, where they were taken from can be seen in the right two columns (darker pixels represent lower values). Since initially the uncertainty is set appropriately for the missing transformation, particularly at the outer patch parts samples are picked from coarse resolutions (first row). With nearly compensated scale and rotation and improved uncertainty, finer details can be used in the second row. When all samples are taken from the highest resolution (third row) the algorithm behaves as the original inverse compositional alignment.

parameter distribution is available, it may also be of advantage to incorporate this in terms of priors in Bayesian estimation as proposed in [6]. To avoid mixing up different effects, in this contribution we focus on the intensity-related aspects when parameter uncertainty is available.

Algorithm and Implementation

We will now give some details of the implementation (see fig. 3 for an overview) and additionally, a second, more efficient approximation for the considerations presented in section 3. As an approximation for the image's scale space, we use the Gauss pyramid with width and height reduced by a factor of 2 per level (as e.g. also used in [3]). Between the pixels of a level and between the levels we interpolate linearly, which is also known as trilinear filtering in computer graphics (see [15]). If anisotropic smoothing is required,

```
Select set $P$ of measurement positions in template and repeat until all positions are taken
from the best resolution or some control points have reached the desired accuracy:
  1. For each $x_T \in P$ propagate parameter uncertainty $\Sigma_{pp}$ to position uncertainty $\Sigma_{x_T x_T}$
  2. Obtain template grey value and gradient (an)isotropically from template pyramid
     according to $\Sigma_{x_T x_T}$
  3. Construct Hessian and Steepest Descent Images (same as in [9])
  4. Repeat until no significant improvement:
       (a) For each $x_T \in P$ obtain image coordinates $x_I$ using $\tilde{p}$
       (b) propagate parameter uncertainty $\Sigma_{pp}$ to position uncertainty $\Sigma_{x_I x_I}$
       (c) obtain (an)isotropic grey values from image pyramid according to $\Sigma_{x_I x_I}$
       (d) Compute residuals, solve for $\Delta p$ and compose $\Delta p$ with $\tilde{p}$
  5. Update covariance $\Sigma_{pp}$
  6. If parameter update or covariance is sufficient break, otherwise go to 1
```

Figure 3: Overview of alignment with uncertainty

we first find the smaller principal vector of the pixel covariance and extract trilinear image values from the scale space, which must then be smoothed in direction of the larger principal vector. This exploits the pyramid and avoids anisotropic filtering with huge masks at full image resolution. We call this method simply *anisotropic* in the remainder.

Since often the parameter distribution is not known exactly but only its approximate shape, since additionally the linearisation of the warp is sometimes only valid in a small range and since anisotropic smoothing is expensive, we propose even in case of anisotropic covariance to simply pick the grey value directly from scale space according to eq. 6: As the trace is the sum of the eigenvalues and the eigenvalues of a covariance matrix are the variances in principal directions, the trace can be seen as a rough upper bound of the maximum variance. We call this approximation the *scale* method in the remainder. In case of isotropic pixel position uncertainty they are the same.

Based upon the gradients and the image intensities we perform the inverse compositional alignment. In the minimum of the error function, we estimate the parameter covariance from the Hessian and the reference variance. This new covariance is then used in the next iteration, for which the template and the image is constructed again as described above (compare fig. 2). Convergence of the system can be declared if all measurements (or some control measurements) are picked from the highest resolution. In this case the algorithm behaves as the original inverse compositional alignment.

4 Experiments

In order to demonstrate the principle of the novel approach, we first show a very simple example, where we create a 512×512 (floating point valued) test pattern image with intensity $I(x,y) = sin(\lambda\ x) + sin(\lambda\ y)$ as depicted in fig. 4. This image is rotated around its center and afterwards Gaussian noise ($\sigma_I = 2\%$ of the sine amplitude) is added to each

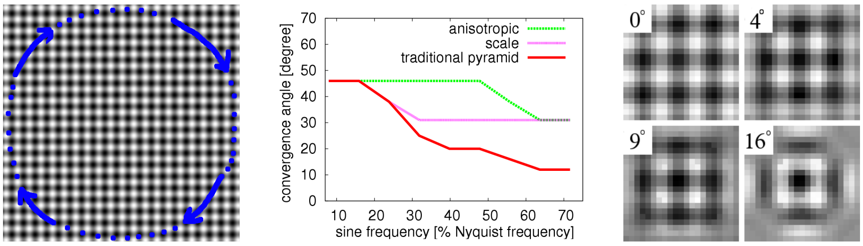

Figure 4: Top Row: The left image shows a sum of a horizontal and a vertical sine-pattern. Such images have been created for different frequencies and each was rotated around its center as indicated by the arrows. In the center plot, the maximum angle for which gradient-based Euclidian parameter estimation converged for a 21×21 center patch is depicted in dependence of the sine-frequency. Scale and anisotropic are the new methods of the previous sections which we compare with a traditional pyramid approach. Particularly when very fine image structures close to the Nyquist frequency are present, both novel approaches outperform the rigid pyramid with respect to the convergence radius. The template values of the center patch computed for rotational uncertainties of 0,4,9,16 degrees with the anisotropic method can be seen in the right image.

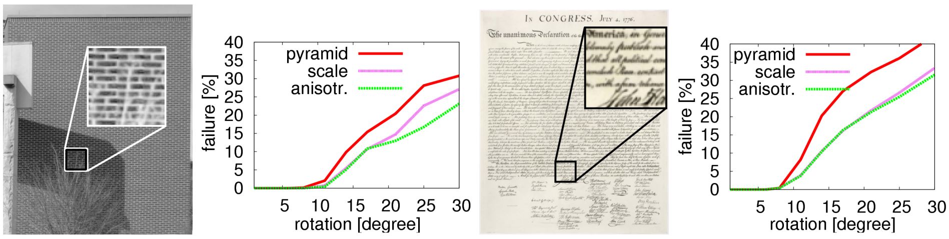

Figure 5: The images *bricks* and *declaration* have been rotated and the fraction of diverged alignments at SIFT-feature positions in the first image have been counted (right beneath each image). The start position in the second image was correct, but the rotation was set to zero to obtain a rotation convergence radius estimate. For the traditional pyramid method, the best pyramid layer is displayed, which provides still worse results than automatic individual smoothing. Note that these images contain very fine structures (see detail magnifications) which are almost filtered out in the classical coarse-to-fine strategy. In the novel approaches, they are used if possible (see also fig. 2).

pixel. We then compute a 3-parametric Euclidian warp (α, dx, dy) using the anisotropic and the scale method and a traditional pyramid-based approach for comparison, where the estimation is first performed on pyramid level 3 and then the results are down-propagated and refined on the next better resolution. We use 21×21 samples in a patch centered in the image and provide 0° as a rotation prediction. With increasing sine frequency the pyramid approach converges only for smaller and smaller angles, while the anisotropic filtering nearly always catches rotations of up to 45°. The scale approach has slightly worse convergence than the anisotropic but still better than the pyramid approach. Next,

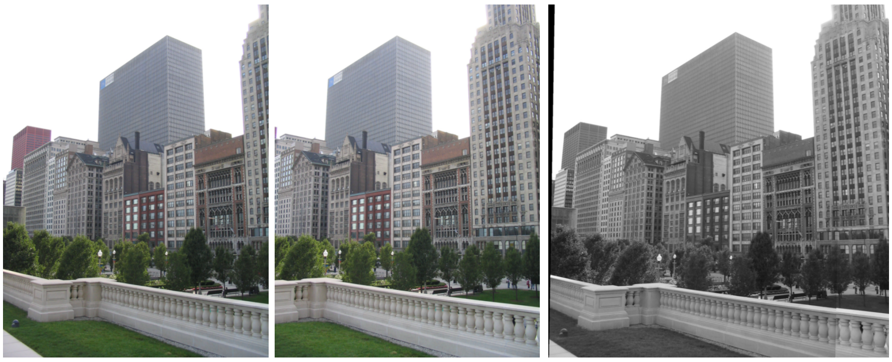

Figure 6: The two left images (1536×2048, ≈ 40° field of view) have been taken with a digital camera which purely rotated. An estimate of the rotation was given within 1° accuracy, from which an infinite homography could be predicted. Given prediction and uncertainty the homography has been optimised and the resulting parameters have been used to stitch the images (right). The optimisation has been run upon 20×20 samples only, distributed uniformly across the 3MPixel image. No heuristic smoothing or manual selection of a "good" pyramid level was applied. Note that this is an extremely challenging situation because of the frequency content. Remaining errors may be due to lens distortion, camera movement and changed illumination.

the images of fig.5 have been chosen, where SIFT features were detected followed by a rotation of the images. Around each feature 21×21 samples have been used in a square window 10 times the detection σ (cf. to [3]). Then Euclidian parameters have been estimated with correct position prediction but with no rotation prediction. When the rotation was estimated worse than 0.05 rad, a failure has been recorded. The graphs show that for very small rotation errors also the pyramid approach converged, but for larger rotational errors the novel approaches diverge less frequently, because fine structures are better exploited here. In the next experiment the automatic scale approach is demonstrated based on extremely sparse samples. We applied gradient based homography estimation for a real pair of photos containing high-frequency patterns of skyscrapers. No heuristic smoothing or *some good pyramid level* had to be selected. Instead, a prediction for the homography parameters was approximated by propagating the rotational uncertainty of 1° (see fig. 6 for details). For such warps with higher numbers of parameters heuristic smoothing becomes really involved, while our framework solves this problem automatically.

5 Conclusion

Image pyramids have been used in gradient-based displacement estimation for a long time to increase the convergence radius. When more complex parametric image transformations were considered, the pyramid concept has simply been adopted in the literature so far or the images under inspection had to be provided "smooth enough" for convergence. In this contribution we developed a novel framework, which incorporates parameter uncertainty into the registration process working in scale space. Given a parameter guess and its approximate uncertainty, the system selects the required amount of smoothing automatically on a per-sample basis, which allows to keep more detail of the original image and therefore is less susceptible to the aperture problem. It can be seen as a generalization of the pyramid concept from displacement. Although the evaluation showed superiority using local feature alignment, the concept can be applied to a much broader range of parameter estimation applications as camera tracking or homography estimation.

References

[1] T. Lindeberg, "Scale-space theory: A basic tool for analysing structures at different scales.", Journal of Applied Statistics, 21(2):224-270, 1994

[2] B. D. Lucas and T. Kanade."An Iterative Image Registration Technique with an Application to Stereo Vision." Int. Conf. on Artificial Intelligence, pp. 674-679, 1981

[3] D. G. Lowe "Distinctive image features from scale-invariant keypoints" International Journal of Computer Vision, 60, 2 (2004), pp. 91-110

[4] J. Shi and C. Tomasi. "Good features to track". Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 593-600, June 1994.

[5] W.J.Christmas "Spatial filtering Requirements for Gradient-based optical flow measurement", Proc. British Machine Vision Conference 1997, Southampton, UK, 1997

[6] N.Molton, A.Davison, I.Reid "Parameterisation and Probability in Image Alignment", OUEL Report 2266/03, University of Oxford, 2003

[7] J.R.Bergen, P.Anandan, K.J.Hanna, R.Hingorani "Hierarchical Model-Based Motion Estimation", Proceedings of ECCV92 (LNCS 588), pp. 237 - 252

[8] Mikolajczyk, K. , Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Van Gool, L.,"A Comparison of Affine Region Detectors", International Journal of Computer Vision 65(1-2), pp.43–72, 2005

[9] Baker, S. and Matthews, I., "Lucas-Kanade 20 Years On: A Unifying Framework", IJCV 56(3), pp.221-255, Springer-Verlag, 2004

[10] M. Lefébure and L.D. Cohen. "Image Registration, Optical Flow and Local Rigidity", Journal of Mathematical Imaging and Vision 14(2), pp.131-147, 2001

[11] K.Koeser and R.Koch,"Differential Spatial Resection", to be published at European Conference on Computer Vision, Marseille, 2008

[12] K.Koeser, C.Beder and R.Koch,"Conjugate Rotation: Parameterization and Estimation from an Affine Feature Correspondence", CVPR, Anchorage, 2008

[13] R.Steele and C. Jaynes."Feature Uncertainty Arising From Covariant Image Noise", Proc. of CVPR 2005, pp.1063–1070

[14] L.Dorini, S.Goldenstein. "Unscented KLT: Nonlinear Feature and Uncertainty Tracking", Computer Graphics and Image Proc. (SIBGRAPI), Brazil, 2006.

[15] L. Williams: "Pyramidal Parametrics", Computer Graphics 17(3), pp.1–11, 1983