CHAPTER 4

4. FAST AND ROBUST FRAME ALIGNMENT

4.1 Introduction

Phase correlation is a well-established frequency domain method to estimate rigid 2-D translational motion between pairs of (frame) images. In this chapter, a novel variant of the phase correlation approach is proposed, in which 2-D translation is estimated by projection-based subspace phase correlation (SPC). Conventional wisdom has often suggested that such an approach can only amount to a compromise solution between accuracy and efficiency. In this work, however, it demonstrates the opposite by providing evidence that suggests that better performance can be achieved, especially in terms of subpixel accuracy as well as robustness to interference from secondary motion components and noise. Comprehensive results from synthetic data, MRI images and an image sequence have fully validated the proposed methodology. Due to its substantially lower computational complexity, the proposed method offers additional advantages in terms of efficiency and can lend itself to very fast implementations for a wide range of applications where speed is at a premium.

The remaining part of this chapter is organised as follows. Section 4.2 contains introductory concepts related to the phase correlation approach. In Section 4.3, subspace phase correlation is presented and its robustness is demonstrated. Section 4.4 discusses the treatment of non-overlapping regions, in which sub-image-based subspace phase correlation is introduced to enhance its robustness. Experimental results are given in Section 4.5 using synthetic data, MRI image data and standard image sequence data. Comparisons with existing techniques are also provided to confirm the superiority against proposed techniques. Finally, brief conclusions are drawn in Section 4.6.

4.2 Phase Correlation Method

The baseline method of phase correlation is based on the Fourier shift theorem, which states that a shift of two functions will lead to linear phase differences in (transformed) Fourier domain. Let $r(x,y)$ and $g(x,y)$ be two real-valued 2-D functions satisfying $r(x,y) = g(x+x_0, y+y_0)$ and absolutely integrable over $\Re^2$. Their corresponding Fourier transforms are denoted as $G(u,v)$ and $R(u,v)$. Then, we have

$$R(u,v) = G(u,v)e^{j2\pi(ux_0+vy_0)} \tag{4-1}$$

This can be re-written as

$$P(u,v) = \frac{R(u,v)G^*(u,v)}{\left|R(u,v)G^*(u,v)\right|} = e^{j2\pi(ux_0+vy_0)} \tag{4-2}$$

where * is the complex conjugate, $j = \sqrt{-1}$, and $P(u,v)$ is referred to as the cross power spectrum of the two signals.

If applying the inverse Fourier transform F^{-1} to $P(u,v)$, a phase correlation surface (PCS) can be obtained as given below, which is essentially a Dirac function centered at (x_0, y_0).

$$p(x,y) = F^{-1}\big(P(u,v)\big) = \delta(x - x_0, y - y_0) \tag{4-3}$$

If the two functions under consideration are not perfect replicas of each other then the surface is noisy but crucially still contains a dominant peak whose location (x_0, y_0) corresponds to the shift parameters and can be recovered as

$$(x_0, y_0) = \arg\max_{x,y} | \, p(x,y) \, | \tag{4-4}$$

The peak value can be substantially less than unity (the expected value) due to non-overlapping regions between the two signals. Peak height is an indication of confidence to the estimate obtained especially in the presence of secondary motion components which trigger lower peaks as well as noise which causes ripples on the correlation surface. To enhance the peak identification accuracy in the presence of such interferening features pre-processing in the shape of windowing or filtering is often considered. In this chapter, however, such pre-processing measures have not been considered in order to obtain cleaner and more straightforward comparisons with competing methods i.e. comparisons which are not conditional upon using a specific pre-processing regime. Details of the proposed algorithm are given in the next section.

4.3 Subspace Phase Correlation

Although the concepts of subspace and projection are not new in phase correlation, such as the work reported in [142] and [153], the essence of the proposed algorithm still is original, considering the fact that all existing work uses 2-D phase correlation to identify image displacement, yet the proposed algorithm uses only 1-D phase correlation to estimate 2-D offsets as explained below. The proposed method is shown to yield higher levels of accuracy and robustness and is not liable to the simple trade-off between accuracy and efficiency as reported in [160-162].

4.3.1 Subspace Phase Correlation

Let $r_x(x)$ and $g_x(x)$ denote respectively subspace projections of $r(x,y)$ and $g(x,y)$ onto x axis, and let their corresponding 1-D Fourier transforms be denoted as $R_x(u)$ and $G_x(u)$,

$$r_x(x) = \int_{y=-\infty}^{\infty} r(x,y)dy\,,\ R_x(u) = \int_{x=-\infty}^{\infty} r_x(x)e^{-j2\pi ux}dx \tag{4-5}$$

$$g_x(x) = \int_{y=-\infty}^{\infty} g(x,y)dy\,, G_x(u) = \int_{x=-\infty}^{\infty} g_x(x)e^{-j2\pi ux}dx \tag{4-6}$$

Then, we have

$$\begin{aligned} R_x(u) &= \int_{x=-\infty}^{\infty} [\int_{y=-\infty}^{\infty} r(x,y)dy]e^{-j2\pi ux}dx \\ &= \int_{x=-\infty}^{\infty}\int_{y=-\infty}^{\infty} r(x,y)e^{-j2\pi(ux+0\cdot y)}dxdy = R(u,0) \end{aligned} \tag{4-7}$$

$$G_x(u) = \int_{x=-\infty}^{\infty} [\int_{y=-\infty}^{\infty} g(x,y)dy]e^{-j2\pi ux}dx = G(u,0) \tag{4-8}$$

where $R(u,v)$ and $G(u,v)$ are the 2-D Fourier transforms of $r(x,y)$ and $g(x,y)$, respectively.

From (4-1), we have

$$R(u,0) = G(u,0)e^{j2\pi ux_0} \tag{4-9}$$

If substituting (4-6) and (4-7) in (4-9), we have

$$R_x(u) = G_x(u)e^{j2\pi ux_0} \tag{4-10}$$

which clearly shows that $R_x(u)$ and $G_x(u)$ are related by a phase shift that originates from $r_x(x) = g_x(x + x_0)$.

As a result, 1-D phase correlation can be used to estimate the shift in the x direction as shown below:

$$P_x(u) = \frac{R_x(u)G_x^*(u)}{\left|R_x(u)G_x^*(u)\right|} = e^{j2\pi ux_0} \tag{4-11}$$

$$p_x(x) = F^{-1}\left(P_x(u)\right) = \delta(x - x_0) \tag{4-12}$$

Similarly, y_0 can also be estimated from 1-D phase correlation using the 1-D Fourier transforms of projected signal to y direction as follows:

$$p_y(y) = F^{-1}\left(P_y(v)\right) = \delta(y - y_0) \tag{4-13}$$

$$P_y(v) = \frac{R_y(v)G_y^*(v)}{\left|R_y(v)G_y^*(v)\right|} = e^{j2\pi vy_0} \tag{4-14}$$

where $R_y(v)$ and $G_y(v)$ are respectively 1-D Fourier transforms of the projected signals, $r(x, y)$ and $g(x, y)$, onto the y axis.

4.3.2 Peak Height Analysis

If we take into account the expressions describing 2-D phase correlation (4-2) and (4-3) on the one hand and those describing the subspace variant (4-11) to (4-14), we can easily establish that:

$$P(u, v) = P_x(u)P_y(v) \tag{4-15}$$

$$p(x, y) = p_x(x)p_y(y) \tag{4-16}$$

Since all the PCS surfaces are upper bound to unity, i.e. $|\, p_x(x)\,| \leq 1$, $|\, p_y(y)\,| \leq 1$ and $|\, p(x, y)\,| \leq 1$, we have

$$|\, p_x(x)\,| = |\, p(x, y) / p_y(y)\,| \geq |\, p(x, y)\,| \tag{4-17}$$

$$|\, p_y(y)\,| = |\, p(x, y) / p_x(x)\,| \geq |\, p(x, y)\,| \tag{4-18}$$

The above is a simple illustration of the fact that subspace phase correlation tends to yield higher peaks than those obtained from 2-D phase correlation.

4.3.3 Robustness Analysis

If there is additive zero-mean random noise $n(x, y)$ in the shifted signal (due to quantitative error for example), i.e.

$$r(x,y) = g(x+x_0, y+y_0) + n(x,y) \tag{4-19}$$

Then, for 2-D phase correlation, the cross correlation function (4-2) becomes:

$$P(u,v) = \frac{[G(u,v)e^{j2\pi(ux_0+vy_0)} + N(u,v)]G^*(u,v)}{\left|[G(u,v)e^{j2\pi(ux_0+vy_0)} + N(u,v)]G^*(u,v)\right|} \tag{4-20}$$

Let $F(u,v) = G^*(u,v)N(u,v)$, we have

$$P(u,v) = e^{j2\pi(ux_0+vy_0)} + \frac{F(u,v) - |F(u,v)|e^{j2\pi(ux_0+vy_0)}}{G^*(u,v)G(u,v) + |F(u,v)|} \tag{4-21}$$

Since $F(u,v) \neq 0$, the PCS function of the inverse Fourier transform of $P(u,v)$ is no longer a simple delta function which inevitably causes inaccuracy.

For the proposed subspace phase correlation, however, it can be demonstrated that the PCS functions are more robust to such noise models. Without loss of generality, the range of variables in the discrete domain are denoted as $x \in [0, M-1]$ and $y \in [0, N-1]$. From (4-5) and (4-19), we have that

$$\begin{aligned} r_x(x) &= \sum_{y=0}^{N-1}[g((x+x_0)\otimes M, (y+y_0)\otimes N) + n(x,y)] \\ &= \sum_{y=0}^{N-1} g((x+x_0)\otimes M, (y+y_0)\otimes N) + N[\frac{1}{N}\sum_{y=0}^{N-1} n(x,y)] \\ &= g_x(x+x_0) + N \cdot E(n(x,y)) \\ &= g_x(x+x_0) \end{aligned} \tag{4-22}$$

where $\otimes$ is the modulo operator, and $E(n(x,y))$ refers to the average value (mean) of all contributing samples of $n(x,y)$.

When N is large enough, considering that $n(x,y)$ is zero-mean noise, the following approximation is approached:

$$E(n(x,y)) = \frac{1}{N}\sum_{y=0}^{N-1} n(x,y) = 0 \tag{4-23}$$

This indicates that the projected signal virtually eliminates the influence of the random noise component which renders the subspace phase correlation more robust in relation to conventional 2-D phase correlation.

4.4 Implementation and Subpixel Accuracy

There exist two further issues that need to be addressed: (i) dealing with non-overlapped areas between image pairs and (ii) obtaining subpixel accuracy.

4.4.1 Dealing with Non-overlapped Areas

Following [142], there are overlapped and non-overlapped areas in two images under registration (as illustrated in Figure 4.1). If $r_0(x,y)$ and $g_0(x,y)$ are denoted as the original two images (without non-overlapped areas) which satisfy $r_0(x,y) = g_0(x+x_0, y+y_0)$, and if the corresponding images with non-overlapped areas are $r(x,y)$ and $g(x,y)$, then their relationship can be established as

$$g(x,y) = g_0(x,y) - N_g(x,y) \tag{4-24}$$

$$\begin{aligned} r(x,y) &= r_0(x,y) - N_r(x,y) \\ &= g_0(x+x_0, y+y_0) - N_r(x,y) \end{aligned} \tag{4-25}$$

where $N_g(x,y)$ and $N_r(x,y)$ represent corresponding non-overlapped areas in $g(x,y)$ and $r(x,y)$, respectively.

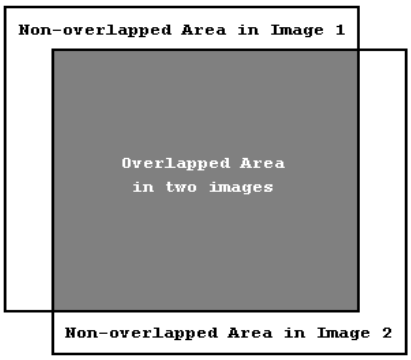


Figure 4.1. Overlapped and non-overlapped areas in two images.

Due to these non-overlapped areas, the relationship between the two images is no longer a simple shift, i.e. $r(x,y) \neq g(x+x_0, y+y_0)$. However, $N_g(x,y)$ and $N_r(x,y)$ can be considered as noise attached to the two images, and normally the mean of this noise is non-zero due to the arbitrary nature of the image contents in the non-overlapped areas. As a result, a refinement is needed to make the subspace approach robust in this case.

The refinement consists of considering gradient-based information as follows. Let $g_x(x)$ be the horizontal projection of $g(x,y)$ (see in Section 4.3, eqs. 4-4 to 4.6), and its gradient signal is defined as follows:

$$g_x^{(1)}(x) = g_x(x+1) - g_x(x) \tag{4-26}$$

If we substitute (4-6) and (4-24) in (4-26), we have

$$
\begin{aligned}
g_x^{(1)}(x) &= [\sum_{y=0}^{N-1} g((x+1)\otimes M, y)] - [\sum_{y=0}^{N-1} g(x,y)] \\
&= \sum_{y=0}^{N-1} [g_0((x+1)\otimes M, y) - g_0(x,y)] - \\
&N[\frac{1}{N}\sum_{y=0}^{N-1} N_g((x+1)\otimes M, y) - \frac{1}{N}\sum_{y=0}^{N-1} N_g(x,y)] \\
&= g_0^{(1)}(x)
\end{aligned}
\tag{4-27}
$$

where $\otimes$ is again the modulo operator, and $g_0^{(1)}(x)$ is the local gradient of the horizontal projection of $g_0(x,y)$. Here the following assumption is utilized:

$$
\sum_{y=0}^{N-1} N_g((x+1)\otimes M, y) = \sum_{y=0}^{N-1} N_g(x,y) = N\mu_g^{(y)} \tag{4-28}
$$

where $\mu_g^{(y)}$ denotes the mean of $N_g(x,y)$ in y-direction, i.e. the mean value is estimated by the average of all the contributing samples. Note than $N_g(x,y)$ will not be the same in both directions due to the fact that the non-overlapping material in two directions is different.

Similarly, it can be proved that the horizontal projection of $r(x,y)$, denoted as $r_x(x)$, satisfy the following

$$
r_x^{(1)} = r_0^{(1)}(x) = g_0^{(1)}(x + x_0) \tag{4-29}
$$

where $r_x^{(1)}$ is local gradient of $r_x(x)$, and $r_0^{(1)}(x)$ is the local gradient of the horizontal projection of $r_0(x,y)$.

From (4-27) and (4-29) it can be seen that that registration of $r(x,y)$ and $g(x,y)$ (using local gradients) essentially amounts to the registration of $r_0(x,y)$ and $g_0(x,y)$

(also using local gradients), and that the latter corresponds to an ideal shift case i.e. without overlapping regions. Consequently, subspace phase correlation using local gradients can overcome the problems associated with overlapping.

In the same way, offset in y-direction can also be estimated by applying the same treatment in the vertical direction. With these estimated offsets in two directions, a 2-D motion vector can be composed and its subpixel accuracy is further determined as discussed next.

4.4.2 Subpixel Accuracy

In 2-D phase correlation, the most reliable displacement (motion) vector is determined as the location of the highest peak on the 2-D PCS of $p(x, y)$. In subspace phase correlation, displacement (motion) components in the horizontal and vertical directions are determined separately from two 1-D PCSs namely $p_x(x)$ and $p_y(y)$, respectively. It is obvious that both $p_x(x)$ and $p_y(y)$ are real. If there is only one dominant motion in the two signals, $p_x(x_0)$ and $p_y(y_0)$ can be easily located as the only peak present in each of $p_x(x)$ and $p_y(y)$. If we denote (x_0, y_0) as the integer estimate of 2-D displacement, a subpixel offset (x'_0, y'_0) can be determined as follows.

Let $p_x(x_0)$ be the highest peak in $p_x(x)$ and its two immediate neighbors on either side are $p_x(x_0 - 1)$ and $p_x(x_0 + 1)$. Then, x'_0 is estimated either between x_0 and $x_0 - 1$ or between x_0 and $x_0 + 1$ depending upon whether $p_x(x_0 - 1) > p_x(x_0 + 1)$ or not.

Similarly, y'_0 is estimated either between y_0 and $y_0 - 1$ or between y_0 and $y_0 + 1$ depending upon whether $p_y(y_0 - 1) > p_y(y_0 + 1)$ or not.

If we define $p_m(x_0)$ and $p_m(y_0)$ as given below,

$$\begin{aligned} p_m(x_0) &= p_x(x_0 - 1) - p_x(x_0 + 1) \\ p_m(y_0) &= p_y(y_0 - 1) - p_y(y_0 + 1) \end{aligned} \tag{4-30}$$

The following linear interpolation can be used to estimate the subpixel displacements:

$$x'_0 = \begin{cases} \dfrac{x_0 p_x(x_0) + (x_0 - 1) p_m(x_0)}{p_x(x_0) + p_m(x_0)} & IF \quad p_m(x_0) > 0 \\ x_0, \quad IF \quad p_m(x_0) = 0 \\ \dfrac{x_0 p_x(x_0) + (x_0 + 1)[-p_m(x_0)]}{p_x(x_0) + [-p_m(x_0)]}, otherwise \end{cases} \tag{4-31}$$

$$y'_0 = \begin{cases} \dfrac{y_0 p_y(y_0) + (y_0 - 1) p_m(y_0)}{p_y(y_0) + p_m(y_0)} & IF \quad p_m(y_0) > 0 \\ y_0 \quad IF \quad p_m(y_0) = 0 \\ \dfrac{y_0 p_y(y_0) + (y_0 + 1)[-p_m(y_0)]}{p_y(y_0) + [-p_m(y_0)]}, otherwise \end{cases} \tag{4-32}$$

Please note that, if the heights of two side lobes are very close to each other, i.e. $p_m(x_0) \approx 0$ or $p_m(y_0) \approx 0$, the initial integer offset x_0 or y_0 will be main surviving value out of the interpolation operation above. Besides, if $p_m(x_0) < 0$ or $p_m(y_0) < 0$ then $-p_m(x_0)$ or $-p_m(y_0)$ is then used for interpolation in (4-31) and (4-32).

It is worth mentioning that the subpixel scheme described above is different from the ones proposed in [142] and [163]. While their techniques use only the main peak and one selected side lobe to estimate a subpixel offset, the proposed solution, on the

contrary, utilizes the two side lobes as it is believed that this can help to compensate energy leaking on the correlation surface leading to more robust estimation. The performances of these two solutions are compared in details in Section 4.5.

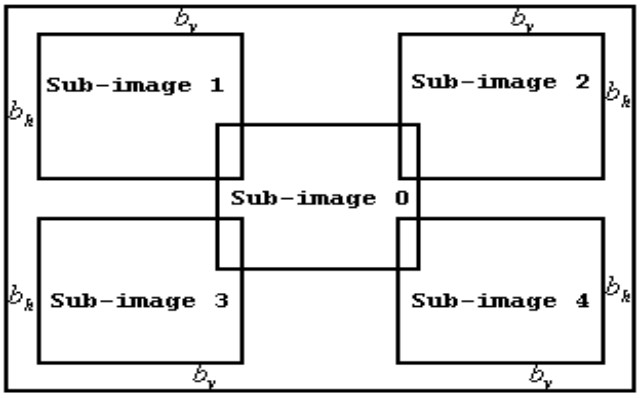


Figure 4.2. Definition of five sub-images in each test image.

4.4.3 Further Improving Robustness

To further improve the robustness of the proposed approach, five sub-images are defined in each of two images undergoing registration. The sub-image partition scheme is shown in Fig. 4.2. The five sub-images are of the same size determined as follows. Let the size of the original image and the sub-image be (M,N) and (m,n), respectively, where $m=2^{m_0}$ and $n=2^{n_0}$ to cope with the requirement from fast Fourier transform. Then, (m,n) should satisfy

$$2(2^{m_0}+b_h) \le M < 2(2^{m_0+1}+b_h)$$

$$2(2^{n_0}+b_v) \le N < 2(2^{n_0+1}+b_v)$$

where b_h and b_v refer to columns or rows of boundary pixels as shown in Fig. 4.2, and sub-image 0 is the central sub-image. In general b_h and b_v should be slightly larger than the offsets in horizontal and vertical directions, and $b_h = b_v = 5$ is used in the proposed system.

For each pair of images under registration, five groups of results can be generated from the corresponding pair of sub-images. The final estimate is calculated as the median value of the corresponding set of five estimates. This is applied independently in the horizontal and vertical directions.

4.5 Results and Discussion

The performance of proposed method was determined by using both synthetic and real data. For synthetic data, subpixel shifts of images are generated by linear interpolation of the original images, respectively. As for real data, a set of MRI images of subpixel displacements are employed. Further, results using video sequence data are also presented.

Using ground truth (real shifts) as a reference, an error vector between a real shift and the corresponding estimate is obtained for each method along the x and y directions. Let δ_x and δ_y denote the corresponding error vectors, i.e. $\delta_x(i) = x(i) - \hat{x}(i)$ and $\delta_y(i) = y(i) - \hat{y}(i)$, where $x(i)$ and $y(i)$ are the i^{th} real offsets, and $\hat{x}(i)$ and $\hat{y}(i)$ are their

estimates. The mean μ and standard deviation σ of error vectors are used to measure accuracy. Additionally, the mean square error (MSE) between the estimates and the ground truth is used as an overall measurement and this is consistent with the work reported in [154]. Details on these tests and results are compared and discussed below.

4.5.1 Synthetic Data

The two original images illustrated in Figure 4.3, are manually shifted by linear interpolation to obtain subpixel displacements of the same values as those given in [149] and [150] (see Table 4.1). It should be noted that this approach is different from that followed in [149] and [150], in which subpixel shifts are obtained by lowpass-filtering and down-sampling of a real high resolution.

(a) "Paris" (b) "Pentagon"

Figure 4.3. Two test images used to generate subpixel shifts.

In Figure 4.3, two original test images, "Paris" and "Pentagon" are shown. These are the same as those used in [149] and [150]. These are 8bpp images and their sizes are

200×200 and 512×512 pixels, respectively. The generated subpixel offsets and the registration results are presented in Table I, in which the achieved results are compared with those obtained from Hoge [142], Foroosh [149] and Stone [151]. All three competing approaches use 2-D phase correlation and are compared against the proposed gradient-based subspace phase correlation as well as the 2-D phase correlation with the improved subpixel accuracy scheme. The size of each sub-image is "64×64" for "Paris" and "128×128" for "Pentagon" during registration in both 1-D and 2-D phase correlation using the proposed method.

From the results listed in Table 4.1, the following observations can be made. Firstly, Stone's approach yields the worst results in this test in terms of absolute mean value of registration error, standard deviation and MSE, followed by Hoge's approach. Results using 2-D phase correlation from Foroosh's approach are slightly worse than those using 2-D phase correlation, and this proves the effectiveness of the proposed subpixel solution. Finally, gradient-based subspace (1-D) phase correlation produces the best results in the comparison set.

While the results in Table 4.1 seem to suggest that the results of the proposed method are slightly worse than those reported in [150], the comparison is not fair due to the fact that the method in [150] is not simulated as it requires not only 2-D phase correlation but also the expectation maximization (EM) for iterative optimization.

However, the method in [150] is further compared with ours using real MRI images and the results are discussed in the next section.

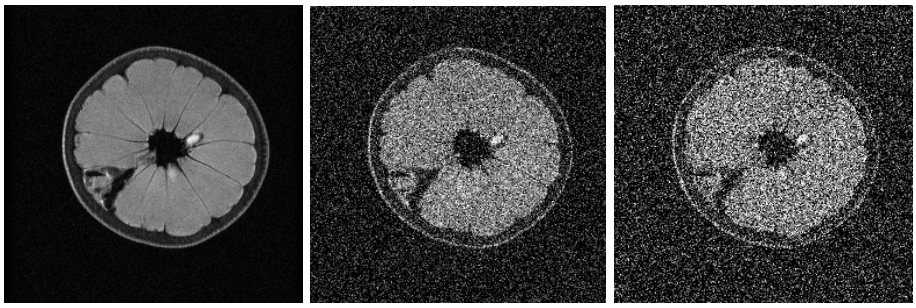
Table 4.1. Table of results for shifts of the images in Fig. 4.3 using linear interpolation.

Image Pairs	True Shifts (x,y)	Hoge [142]	Foroosh et al [149]	Stone et al [151]	Proposed method	
					1-D	2-D
Paris	(0.50,-0.50)	(0.228,-0.241)	(0.438,-0.315)	(0.833,-1.39)	(0.558,-0.535)	(0.450,-0.470)
	(0.25,0.50)	(0.088,0.232)	(0.156,0.455)	(0.140,0.014)	(0.256,0.471)	(0.235,0.493)
	(-0.25,-0.50)	(-0.087,-0.237)	(-0.117,-0.357)	(-0.099,-1.372)	(-0.256,-0.543)	(-0.234,-0.497)
	(0.00,0.75)	(-0.001,0.897)	(0.041,0.884)	(-0.248,0.452)	(0.016,0.736)	(-0.089,0.690)
Pentagon	(0.167,-0.50)	(0.059,-0.255)	(0.103,-0.426)	(0.294,-0.789)	(0.174,-0.547)	(0.155,-0.545)
	(0.67,0.25)	(0.863,0.094)	(0.732,0.158)	(1.703,-0.203)	(0.652,0.252)	(0.641,0.255)
	(-0.33,-0.167)	(-0.134,-0.059)	(-0.274,-0.103)	(-0.062,0.113)	(-0.340,-0.174)	(-0.336,-0.166)
	(0.33,0.33)	(0.135,0.135)	(0.235,0.237)	(-0.109,0.023)	(0.341,0.341)	(0.338,0.338)
Error analysis	**Mean** (x,y)	(0.023,-0.050)	(0.003,-0.046)	(-0.139,0.414)	(-0.008,0.02)	(0.02,0.008)
	Std-dev(x,y)	(0.188,0.222)	(0.086,0.110)	(0.447,0.371)	(0.023,0.02)	(0.034,0.03)
	MSE (x,y)	(0.032,0.046)	(0.007,0.013)	(0.194,0.292)	(0.0005,0.0008)	(0.001,0.0008)

4.5.2 Real MRI Data

The MRI data set used in the experiments is courtesy of W. S. Hoge and contains five MRI images of a grapefruit (size of 256×256 in 8-bit grey format) [142]. The true shifts between each pair of images are known and subsequently used as ground truth for performance evaluation. The first MRI image is shown in Figure 4.4(a), along with two other images obtained by manually adding Gaussian noise. In this group of experiments, subpixel registration performance comparisons are carried out using noise-free original images. The proposed method is compared against the techniques of Balci and Foroosh [150], Hoge [142], and Foroosh et al [149] and tabulate the results in Table 4.2. In addition, results from Balci and Foroosh's work in [150] are directly

quoted. Since in this group of test images there doesn't exist non-overlapped areas, the whole image is used for registration (rather than the five sub-image scheme).



(a) (b) (c)

Figure 4.4. Three examples of test images: (a) original MRI image (Courtesy of W. S. Hoge); (b) and (c) are two noisy versions of (a) with additive Gaussian noise.

In Table 4.2, the mean, standard deviation and range errors in x and y directions are given for comparison purposes. It can be clearly seen that the overall accuracy along the y-axis is better than that along x-axis, which is possibly due to the difference in generating displacements in different directions (see [142] for details). In both x and y directions, the approach from [150] generates maximum error values, followed by the methods from [149] and [142].

The overall-results from the proposed two approaches i.e. using subspace phase correlation and 2-D phase correlation are quite similar. Again, this clearly demonstrates that subpixel performance is more robust than that of [142], even when using the same 2-D phase correlation. Considering the MSE-based evaluation, the proposed approach achieves the minimum error in the comparison set. In addition, it is

interesting to point out that the gradient-based subspace phase correlation produces slightly worse results than that of 2-D approach using the proposed subpixel solution in this group of experiments. This is because that there exist no non-overlapped areas as stated earlier.

Table 4.2. Pairwise registration results of the five MRI images.

Image Pairs	Physical offset (x,y)	Foroosh et al [149]	Hoge [142]	Balci & Foroosh [150]	Proposed method	
					1-D	2-D
(1,2)	(-2.40,-4.00)	(-2.23,-4.23)	(-2.03,-4.01)	(-2.11,-4.00)	(-2.07,-4.10)	(-2.09,-4.02)
(1,3)	(-4.80,-8.00)	(-4.07,-8.24)	(-4.13,-8.01)	(-3.90,-7.49)	(-4.34,-8.05)	(-4.34,-8.01)
(1,4)	(-7.20,-4.32)	(-6.59,-4.41)	(-6.81,-4.17)	(-6.22,-3.93)	(-6.60,-4.34)	(-6.58,-4.38)
(1,5)	(-7.20,-12.00)	(-6.59,-12.26)	(-6.82,-12.02)	(-6.39,-11.42)	(-6.58,-12.15)	(-6.59,-12.08)
(2,3)	(-2.40,-4.00)	(-2.10,-3.60)	(-2.10,-3.99)	(-2.18,-3.87)	(-2.26,-3.86)	(-2.27,-3.96)
(2,4)	(-4.80,-0.32)	(-4.55,-0.39)	(-4.28,-0.15)	(-4.16,-0.31)	(-4.52,-0.30)	(-4.54,-0.356)
(2,5)	(-4.80,-8.00)	(-4.55,-8.24)	(-4.78,-8.00)	(-4.13,-7.73)	(-4.55,-7.92)	(-4.54,-8.01)
(3,4)	(-2.40,3.68)	(-2.00,3.61)	(-2.17,3.84)	(-2.34,3.55)	(-2.42,3.62)	(-2.40,3.65)
(3,5)	(-2.40,-4.00)	(-2.00,-3.56)	(-2.18,-4.51)	(-2.49,-3.83)	(-2.42,-4.07)	(-2.41,-4.00)
(4,5)	(0.00,-7.68)	(-0.26,-7.92)	(0.01,-7.85)	(-0.03,-7.84)	(0.02,-7.66)	(0.02,-7.64)
Error analysis	**Mean** (x,y)	(-0.346,0.061)	(-0.309,0.022)	(-0.445,-0.177)	(-0.266,0.019)	(-0.266,0.017)
	Std-dev(x,y)	(0.281,0.265)	(0.207,0.200)	(0.401,0.257)	(0.240,0.086)	(0.238,0.038)
	MSE (x,y)	(0.190,0.067)	(0.134,0.037)	(0.343,0.091)	(0.123,0.007)	(0.122,0.002)

Table 4.3. Pairwise registration results of the MRI images using subspace phase correlation without the use of local gradient.

Image pairs	Estimated results	Image pairs	Estimated results
(1,2)	**(-2.07,-4.10)**	**(2,4)**	**(-4.52,-0.30)**
(1,3)	**(-4.34,-8.05)**	**(2,5)**	**(-4.55,-7.92)**
(1,4)	**(-6.60,-4.34)**	**(3,4)**	**(-2.42,3.62)**
(1,5)	**(-6.58,-12.14)**	**(3,5)**	**(-2.42,-4.06)**
(2,3)	**(-2.26,-3.86)**	**(4,5)**	**(0.02,-7.66)**

Considering the above fact subspace phase correlation method is further tested below on projected signals without the use of local gradient information and tabulate the results in Table 4.3. For synthetic data, this will normally lead to performance

compromise. However, for the MRI data the results in Table 4.3 are similar to those in Table 4.2 using local gradient information, which shows that the corresponding displacements can still be successfully estimated.

4.5.3 Robustness Analysis

To evaluate the robustness, synthetic zero-mean Gaussian noise is added to the test images. Before adding noise, the intensity level of the original images is normalized within $[0,1]$. Then, zero-mean Gaussian noise is generated with its variance changing linearly in eight levels within the interval $[0.005,0.04]$, hence eight noisy samples are obtained for each of the five original images. Example images with additive noise are shown in Figure 4.4(b) and 4.4(c), where the variances of Gaussian distribution were 0.02 and 0.04, respectively. For the original image in Fig 4.4(a), the noise level of its 8 noisy samples are further measured in terms of the signal to noise ratio (SNR), which are 29.2, 22.8, 18.9, 16.2, 14.0, 12.3, 10.9 and 9.5 DB, respectively.

Again, pair-wise registration is performed, thus totally 80 pairs of noisy images are used for 8 different noise levels (Gaussian variance values). For each noise level, all the estimates from 10 pairs of images are measured using the MSE criterion. Hoge's method was found to fail in most of the noisy cases. Results obtained from the proposed approach are compared with those from Foroosh et al [149] as shown in Figure 4.5.

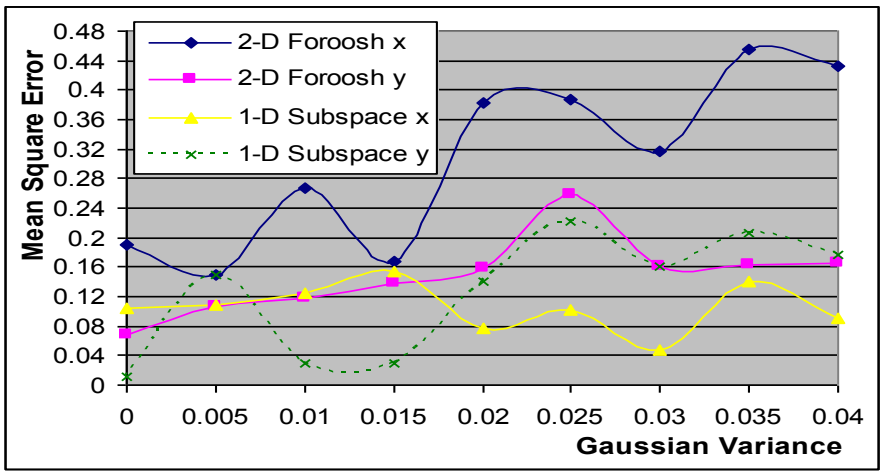


Figure 4.5. Mean square errors (y-axis) vs. Gaussian variance (x-axis).

As seen from Fig. 4.5, in general, Foroosh's method generates higher MSE, though there are some exceptions along the y-axis. Along the x-axis, the subpixel scheme using subspace phase correlation usually yields consistently lower levels of MSE than the 2-D approach. However, along the y-axis, subspace correlation occasionally produces higher MSE than the 2-D approach.

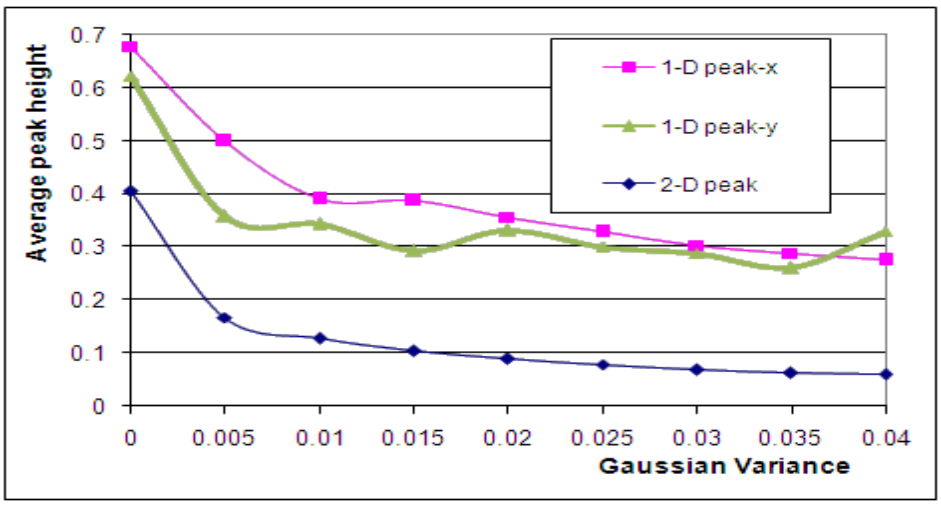


Figure 4.6. Average height of the most dominant peak (y-axis) vs. Gaussian variance (x-axis).

The height of the most dominant peak can be considered as an indication of robustness. Figure 4.6 shows the average height of the most dominant peak as a function of increasing Gaussian variance for 2-D correlation and subspace correlation (along the x- and y- axes). With increasing Gaussian variance, the average height from 2-D or subspace correlation decreases subsequently. However, it is obvious that subspace correlation generates much higher average peaks than 2-D correlation.

4.5.4 Moving Sequence Data

In this group of experiment, the well-known "flower garden" sequence is used. Only the luminance component (8bpp) was used and 115 frames of 320×240 pixels each were processed. There is apparent camera panning in the sequence and this causes a dominant global motion component and a resulting large frame difference (see Fig. 4.7). Using the phase correlation approach, the camera shift can be estimated as a global motion vector. These estimated global motion vectors are subsequently used to motion compensate the sequence and the motion compensated errors are then used as a measure for comparisons. In particular, the sums of absolute frame difference before and after global motion compensation are compared against the original frame by using the signal-to-noise ratio as defined below:

$$SNR(\Delta f_{1,2}, f_2) = -20\log\frac{\sum_{i,j} |\Delta f_{1,2}(i,j)|}{\sum_{i,j} |f_2(i,j)|} \tag{4-33}$$

where f_1, f_2 are two frames and $\Delta f_{1,2}$ refers to their difference (with or without global motion compensation).

Figure 4.7. Original two images from "flower garden" sequence at frame #20 and #21 and their raw difference (top row) and three motion-compensated results (the bottom row), estimated by, from left to right, using Foroosh [149], proposed 2-D and subspace phase correlation, respectively. All the difference images are enhanced for better visualization (brightness +10% and contrast -10%).

In total, five sub-images are extracted for each frame during registration. The gradient-based subspace phase correlation, the 2-D phase correlation method using the proposed subpixel strategy and Foroosh's method are compared. Before motion compensation, the average (computed over all processed frames) SNR of the raw frame-difference signal is 23.76dB. After motion compensation, the average SNR becomes 32.93, 32.29 and 31.79dB from proposed subspace correlation and 2-D correlation as well as Foroosh's method, respectively. Considering 2-D phase correlation, an improvement of 0.5dB is offered. However, an additional 0.64 dB is obtained by using the proposed

subspace phase correlation. This confirms that subspace phase correlation combined with the improved subpixel strategy is the approach that yields enhanced performance.

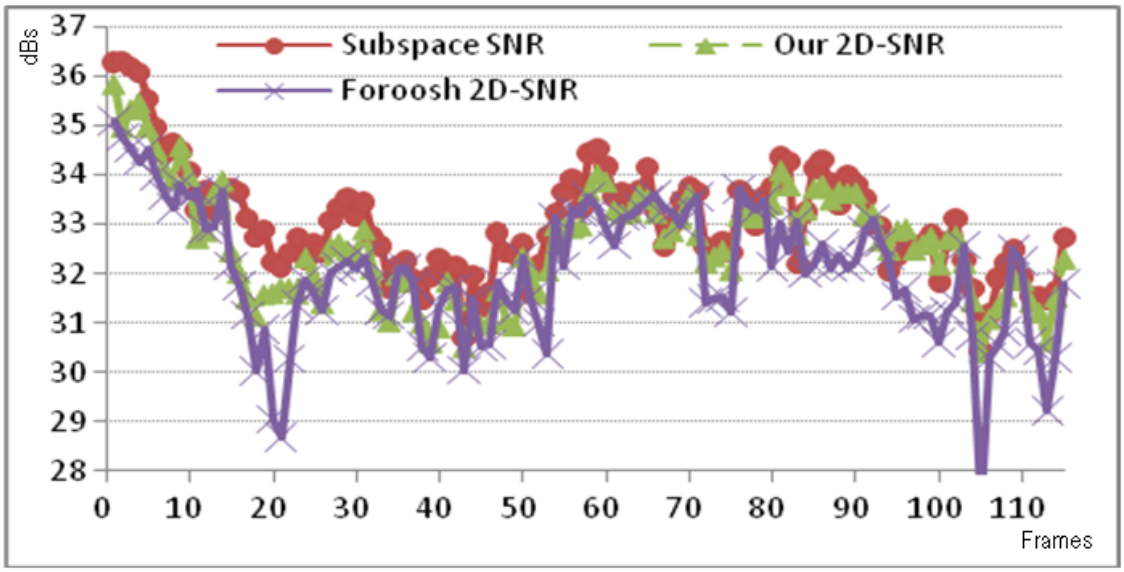


Figure 4.8. Performance in terms of SNR (in dBs) vs. frame number using global motion compensation for the "flower garden" sequence.

The performances of the three methods using motion compensation are also shown in Fig. 4.7. It should be noted that in this sequence global motion conforms to a projective model while compensation is carried out using a translational model. As a consequence the quality of compensation is not very high as can be seen from the difference images. Nevertheless, parts of the background furthest to the imaging plane have been accurately estimated and the resulting error is low. It is obvious that subspace phase correlation yields the best performance with the lowest compensation error, followed by 2-D method using proposed subpixel solution and Foroosh's phase correlation method. A pictorial manifestation of this comparison is shown in Fig. 4.8.

4.5.5 Computational Complexity

In both 2-D phase correlation and subspace correlation, the fast Fourier transform (FFT) is the main computational load. In some approaches, additional processing is required such as windowing, partial differencing or even singular value decomposition [142] and iterative optimization [150]. If the original images are of $N \times N$, then the computing complexity of the FFT in 2-D and subspace correlation is $O(N^2 \log_2 N)$ and $O(N \log_2 N)$, respectively. In summary, the subspace scheme is of substantially lower complexity and consequently faster to implement. In addition, all the components in the proposed approach, including the FFT, subspace projection, local gradient and linear interpolation are suitable for hardware implementation to further improve the efficiency.

4.6 Summary

In this chapter, a novel extension to the phase correlation image registration approach has been described in this chapter. The subspace phase correlation variant was presented and proved to be not only more efficient but also more effective and robust than conventional 2-D phase correlation formulations. The robustness to both zero-mean noise and non-zero-mean noise has been proved theoretically and empirically. Furthermore, the proposed subpixel accuracy enhancement scheme is also effective in yielding more accurate results from a variety of test data including synthetic, MRI and

moving sequence data. Finally, the fact that the proposed algorithm is suitable for hardware implementation makes it a good candidate for a wide variety of applications such as frame alignment, online registration and camera stabilization.