GRADIENT SCHEMES FOR ROBUST FFT-BASED MOTION ESTIMATION

Georgios Tzimiropoulos

University of Nottingham
Computer Science
Nottingham, NG8 1BB

Vasileios Argyriou

Kingston University London
Computer Science
Surrey, KT1 2EE

ABSTRACT

In this work, the focus is on gradient-based correlation schemes which constitute an alternative to phase correlation for FFT-based motion estimation. In particular, our contribution is threefold. First, we present an analysis which highlights the key features of gradient schemes. Second, we introduce an illumination invariant to gradient correlation. Third, we provide a comparison of gradient schemes in the application of block matching for video coding and draw several useful observations and conclusions.

Index Terms— motion estimation, correlation methods, block matching

1. INTRODUCTION

The estimation of the relative motions between two or more images is probably at the heart of any autonomous system which aims at the efficient processing of visual information. Motions in images are induced due to camera displacements or displacements of the individual objects composing the scene. Image registration techniques address the problem of estimating all possible motions and finally aligning the images. In this work, the focus is on methods for the registration of images assumed to be related by a pure translation. Translation estimation is fundamental in a number of applications such as image super-resolution, medical image registration and video coding.

Recently, correlation-based motion estimation in the Fourier domain has received considerable attention in the computer vision community. A frequency domain approach offers speed through the use of FFT routines and enjoys a high degree of accuracy through the use of supplementary schemes which can achieve subpixel registration.

Perhaps the most popular Fourier-based method is phase correlation (PC) [1]. The method capitalizes on the shift property of the Fourier Transform (FT) [1] and the importance of phase in images [2]. It features several remarkable properties: strong response to edges and image salient features, immunity to uniform variations of illumination, insensitivity to changes in spectral energy and most importantly excellent peak localization accuracy. Recently, several subpixel extensions have

been proposed [9]-[23]. In [15] and [7], authors propose to perform the unwrapping after applying a rank-1 approximation to the phase difference matrix and then it is extended for noisy data. To estimate the subpixel shifts, Stone et al. [13] fit the phase values to a 2-D linear function using linear regression, after masking out frequency components corrupted by aliasing. Foroosh et al. [14] showed that the PC function is the Dirichlet kernel and provided analytic results for the estimation of the subpixel shifts using the $sinc$ approximation. Notice that the above methods either assume aliasing-free images or cope with aliasing by frequency masking which requires fine tuning.

Our approach to FFT-based motion estimation, although similar in concept, differs from phase correlation and its extensions, [8]. The main idea is very simple: extract image gradients, combine them in a complex gray level edge map which retains both magnitude and orientation information and then correlate. In this paper, we briefly review gradient-based correlation schemes namely gradient correlation (GC) [3],[8] and orientation correlation (OC) [4]. We present a spatial domain analysis which shows that under a certain assumption, the use of image gradients tailors correlation to the nature of images and provides a mechanism to reject outliers induced by realworld registration problems. We define an illumination invariant extension to GC, the normalized gradient correlation (NGC). We also present an analysis of GC in the presence of uncorrelated white noise. Finally, we evaluate gradient correlation schemes and phase correlation in the application of block motion estimation for video coding and draw several useful observations and conclusions.
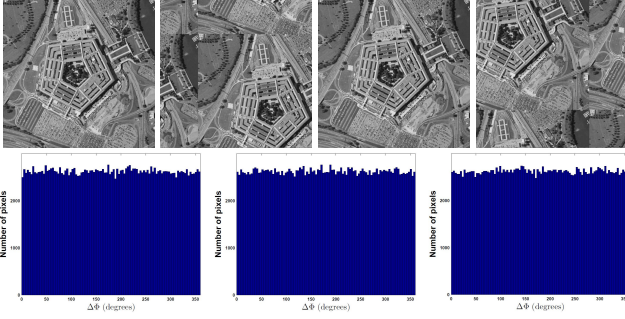
2. GRADIENT SCHEMES FOR ROBUST MOTION ESTIMATION

2.1. Gradient and orientation correlation

Let $I_i(\mathbf{x})$, $\mathbf{x} = [x, y]^T \in \mathcal{R}^2$, $i = 1, 2$ be two image functions, related by an unknown translation $\mathbf{t} = [t_x, t_y]^T \in \mathcal{R}^2$

$$I_1(\mathbf{x} + \mathbf{t}) = I_2(\mathbf{x}) \qquad (1)$$

To estimate the translational displacement, we use gradient based correlation schemes. Gradient correlation (GC) com-

Fig. 1. (Top row) The 512×512 pentagon image and the pentagon image circularly shifted by $\mathbf{t} = [t_x; t_y]^T$, where $t_x = t_y = 128; 256$ and 384 respectively. (Bottom row) For each shift, we compute the difference in orientation $\Delta\Phi$ between the original image and the shifted image.

bines the magnitude and orientation of image gradients [3]

$$\text{GC}(\mathbf{u}) \triangleq G_1(\mathbf{u}) \star G_2^*(-\mathbf{u}) = \int_{\mathcal{R}^2} G_1(\mathbf{x})G_2^*(\mathbf{x}+\mathbf{u})d\mathbf{x} \quad (2)$$

where

$$G_i(\mathbf{x}) = G_{i,x}(\mathbf{x}) + jG_{i,y}(\mathbf{x}) \quad (3)$$

and $G_{i,x} = \nabla_x I_i$ and $G_{i,y} = \nabla_y I_i$ are the gradients along the horizontal and vertical direction respectively. Orientation correlation (OC) considers orientation information solely [4]

$$\text{OC}(\mathbf{u}) \triangleq O_1(\mathbf{u}) \star O_2^*(-\mathbf{u}) = \int_{\mathcal{R}^2} O_1(\mathbf{x})O_2^*(\mathbf{x}+\mathbf{u})d\mathbf{x} \quad (4)$$

where

$$O_i(\mathbf{x}) \leftarrow \begin{cases} G_i(\mathbf{x})/|G_i(\mathbf{x})|, & \text{if } |G_i(\mathbf{x})| \neq 0 \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

From (2) and (4) we estimate t as the argument which maximizes the correlation functions.

For finite discrete images of size $N \times N$, all correlations are efficiently implemented in the frequency domain, by zero padding the images to size $(2N-1) \times (2N-1)$ and using FFT routines to compute the forward and inverse FTs. If no zero padding is used, the match is cyclic and, in this case, the algorithms complexity is $O(N^2 \log N)$.

2.2. Spatial domain analysis

In [3], the authors presented a frequency domain analysis of GC. To complete the picture and provide additional insight, in this section, we present a spatial domain analysis of gradient schemes as follows.

Plugging (3) into (2), we can easily derive

$$\text{GC}(\mathbf{u}) = G_{1,x}(\mathbf{u}) \star G_{2,x}(-\mathbf{u}) + G_{1,y}(\mathbf{u}) \star G_{2,y}(-\mathbf{u}) \\ +j\{-G_{1,x}(\mathbf{u}) \star G_{2,y}(-\mathbf{u}) + G_{1,y}(\mathbf{u}) \star G_{2,x}(\text{-}\mathbf{u})\} \quad (6)$$

The imaginary part in the above equation is equal to zero, therefore

$$\text{GC}(\mathbf{u}) = G_{1,x}(\mathbf{u}) \star G_{2,x}(-\mathbf{u}) + G_{1,y}(\mathbf{u}) \star G_{2,y}(-\mathbf{u}) \quad (7)$$

Using the polar representation of complex numbers, we define $R_i = \sqrt{G_{i,x}^2 + G_{i,y}^2}$ and $\Phi_i = \arctan G_{i,y}/G_{i,x}$. Based on this representation, (5) takes the form

$$\text{GC}(\mathbf{u}) = \int_{\mathcal{R}^2} R_1(\mathbf{x})R_2^*(\mathbf{x}+\mathbf{u})\cos[\Phi_1(\mathbf{x}) - \Phi_2(\mathbf{x}+\mathbf{u})]d\mathbf{x} \quad (8)$$

Each term in (6) has its own special importance. The magnitudes Ri reward pixel locations with strong edge responses and suppress the contribution of areas of constant intensity level which do not provide any reference points for motion estimation. Orientation information is embedded in the cosine kernel. This term is responsible for the dirac-like shape of GC and its ability to reject outliers induced by the presence of dissimilar parts in the two images.

To roughly show the latter point, let us first assume that at $\mathbf{u} \neq \mathbf{t}$ the orientation difference function $\Delta\Phi(\mathbf{u}, \mathbf{x}) = \Phi_1(\mathbf{x}) - \Phi_2(\mathbf{x}+\mathbf{u})$ is uniformly distributed over $[-\pi, \pi)$. This assumption appears to be reasonable, since for displacements other than the correct, the images do not match and therefore we expect that differences in gradient orientation can take any value in the range $[0, 2\pi)$ with equal probability. For example, Fig. 1 (a) shows the 'Pentagon' image. Circularly shifting the image in three different fashions yields the images in Fig. 1 (b)-(d). For each shift, we compute the difference in orientation $\Delta\Phi$ between the original and the shifted image. For each case, Fig. 1 (e)-(g) show the histogram with the distribution of $\Delta\Phi$. In all cases, we can observe that $\Delta\Phi$ is well-described by a uniform distribution.

Let us further impose $R_i = 1, i = 1, 2$, that is we essentially compute a modified orientation correlation function (mOC) where, in contrary to (4), the orientation of all pixels is taken into consideration

$$mOC(\mathbf{u}) = \int_{\mathcal{R}^2} \cos[\Phi_1(\mathbf{x}) - \Phi_2(\mathbf{x}+\mathbf{u})]d\mathbf{x} \quad (9)$$

To model dissimilar parts, we modify the perfect translational model of (1) as follows

$$I_1(\mathbf{x}+\mathbf{t}) = I_2(\mathbf{x}), \mathbf{x} \in \Omega \subseteq \mathcal{R}^2 \quad (10)$$

That is after shifting I_1 by $\mathbf{t}$, I_1 and I_2 match only in $\mathbf{x} \in \Omega$. Since $\forall \mathbf{u} \neq \mathbf{t} \Delta\Phi(\mathbf{u}, \mathbf{x})$ is uniformly distributed, the integral in (7) will be equal to zero. At $\mathbf{u} = \mathbf{t}$), we have

$$mOC(\mathbf{t}) = \int_{\Omega} \cos \Delta\Phi(\mathbf{t}, \mathbf{x})d\mathbf{x} \\ + \int_{\mathcal{R}^2 - \Omega} \cos \Delta\Phi(\mathbf{t}, \mathbf{x})d\mathbf{x} = \int_{\Omega} d\mathbf{x} \quad (11)$$

since $\Delta\Phi(\mathbf{t}, \mathbf{x}) = 0 \forall \mathbf{x} \in \Omega$ and $\Delta\Phi(\mathbf{t}, \mathbf{x})$ is uniformly distributed if $\mathbf{x} \in \mathcal{R}^2 - \Omega$. Overall, mOC will be non-zero only for $\mathbf{u} = \mathbf{t}$, and its value at that point will be the contribution from the areas in the two images that match solely.

Essentially, using image gradients to perform correlation, the errors induced by outliers are mapped to a uniform distribution for which correlation is well-known to feature robust performance. Our analysis does not impose any bound to the number of outliers. In fact, as their number increases, one would expect that accuracy is enhanced, since $\Delta\Phi$ will better approximate the uniform distribution. In practice, the dynamic range of the algorithm is limited by errors in estimating the image gradients and aliasing effects induced by the FFT. The above analysis agrees with experimental results which have shown that gradient-based correlation schemes are able to estimate translational displacements reliably even when the overlap between the given images is less than 20%. Note that phase correlation is able to register images when the overlap is of the order of 40% [5].

2.3. Normalized gradient correlation

Following the above analysis, we introduce the normalized gradient correlation (NGC)

$$\text{NGC}(\mathbf{u}) \triangleq \frac{\int_{\mathcal{R}^2} R_1(\mathbf{x}) R_2^*(\mathbf{x}+\mathbf{u}) \cos[\Phi_1(\mathbf{x}) - \Phi_2(\mathbf{x}+\mathbf{u})] d\mathbf{x}}{\int_{\mathcal{R}^2} |R_1(\mathbf{x}) R_2(\mathbf{x}+\mathbf{u})| d\mathbf{x}} \tag{12}$$

NGC has two interesting properties:

1. $0 \leqslant |\text{NGC}(\mathbf{u})| \leqslant 1$.

2. Invariance to affine changes in illumination.

The first property provides a measure to assess the correctness of the match. To show the second property consider $I_2'(\mathbf{x}) = aI_2(\mathbf{x}) + b$ with $a, b \in \mathcal{R}$. Then, by differentiation $G_2' = aG_2$; therefore the brightness change due to b is removed. Additionally, $R_2' = aR_2$ and $\Delta\Phi_2' = \Delta\Phi_2$; thus the effect of the contrast change due to a will cancel out in (9). Both properties make NGC, in addition to image registration, particularly useful for template matching applications [6].

2.4. Analysis in the presence of uncorrelated additive white noise

In general, signal differentiation exacerbates noise effects. Nevertheless, under the assumption of white noise, correlation is not affected by the degradation of the signal-to-noise ratio (SNR). To illustrate this, consider the case of a $1D$ signal s corrupted by additive white Gaussian noise n

$$r(t) = s(t) + n(t) \tag{13}$$

where the noise is assumed to be uncorrelated with the signal. The noise autocorrelation is given by $R_n(\tau) = \sigma_n^2 \delta(\tau)$ where σ_n^2 is the noise variance.

If we perform differentiation, then

$$d(t) = s_d(t) + n_d(t) \tag{14}$$

where $d(t) \triangleq \frac{dr(t)}{dt}$, $s_d(t) \triangleq \frac{ds(t)}{dt}$ and $n_d(t) \triangleq \frac{dn(t)}{dt}$. Obviously, n_d is uncorrelated with sd. Its autocorrelation is given by [7]

$$R_{n_d}(\tau) = -\frac{d^2 R_n(\tau)}{d\tau^2} = -\sigma_n^2 \frac{d^2 \delta(\tau)}{d\tau^2} \tag{15}$$

Assume that we are given two signals related by a shift ξ, that is $s_2(t) = s_1(t + \xi)$, and corrupted by additive white Gaussian noise. The cross-correlation of d_1 and d_2 is

$$\begin{aligned} R_{d_1 d_2}(\tau) &= E\{d_1(t) d_2(t+\tau)\} \\ &= E\{[s_{d_1}(t) + n_d(t)] \\ &\quad \times [s_{d_1}(t+\xi+\tau) + n_d(t+\tau)]\} \\ &= E\{s_{d_1(t)} s_{d_1}(t+\xi+\tau)\} - \sigma_n^2 \frac{d^2 \delta(\tau)}{d\tau^2} \end{aligned} \tag{16}$$

since noise and signals are assumed to be uncorrelated. The above result, which easily extends to the $2D$ case, shows that uncorrelated white noise does not affect the estimation process. In contrary, one can show that white noise strongly affects the performance of phase correlation [1].

3. RESULTS

We evaluated GC, OC, NGC and phase correlation (PC) in the application of block motion estimation for video coding. We used 5 well-known 352×288 CIF sequences (Coastguard, Bus, Paris, Foreman, Highway) and the 352×240 Football SIF sequence. Coastguard is a sequence with slow object and camera translational motions. Bus contains fast rigid translations and occlusions. Paris is a typical dialogue scene with fast hand and face motions while the camera remains fixed. Foreman is a head-and-shoulders scene with varying foreground and slow background motions. Highway is characterized by significant scene depth and projective motion. Football is an active sports scene containing fast non-rigid body motions and occlusions.

We used standard MPEG size 16×16 blocks while we obtained sub-pixel accuracy registration by using separable variable fitting on the correlation function [3]. For each sequence, we assessed the performance of each method by applying motion compensation using the estimated motion parameters and computing the well-established peak signal-tonoise ratio (PSNR) of the motion compensated prediction error. In addition to the noise-free case, we repeated the experiments for the case of additive white Gaussian noise with a noise level of 20 dB.

For each method and case, Table 1 presents the mean value of the PSNR over the first 150 frames of each sequence. For the noise free case, Fig. 2 illustrates plots of the PSNR as a function of the frame number for the first 21 frames of each

Table 1. Mean PSNR values for GC, NGC, OC and PC for the application of block motion estimation.

	Coast.	Bus	Paris	Fore.	High.	Foot.
GC	<u>31.20</u>	25.47	<u>32.04</u>	<u>32.90</u>	<u>34.03</u>	<u>22.35</u>
NGC	30.92	25.13	31.94	32.29	33.93	22.22
OC	30.87	<u>25.56</u>	31.60	31.18	32.50	21.73
PC	27.65	23.08	30.66	29.06	31.69	20.54
GC + noise	17.69	<u>17.13</u>	17.46	17.63	<u>18.35</u>	<u>16.42</u>
NGC + noise	<u>17.69</u>	17.11	<u>17.47</u>	<u>17.64</u>	18.34	16.41
OC + noise	17.39	16.64	16.94	17.09	18.07	15.84
PC + noise	16.37	15.27	16.11	16.77	18.01	15.45

Fig. 2. PSNR vs Frame No. for the noise free case. (a) Coastguard (b) Bus (c) Paris (d) Foreman (e) Highway (f) Football.

sequence. From the obtained results, we conclude that, overall, GC achieves the best performance. Additionally, a close examination of Table 1 reveals1:

1. With the exception of Bus, GC and NGC perform consistently better than OC. Interestingly, the gain becomes significant large for sequences where the motion deviates from the translational model (Paris, Foreman, Highway, Football). We conclude that in these cases both magnitude and orientation information of the image gradients must be used for more robust performance.
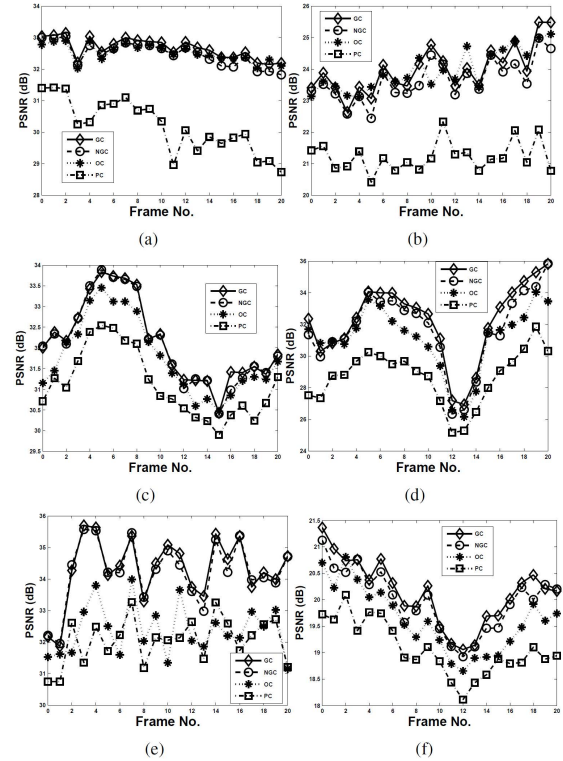
2. GC has a gain of approximately 0.1-0.3 dB over NGC for Coastguard, Bus, Paris, Highway and Football while the gain for Foreman is about 0.6 dB. As expected, the normalization does not offer any advantages in the specific application. In contrary, it seems that the spectral leakage and aliasing effects induced by the computation of the correlation of the absolute gradients make the computation of NGC less accurate. Note however that, for template matching applications, the normalization is necessary.

3. All gradient schemes outperform phase correlation. Compared to GC, the loss in performance varies approximately from 1.5 dB to 3 dB.

4. For the noisy case, GC and NGC attain the same level of performance. Compared to OC, the gain varies approximately from 0.3 dB to 0.6 dB. With the exception of Highway, the gain in performance over PC is more than 1 dB.

4. CONCLUSIONS

In this work, we presented an analysis which highlights the key features of gradient-based correlation schemes. In particular, we demonstrated that the use of image gradients equips correlation with robustness against outliers induced by the presence of dissimilar parts in the two images. Additionally, we showed that uncorrelated white noise does not affect gradient correlation. We also introduced a normalized illumina-

tion invariant extension to gradient correlation. Finally, we evaluated gradient schemes along with phase correlation in the application of block motion estimation. Our results show that, among all methods, GC provides the best results in terms of the PSNR of the motion compensated prediction error.

5. REFERENCES

[1] C.D. Kuglin and D.C. Hines, The phase correlation im-age alignment method, in Proc. IEEE Conf. Cybernetics and Society, 1975, pp. 163165.

[2] A.V. Oppenheim and J.S. Lim, The importance of phase in signals, Proc. of the IEEE, vol. 69, no. 5, pp. 529541, 1981.

[3] V. Argyriou and T. Vlachos, Estimation of sub-pixel motion using gradient cross-correlation, Electronic Letters, vol. 39, no. 13, pp. 980982, 2003.

[4] A.J. Fitch, A. Kadyrov, W.J. Christmas, and J. Kittler, Orientation correlation, in BMVC, 2002, pp. 133142.

[5] Y. Keller and A. Averbuch, A projection-based extension to phase correlation image alignment, Signal Process., vol. 87, pp. 124133, 2007.

[6] J. P. Lewis, Fast normalized cross-correlation, Vis. Interf., pp. 120123, 1995.

[7] A.L. Garcia, Probability and random processes for

electrical engineering, Singapore: Pearson Education, 2004.

[8] Tzimiropoulos, G. Argyriou, V. and Stathaki, T., "Subpixel Registration With Gradient Correlation," in Image Processing, IEEE Transactions on , vol.20, no.6, pp.1761-1767, June 2011

[9] J. G. Liu, H. Yan and G. Morgan "PCIAS subpixel technology", Meas. Control, vol. 45, no. 7, pp.207 -211 2012

[10] Xue Wan; Jianguo Liu; Hongshi Yan, "The Illumination Robustness of Phase Correlation for Image Alignment," in Geoscience and Remote Sensing, IEEE Transactions on , vol.53, no.10, pp.5746-5759, Oct. 2015

[11] Xiaohua Tong; Zhen Ye; Yusheng Xu; Shijie Liu; Lingyun Li; Huan Xie; Tianpeng Li, "A Novel Subpixel Phase Correlation Method Using Singular Value Decomposition and Unified Random Sample Consensus," in Geoscience and Remote Sensing, IEEE Transactions on , vol.53, no.8, pp.4143-4156, Aug. 2015

[12] Jinchang Ren; Jianmin Jiang; Vlachos, T., "High-Accuracy Sub-Pixel Motion Estimation From Noisy Images in Fourier Domain," in Image Processing, IEEE Transactions on , vol.19, no.5, pp.1379-1384, May 2010

[13] H. S. Stone, M. Orchard, E.-C. Chang, and S. Martucci, A fast direct fourier-based algorithm for subpixel registration of images, IEEE Trans. Geosci. Remote Sens., vol. 39, no. 10, pp. 22352243, Oct. 2001.

[14] H. Foroosh, J. Zerubia, and M. Berthod, Extension of phase correlation to sub-pixel registration, IEEE Trans. Image Process., vol. 11, no. 2, pp. 188200, Feb. 2002.

[15] W. S. Hoge, Subspace identification extension to the phase correlation method, IEEE Trans. Med. Imag., vol. 22, no. 2, pp. 277280, Feb. 2003.

[16] M. Balci and H. Foroosh, Subpixel estimation of shifts directly in the fourier domain, IEEE Trans. Image Process., vol. 15, no. 7, pp. 19651972, Jul. 2006.

[17] P. Vandewalle, S. Susstrunk, and M. Vetterli, A frequency domain approach to registration of aliased images with application to superresolution, EURASIP J. Appl. Signal Process., vol. 2006, pp. 114, 2006.

[18] M. Brand, Fast low-rank modifications of the thin singular value decomposition, Linear Algebra and Its Applications, vol. 415, no. 1, pp. 2030, 2006.

[19] J. Ren, T. Vlachos, and J. Jiang, Subspace extension to phase correlation approach for fast image registration, in Proc. IEEE ICIP, 2007, pp. 481484.

[20] K. Levenberg, A method for the solution of certain problems in leastsquares, Quarterly Appl. Math., vol. 2, pp. 164168, 1944.

[21] G. Tzimiropoulos, V. Argyriou, S. Zafeiriou, and T. Stathaki, Robust FFT-based scale-invariant image registration with image gradients, IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 10, pp. 18991906, Oct. 2010.

[22] L. G. Brown, A survey of image registration techniques, ACM Computing Surveys (CSUR), vol. 24, no. l4, p. 376, 1992.

[23] B. D. Lucas and T. Kanade, An iterative image registration technique with an application to stereo vision, in Int. Joint Conf. Artif. Intell., 1981, vol. 3, pp. 674679.