Iterative gradient-based shift estimation: to multiscale or not to multiscale?

Martin Rais[1 2 ⋆], Jean-Michel Morel[2], and Gabriele Facciolo[2]

[1] DMI, UIB, Spain
[2] CMLA, ENS-Cachan, France

Abstract. Fast global shift estimation is a critical preprocessing step on many high level tasks such as remote sensing or medical imaging. In this work we deal with a simple question: should we use an iterative technique to perform shift estimation or should we use a multiscale approach. Based on the obtained results, both methodologies proved to lose accuracy as the noise increases, however this accuracy loss increases with the shift magnitude. The conclusion is that a multiscale strategy should be used when the shift magnitude is higher than approximately a fifth of a pixel.

Keywords: Shift estimation. Multiscale. Iterative.

1 Introduction

Given two images shifted by some unknown displacement $\mathbf{v}$, the problem of shift estimation is to compute this displacement. Problems such as low SNR conditions, lack of image structure and quantization errors make this task non trivial. Several issues appear as well when seeking for accurate subpixel shifts. Nevertheless, precise and real-time shift estimation methods are required in many fields, such as remote sensing [11, 4, 15] or medical imaging [16, 6].

As mentioned in [17], there are mainly four types of shift estimation methods that achieve subpixel accuracy: correlation interpolation, intensity interpolation, differential methods and phase correlation.

Correlation interpolation methods achieve subpixel accuracy by fitting an interpolation surface to the samples of a discrete correlation function, and then, the maximum of this surface is searched. This methodology not only implies calculating the discrete correlation between images, which is a resource consuming task, but also to interpolate it. A more straightforward way to achieve subpixel accuracy is to interpolate selected parts of the input images to create a much

⋆ During this work, the author had a fellowship of the Ministerio de Economia y Competividad (Spain), reference BES-2012-057113, for the realization of his Ph.D. thesis. This work was also partially supported by the Centre National d'Etudes Spatiales (MISS Project), the European Research Council (Advanced Grant Twelve Labours), the Office of Naval Research (under Grant N00014-97-1-0839), Direction Générale de l'Armement, Fondation Mathématique Jacques Hadamard and Agence Nationale de la Recherche (Stereo project).

denser grid. Then, the task is to match these grids between images, which requires knowing beforehand which part of the input images to interpolate and then match, something which is not always available.

To date, there are mainly two fast and accurate shift estimation methods from which several branches have emerged. The first one is based on the phase correlation technique [12, 3, 5], in which the displacement is estimated using the cross-power spectrum between both images. This technique, although able to obtain quite accurate results, requires at least the computation of the DFT for both input images, a task that could be prohibitive depending on the context.

On the other hand, differential methods are normally faster than Fourier-based methods, since they do not require computing the DFT. By using a differential technique, the difference between two frames is related with the spatial intensity gradient of the first image. Given $I_1(x, y)$ and $I_2(x, y)$, and denoting the components of the translation between both images by v_x and v_y, we have

$$I_2(x, y) = I_1(x - v_x, y - v_y). \tag{1}$$

Using the first order Taylor expansion yields

$$I_1(x, y) - I_2(x, y) \approx v_x \frac{\partial I_1(x, y)}{\partial x} + v_y \frac{\partial I_1(x, y)}{\partial y} \tag{2}$$

which is a formula known under the name of optical flow equation [6]. Since the higher terms of the Taylor approximation were removed, this relation performs well only when the translation is small, in particular when it is less than one pixel. The unknown shift $\mathbf{v}$ is estimated by minimizing the error in this equation. This can be done by minimizing the L_2 norm. Using linear least squares is a classical solution, introduced by Lucas-Kanade [7], and has linear complexity,

This estimator, however, ignores the higher terms of the Taylor development and the fact that the underlying input images have noise, which biases the results. A complete study on this bias was performed by Robinson and Milanfar [13], followed by Pham *et al.* [9]. In these works, an explicit formula for the Lucas-Kanade estimator bias was derived. However, these authors address the bias in two completely different ways.

Robinson and Milanfar tried to reduce the influence of the estimator's bias by designing a gradient estimation filter (i.e. antisymmetric) which minimized its bias derivation in the Fourier domain based on the selection of pre-filters and on the prior knowledge of the image spectrum and some constraint about the shift [14]. Surprisingly, this article proposed to minimize the estimator bias by attacking the approximation error in the data model due to the linear signal approximation performed by the Taylor development, while completely ignoring the noise. In fact, low Signal-to-Noise Ratio situations are discarded since they claim that in many image registration applications, the effective SNR falls into a high SNR regime. For this reason, they achieve poor results on images with SNR lower than 20dB. Furthermore, none of these previous approaches work under aliased situations or badly sampled images, which are possible (yet undesired) in computer vision problems.

Pham *et al.*, on the other hand, derived a complete 2D gradient-based shift estimation bias in the spatial domain. However, instead of dealing with it explicitly, they note the linear dependence between the estimator's bias with both the shift magnitude and the noise. They also remark that the bias due to the noise is proportional to the shift magnitude as well. Thus, they propose to reduce the bias by iteratively computing the shift and resampling the second image onto the first. Based on their results, with only three iterations they are able to obtain an almost unbiased estimator. This iterative scheme was actually proposed by Lucas and Kanade in [7], and further refined in Baker *et al.* [1] in a complete study of the Lucas-Kanade estimator. However, this iterative scheme involves performing interpolation, which becomes an expensive computation.

A different iterative scheme, such as in Thevenaz *et al.* [16], consists in computing the image pyramid and to perform shift estimation on each level separately. Beginning by the coarser level, the estimated shift is then used to resample the second image on the next finer scale. Although this technique requires the construction of the pyramid, it can allow itself to use more complex interpolation techniques on lower scales due to its reduced cost. Most importantly, the shift estimation performed on each scale could also be made iterative, a price that can be payed when working on coarser scales.

Objective. Both iterative techniques (direct and multiscale) succeed in reducing the bias when enough iterations are applied. However, it is not straightforward, based on the shift magnitude and on the noise conditions of the input images, to estimate which methodology achieves better results on each condition. For example, if the shift magnitude is above one pixel, the multiscale approach will definitely be necessary. Furthermore, under noisy conditions, working on a coarser scale permits to reduce the noise influence, however less pixels (and thus equations) will be available to perform shift estimation.

In this article we evaluate both methodologies by varying the noise, the shift to estimate, the derivative kernels used, the amount of iterations and the underlying interpolation method in order to understand how each methodology performs. What is more, we answer the question of deciding between applying a multiscale approach or sticking to the original solution.

The rest of this paper is organized as follows. In section 2 both methodologies are explained in detail. In section 3 we evaluate each of them, under all possible conditions, and based on this we draw conclusions on section 4.

2 Methods

2.1 Iterative Lucas-Kanade shift estimation method

The Lucas-Kanade algorithm is based on the optical flow equation:

$$I_t(x,y) \approx \nabla I_1(x,y)\mathbf{v}. \tag{3}$$

where $\nabla I_1(x,y) = \left[\frac{\partial I_1(x,y)}{\partial x}, \frac{\partial I_1(x,y)}{\partial y}\right]$ and $\mathbf{v} = \begin{bmatrix} v_x \\ v_y \end{bmatrix}$ is the unknown shift. Then, to estimate the global optical flow between both grayscale images I_1 and I_2, the Lucas-Kanade algorithm assumes a constant flow for the whole image, thus implying a unique translation vector $\mathbf{v}$ for every pixel. This assumption leads to the construction of an overdetermined system of equations, $\mathbf{Av} = \mathbf{b}$, where $\mathbf{A}$ is composed of spatial intensity derivatives and $\mathbf{b}$ has temporal derivatives

$$\mathbf{A} = \begin{pmatrix} \frac{\partial I_1}{\partial x}(p_1) & \frac{\partial I_1}{\partial y}(p_1) \\ \vdots & \vdots \\ \frac{\partial I_1}{\partial x}(p_n) & \frac{\partial I_1}{\partial y}(p_n) \end{pmatrix} \qquad \mathbf{v} = \begin{pmatrix} v_x \\ v_y \end{pmatrix} \qquad \mathbf{b} = -\begin{pmatrix} \frac{\partial I_1}{\partial t}(p_1) \\ \vdots \\ \frac{\partial I_1}{\partial t}(p_n) \end{pmatrix} \tag{4}$$

and p_i with $i = 1 \dots n$ represents the ith pixel and n the number of pixels. To solve this system, (v_x, v_y) is obtained by performing the linear least squares method, using the Moore-Penrose pseudo-inverse. Let I_x, I_y and I_t denote $\frac{\partial I_1}{\partial x}$, $\frac{\partial I_1}{\partial y}$ and $\frac{\partial I_1}{\partial t}$ respectively, the following linear system has to be solved

$$\mathbf{A}^T\mathbf{A}\mathbf{v} = \mathbf{A}^T\mathbf{b} \tag{5}$$

where $\mathbf{A}^T\mathbf{A} = \begin{bmatrix} \sum I_x^2 & \sum I_x I_y \\ \sum I_x I_y & \sum I_y^2 \end{bmatrix}$ is the second moment matrix, and $\mathbf{A}^T\mathbf{b} = \begin{bmatrix} \sum I_t I_x \\ \sum I_t I_y \end{bmatrix}$ is a spatio-temporal gradient correlation term. To solve this system, the matrix $\mathbf{A}^T\mathbf{A}$ must be invertible in which case the solution is $(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$.

It is not a coincidence that the results of the method depend on the inversion of this second moment matrix since the determinant of this matrix is crucial for determining the limits on the estimation performance [9]. A study on this matrix before performing the shift estimation can be used to discard ill-posed cases. This happens for example when the gradient occurs on its majority on a single direction and therefore we are dealing with a potentially unsolvable situation, commonly known as the aperture problem. Last but not least, since the Taylor development is centered at 0, this method performs well only when the translation is small, i.e., shifts larger than 1 would not be correctly estimated.

Lucas and Kanade also suggest iterating the method to obtain better results and converge to the true displacement value. This algorithm is easily understood in the following lines:

$i \leftarrow 0;\ I_2(0) \leftarrow I_2;\ w \leftarrow 0$
while $i \leq k$ **and** $|v(i-1) - v(i-2)| \geq min$ **do**
    $v(i) \leftarrow findshift(I_1, I_2(i))$
    $w \leftarrow w + v(i)$
    $I_2(i+1) \leftarrow Resample(I_2, -w)$
    $i \leftarrow i + 1$

where $findshift$ uses Eq. (5) to solve for $v(i)$ and *Resample* performs interpolation on the input images, which is a costly operation. In particular, if an inappropriate interpolation algorithm is used, it could lead to poor results, implying a non-negligible computational cost.

On the other hand, it was proved in Pham *et al.* [9] that this iterative method is able to significantly reduce the bias, provided an appropriate resampling method is used. Due to this reason, very poor results are obtained when dealing with highly aliased images, as shown in [10] when a single iteration outperforms the multi-iteration method. Nevertheless, with a correct resampling and with a sufficient number of iterations, this method is the only capable, to our best knowledge, of practically removing the bias.

2.2 Multiscale Lucas Kanade shift estimation method

By building a pyramid representation of the input images, Eq. (5) can be applied on each scale to estimate the shift between images, and this estimated shift can in turn be used to resample the second image on the following level of the pyramid. In our case, a dyadic Gaussian pyramid approximation was used [2]. We also evaluated using an exact dyadic Gaussian pyramid [8], filtering with $\sigma = 1.4$ before subsampling, however the results were similar. Starting from the coarse image at scale $n > 1$, the method is summarized in the following lines:

$I_1^{1\dots n} \leftarrow ComputeImagePyramid(I_1, n)$ *// Burt & Adelson's Gaussian Pyramid [2],*
$I_2^{1\dots n} \leftarrow ComputeImagePyramid(I_2, n)$ *// i.e.,* IMPYRAMID *function from Matlab*
$i \leftarrow n;\quad w \leftarrow 0$
while $i > 0$ **do**
  $v(i) \leftarrow findshift(I_1^i, I_2^i)$
  $w \leftarrow w * 2 + v(i) * 2$
  $I_2^{i-1} \leftarrow Resample(I_2^{i-1}, -w)$
  $i \leftarrow i - 1$
$v(i) = findshift(I_1^1, I_2^1)$

3 Results

Both methodologies described in sections 2.1 and 2.2 were evaluated extensively under different noise conditions, shifts and gradient estimators. To show the most representative results, four SNR conditions were evaluated: noiseless, low noise ($\sigma = 5$), medium noise ($\sigma = 25$) and high noise ($\sigma = 50$). Each table is organized in groups of four lines corresponding to each of these four noise configurations. Also, the four most significative shifts in terms of results are shown: a big shift $(0.5, -0.9)$, a medium shift $(0.2, -0.2)$, a small shift $(0.024, 0.052)$ and no shift.

The performance of each algorithm under each condition was evaluated by simulating shifted images obtained from a high resolution satellite image of 10000×10000 pixels. For each noise and shift, 100 experiments were averaged, and each experiment was performed by shifting the large image using Fourier interpolation and taking a 50×50 subimage from a random position away from the edges to avoid ringing artifacts followed by adding white Gaussian noise and evaluating all the methods (Fig. 1). The results shown were later validated using the Cramer-Rao bound (verifying that both $var(\hat{v_x})$ and $var(\hat{v_y})$ are lower than 0.01) so that the averaged values contain only valid shift estimations.

In tables 1 and 2 results are shown for 2 iterations and bicubic interpolation and for 3 iterations with spline interpolation respectively. From these results several conclusions can be drawn. First, as expected, the multiscale method is

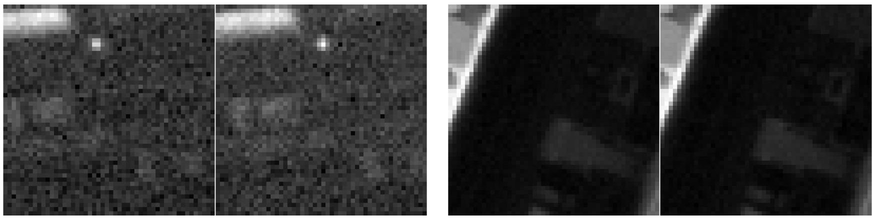

Fig. 1. Two problems: a noisy and an almost unidimensional gradient situation

Shift (px)	**Noise**	**IT2G1**	**IT2G2**	**IT2G3**	**MS2G1**	**MS2G2**	**MS2G3**
(0.5000,-0.9000)	$\sigma = 0$	0.0514	0.0818	0.0472	0.0387	0.1600	**0.0316**
	$\sigma = 75$	0.1375	0.1053	0.1103	0.0744	0.1808	**0.0582**
	$\sigma = 150$	0.2875	0.1414	0.2305	0.1267	0.2130	**0.1009**
	$\sigma = 300$	0.4927	0.2327	0.4292	0.2319	0.2872	**0.1909**
(0.0000,0.0000)	$\sigma = 0$	**0.0000**	0.0000	0.0000	0.0000	0.0000	0.0000
	$\sigma = 75$	**0.0114**	0.0188	0.0132	0.0164	0.0188	0.0175
	$\sigma = 150$	**0.0168**	0.0330	0.0219	0.0300	0.0326	0.0313
	$\sigma = 300$	**0.0191**	0.0539	0.0307	0.0527	0.0549	0.0573
(0.2000,-0.2000)	$\sigma = 0$	0.0115	0.0117	0.0154	**0.0115**	0.0360	0.0159
	$\sigma = 75$	0.0249	0.0267	0.0223	**0.0192**	0.0470	0.0225
	$\sigma = 150$	0.0652	0.0424	0.0487	0.0360	0.0591	**0.0358**
	$\sigma = 300$	0.1295	0.0765	0.1103	0.0738	0.0899	**0.0694**
(0.0240,0.0520)	$\sigma = 0$	0.0040	**0.0019**	0.0052	0.0039	0.0073	0.0054
	$\sigma = 75$	**0.0122**	0.0181	0.0138	0.0156	0.0189	0.0166
	$\sigma = 150$	**0.0198**	0.0326	0.0231	0.0296	0.0326	0.0311
	$\sigma = 300$	**0.0300**	0.0543	0.0354	0.0547	0.0564	0.0581
Avg.	$\sigma = 0$	0.0167	0.0238	0.0170	0.0135	0.0508	**0.0132**
	$\sigma = 75$	0.0465	0.0422	0.0399	0.0314	0.0664	**0.0287**
	$\sigma = 150$	0.0973	0.0624	0.0811	0.0556	0.0843	**0.0498**
	$\sigma = 300$	0.1678	0.1043	0.1514	0.1033	0.1221	**0.0939**

Table 1. Estimation error (in pixels) per shift of every method using **2 iterations** and **bicubic** interpolation from valid estimations. For each shift and estimation method, four SNR conditions were tested. The first three columns are for the iterative method (IT) while the last three are for the multiscale approach (MS) with a single iteration per scale. In each case, three gradient estimation methods were used: backward difference and Gaussian derivative with $\sigma = 1$ and with $\sigma = 0.3$ respectively.

much more robust when the shift magnitude is high. In fact, even at a shift of (0.2,-0.2) it is recommendable to use the multiscale method instead of the standard iterative version. Second, when no shift or a small shift is present, the non-multiscale methods achieve much better accuracies. Apparently, the multiscale algorithms are not suited for such small shifts since their poor performance on lower scales results in less accurate results. This result contradicts several state-of-the-art methods and is worth remarking. Third, regarding the amount of iterations/scales to use, in presence of high noise, performing more iterations

Shift (px)	Noise	IT3G1	IT3G2	IT3G3	MS3G1	MS3G2	MS3G3
(0.5000,-0.9000)	$\sigma = 0$	0.0156	0.0238	**0.0065**	0.0114	0.1007	0.0086
	$\sigma = 75$	0.0646	0.0437	0.0437	0.0293	0.1194	**0.0251**
	$\sigma = 150$	0.1869	0.0727	0.1326	0.0533	0.1488	**0.0497**
	$\sigma = 300$	0.4092	0.1480	0.3337	0.1093	0.2143	**0.1001**
(0.0000,0.0000)	$\sigma = 0$	**0.0000**	0.0000	0.0000	0.0000	0.0000	0.0000
	$\sigma = 75$	**0.0116**	0.0199	0.0133	0.0261	0.0330	0.0260
	$\sigma = 150$	**0.0194**	0.0357	0.0246	0.0468	0.0547	0.0467
	$\sigma = 300$	**0.0250**	0.0621	0.0391	0.0980	0.1065	0.1001
(0.2000,-0.2000)	$\sigma = 0$	**0.0027**	0.0045	0.0043	0.0206	0.0205	0.0235
	$\sigma = 75$	0.0201	0.0229	**0.0182**	0.0304	0.0430	0.0334
	$\sigma = 150$	0.0475	0.0395	**0.0373**	0.0483	0.0593	0.0489
	$\sigma = 300$	0.1096	**0.0718**	0.0913	0.0982	0.1195	0.0977
(0.0240,0.0520)	$\sigma = 0$	**0.0007**	0.0009	0.0013	0.0068	0.0041	0.0080
	$\sigma = 75$	**0.0120**	0.0188	0.0132	0.0219	0.0283	0.0225
	$\sigma = 150$	**0.0207**	0.0348	0.0248	0.0471	0.0548	0.0464
	$\sigma = 300$	**0.0313**	0.0622	0.0412	0.1011	0.1056	0.1001
Avg.	$\sigma = 0$	0.0047	0.0073	**0.0030**	0.0097	0.0313	0.0100
	$\sigma = 75$	0.0271	0.0263	**0.0221**	0.0269	0.0559	0.0268
	$\sigma = 150$	0.0686	**0.0457**	0.0548	0.0489	0.0794	0.0479
	$\sigma = 300$	0.1438	**0.0860**	0.1263	0.1017	0.1365	0.0995

Table 2. Estimation error (in pixels) per shift of every method using **3 iterations** and **spline** interpolation from valid estimations. Table configuration is the same as in Table 1.

in the original scale or using more scales in the multiscale approach gives worse results in terms of accuracy. When dealing with a noisy situation, the resampling operation proved to be negative for the shift estimation algorithm. This result is more accentuated for the multiscale approach. Finally, the multiscale algorithm proved to be a better contender when dealing with noise in general, although this factor is greatly influenced by the shift magnitude. However, except when the shift magnitude is lower than a fifth of a pixel, its use is recommended. Moreover, its computational cost is lower than the iterative counterpart since the resampling is performed on lower resolution images.

4 Conclusions

In this paper we dealt with a simple question never answered in the community: should we use a multiscale strategy to perform gradient based shift estimation or should we directly attack the problem by simply iterating in the original scale. The answer of this question was shown to depend heavily on the shift magnitude more than the SNR of the images. Under small shift magnitudes, performing a multiscale strategy achieves poorer results, in particular due to the lack of displacement on the lower scales that makes the method less accurate. However, when dealing with shifts higher than one fifth of a pixel, the multiscale strategy showed strong improvements over traditional iterative Lucas-Kanade shift estimation. Last but not least, in situations under low SNR, we concluded that performing fewer iterations or using fewer scales achieves improved accuracy, and this result is even more remarked in the multiscale approach. As a future work, experimentation is planned on a larger dataset of images with different characteristics, and by testing several other interpolation methods for image resampling when iterating the algorithm.

References

1. Baker, S., Matthews, I.: Lucas-kanade 20 years on: A unifying framework. Int. J. Comput. Vision 56(3), 221–255 (Feb 2004)
2. Burt, P.J., Adelson, E.H.: The laplacian pyramid as a compact image code. Communications, IEEE Transactions on 31(4), 532–540 (1983)
3. Foroosh, H., Zerubia, J., Berthod, M.: Extension of phase correlation to subpixel registration. IEEE TIP 11(3), 188 –200 (mar 2002)
4. Goshtasby, A., Stockman, G., Page, C.: A Region-Based Approach to Digital Image Registration with Subpixel Accuracy. IEEE Transactions on Geoscience and Remote Sensing GE-24(3) (1986)
5. Guizar-Sicairos, M., Thurman, S.T., Fienup, J.R.: Efficient subpixel image registration algorithms. Opt. Lett. 33(2), 156–158 (Jan 2008)
6. Horn, B.K., Schunck, B.G.: Determining optical flow (1981)
7. Lucas, B.D., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: Proceedings of the 7th International Joint Conference on Artificial Intelligence - Volume 2. pp. 674–679 (1981)
8. Morel, J.M., Yu, G.: Is sift scale invariant? Inverse Problems and Imaging 5(1), 115–136 (2011)
9. Pham, T.Q., Bezuijen, M., Van Vliet, L.J., Schutte, K., Luengo Hendriks, C.L.: Performance of optimal registration estimators. In: Proc. SPIE. vol. 5817, pp. 133–144 (2005)
10. Pham, T., Duggan, M.: Bidirectinal bias correction for gradient-based shift estimation. In: IEEE ICIP. pp. 829–832 (Oct 2008)
11. Rais, M., Thiebaut, C., Delvit, J.M., Morel, J.M.: A tight multiframe registration problem with application to earth observation satellite design. In: 2014 IEEE International Conference on Imaging Systems and Techniques. pp. 6–10 (2014)
12. Reddy, B., Chatterji, B.: An fft-based technique for translation, rotation, and scale-invariant image registration. IEEE TIP 5(8), 1266 –1271 (aug 1996)
13. Robinson, D., Milanfar, P.: Fundamental performance limits in image registration. IEEE TIP 13(9), 1185–1199 (Sept 2004)
14. Robinson, D., Milanfar, P.: Bias minimizing filter design for gradient-based image registration. Signal Processing: Image Communication 20(6), 554 – 568 (2005), special Issue on Advanced Aspects of Motion Estimation
15. Sabater, N., Leprince, S., Avouac, J.P.: Contrast Invariant and Affine sub-pixel Optical Flow. In: IEEE ICIP. pp. 53–56 (2012)
16. Thévenaz, P., Ruttimann, U.E., Unser, M.: A pyramid approach to subpixel registration based on intensity. IEEE TIP 7(1), 27–41 (1998)
17. Tian, Q., Huhns, M.N.: Algorithms for subpixel registration. Comput. Vision Graph. Image Process. 35(2), 220–233 (Aug 1986)