

Robust Phase-Correlation Based Registration of Airborne Videos Using Motion Estimation

Frank de Morsier, Maurice Borgeaud, Christoph Küchler,
Adrian Vogel, Volker Gass and Jean-Philippe Thiran

Abstract This paper presents an algorithm for near-real time registration of airborne video sequences with reference images from a different sensor type. Phase-correlation using Fourier-Melin Invariant (FMI) descriptors allow to retrieve the rigid transformation parameters in a fast and non-iterative way. The robustness to multi-sources images is obtained by an enhanced image representation based on the gradient norm and by the extrapolation of registration parameters by motion estimation between frames. A phase-correlation score, indicator of the registration quality, is introduced to regulate between frame-to-reference image registration and extrapolation from previous frames only. Our Robust Phase-Correlation based registration algorithm using Motion Estimation (RPCME) is compared with a Mutual Information (MI) algorithm for the registration of two different panchromatic airborne videos with Geoeye reference images. The RPCME algorithm registered most of the frames accurately, retrieving much better orientation than MI. It shows robustness and good accuracy to multisource images with the advantage of being a direct (non-iterative) method.

F. de Morsier (✉) · J.-P. Thiran
École Polytechnique Fédérale de Lausanne, LTS 5, Lausanne, Switzerland
e-mail: frank.demorsier@epfl.ch

M. Borgeaud
European Space Agency, ESRIN, Frascati, Italy

C. Küchler · A. Vogel
RUAG Schweiz AG, Emmen, Switzerland

V. Gass
Swiss Space Center, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

1 Introduction

In the management of natural hazards, the quick assessment of the damaged areas is essential to help the coordination of the rescue teams. An airborne vehicle (like Unmanned Aerial Vehicle) equipped with an optical sensor could be launched quickly after the disaster and provide up-to-date images allowing a comparison with the reference images available from any source (space or airborne sensors). The analysis and fusion of remote sensing images requires an accurate registration, i.e. finding the parameters of a transformation that matches each pixel of the images to the physical reality. Thus, the comparison between airborne video frames and reference images requires a registration robust to the differences in sensor sensitivity, weather and light conditions, lens distortions, view angles as well as temporal changes on the ground. A near-real time registration onboard could allow for live damage detection and let the airplane dynamically focus on damaged areas.

The general problem of registration has been extensively investigated in the literature for remote sensing (e.g. fusion of multi-temporal or multimodal images), medical imaging (e.g. fusion of CT and MRI images) and computer vision (Zitova and Flusser 2003). Two main categories of registration can be separated: affine registration (preserving straight lines) and non-affine registration (where local distortions are allowed). In this wide range of methods, focus has been set on affine registration methods which are potentially faster for near-real time scenarios and accurate enough once perspective and lens distortions are corrected.

In the particular case of airborne video registration, the registration of a frame with a reference image can be helped by the previous frames. The motion between consecutive frames can be easily retrieved using standard registration techniques, since they are extremely redundant. In Cannata et al. (2000), a Kalman filter allows for an online registration based on a certain number of previous frames. In Shastry et al. (2005), airborne video registration for traffic-flow monitoring is done using fixed features and helicopter movements are removed from inter-frame registration. In Wu and Luo (2008), a prediction model of the camera movement allows to remove outliers in feature points matching by taking advantage of the relationships among consecutive video frames. A global optimization problem involving frame-to-reference and frame-to-frame constraints can be minimized to get the registration parameters (Kumar et al. 2000, Hirvonen et al. 2001).

However the appearance between the frame and the reference image is often dissimilar from temporal changes and different sensor sensitivity. A more robust representation of the images is needed in order to see these methods working with images from different sources. In Hirvonen et al. (2001), Oriented energy image pyramids are used to register video frames having a different appearance than the reference image. However in this case the registration is obtained by solving a minimization problem which does not guarantee the method to be fast.

This paper motivated by the mentioned limitations aims at presenting a fast registration method for airborne video platforms having a potential for near-real

time registration and being robust to video having strong differences with the available reference image. The phase-correlation (PC), a non-iterative method, allows to retrieve the affine rigid transformation parameters between two images (rotation, scaling and translation) with low computational complexity (Kuglin and Hines 1975). The PC is sensitive to images differences, therefore robustness to changes and multi-sources images is obtained by an enhanced image representation and by the extrapolation of registration parameters using the inter-frame motion.

2 Methodology

This section introduces the methodology for the robust registration of airborne video frames with a reference image having a different appearance. Two different pixel-based approaches for registering images with a subpixel accuracy are presented: Mutual Information and Phase-Correlation. Both are evaluated and compared for the registration of different airborne videos in Sect. 3.

2.1 Mutual Information Registration

The Mutual Information registration algorithm (MI) is a standard registration approach based on image intensities and not on local image features. This registration method is introduced here for comparisons with our proposed algorithm based on phase-correlation (see Sect. 2.2). Mutual information measures the amount of information that one variable contains about the other, i.e. it expresses the probability that the pixels are reflecting the same physical reality in both images. It is really often used in multi-source/multimodal registration because handling well non-linear transformations (no *a priori* on intensity relationship). Therefore we selected this pixel-based method for comparison with our phase-correlation based algorithm. Analogous to the Kullback–Leibler expression, the mutual information is defined as follows:

$$MI = \sum_{I_1, I_2} p(I_1, I_2) \log \frac{p(I_1, I_2)}{p(I_1)p(I_2)} \quad (1)$$

with I_2' being the image after affine transformations, $p(I_1)$ the marginal distribution of I_1 and $p(I_1, I_2')$ the joint probability distribution. These probabilities can be computed from their normalized joint histogram. The registration is based on maximizing the mutual information between the image I_1 and the transformed image I_2' . The implementation has been realized using the HPV interpolation and the Powell's Direction set method, for more details see Lu et al. (2008).

2.2 Phase-Correlation Registration

The phase-correlation (PC) exploits the Fourier shift properties. An image $I_1(x, y)$ translated of (x_0, y_0) becomes $I_2(x, y) = I_1(x - x_0; y - y_0)$ and can be expressed in the Fourier domain as: $\hat{I}_2(\eta, \xi) = e^{-j2\pi(\eta x_0 + \xi y_0)} \cdot \hat{I}_1(\eta, \xi)$. The normalized cross power spectrum $R(\eta, \xi)$ between these two images is a phase change in the Fourier domain.

$$R(\eta, \xi) = \frac{\hat{I}_1 \cdot \hat{I}_2^*}{|\hat{I}_1 \cdot \hat{I}_2^*|} = e^{-j2\pi(\eta x_0 + \xi y_0)} \quad (2)$$

with \hat{I}^* the complex conjugate of \hat{I} .

The phase-correlation $PC(x, y)$, being the inverse Fourier transform of $R(\eta; \xi)$, is ideally a delta function shifted from the center of (x_0, y_0) : $PC(x, y) = \hat{R}(x, y) = \delta(x + x_0, y + y_0)$. Therefore the localization of the PC maximum peak yields to the horizontal and vertical shifts existing between the two images.

Using the same properties, one can retrieve rotation and scaling differences using the Fourier-Melin Invariant descriptors (FMI) of the images. The FMI of an image corresponds to its Fourier transform magnitude mapped in the log-polar space. The magnitude of the Fourier transform is invariant to shifts and the log-polar space maps rotation and scaling into horizontal and vertical shifts (Gibson et al. 2001).

For subpixel displacements between the images, the delta function of the PC is spread over a small neighborhood, corresponding to a Dirichlet kernel (Foroosh et al. 2002). The zero-padding of the normalized cross power spectrum $R(\eta, \xi)$, as suggested by (Marcel et al. 1997), involves the inverse Fourier transform of a large matrix. A computationally costless approach for subpixel accuracy is the interpolation of the peak position using its neighborhood pixels values (Gibson et al. 2001). Let the maximum peak subpixel position be $PC(x_0 + \Delta x; y_0 + \Delta y)$:

$$\Delta x = \frac{PC(x_0 + 1, y_0) - PC(x_0 - 1, y_0)}{PC(x_0, y_0) + PC(x_0 + 1, y_0) + PC(x_0 - 1, y_0)} \quad (3)$$

$$\Delta y = \frac{PC(x_0, y_0 + 1) - PC(x_0, y_0 - 1)}{PC(x_0, y_0) + PC(x_0, y_0 + 1) + PC(x_0, y_0 - 1)} \quad (4)$$

In real situations, the PC result is noisy from image non-rigid distortions and temporal differences, and can result in a wrong location of the maximum peak. However a concentration of high-values are present in the neighborhood of the real peak location. Therefore the peak localization is improved by low-pass filtering the PC with a Gaussian kernel (standard deviation of the Gaussian low-pass filter set to 2 pixels for all the experiments with a PC size of $1,024 \times 1,024$ pixels). A PC result is presented in Fig. 1 before and after low-pass filtering.

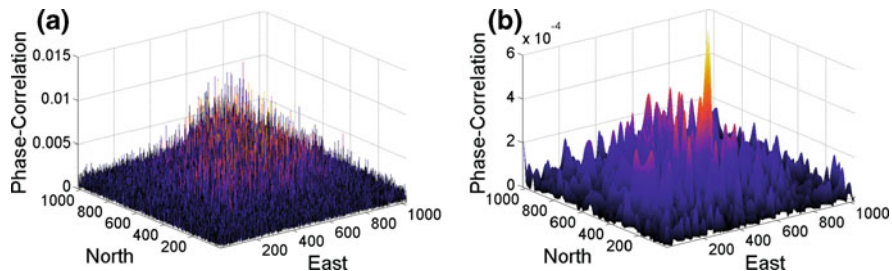


Fig. 1 Phase-correlation result before **a** and after **b** low-pass filtering. The maximum peak corresponding to the translation between two images is more accurate after smoothing

2.3 Robust Image Representation

The differences in illumination from the sunlight, weather conditions and sensors sensitivity disturb the registration process. These changes are mainly concentrated in the low frequencies of the image and can be attenuated by using a high-frequency image representation. The gradient representation ∇I was used by Argyriou and Vlachos (2004) with the vertical derivative mapped to the imaginary part. We assessed that this representation was sensitive to changes between the images and preferred to use the more robust Euclidian norm of the gradient:

$$G(x; y) = |\nabla I| = \sqrt{(\partial_x I)^2 + (\partial_y I)^2}.$$

The gradient norm representation introduces artifacts in the FMI descriptor, because of the edges cutted at the image boundaries. This introduces high frequency values in the Fourier magnitude which are altering the FMI descriptor and would result in registering the frame of the images. However these artifacts can be attenuated by applying a Tukey window on the gradient representation (Vandewalle et al. 2006). The Tukey window is a truncated Gaussian letting unchanged the central part of the image and attenuating only the borders.

The log-polar transformation, applied on the Fourier transform magnitude of the gradient norm image, maps the low frequencies into an important part of the log-polar image. These low frequencies gives not much information on orientation and are too distorted to be useful for scaling. Reddy and Chatterji (1996) suggested to use a high-pass filter before the log-polar mapping to discard low frequencies.

2.4 Registration Extrapolation from Inter-Frame Motion

The consecutive frames of a video sequence are highly redundant and extremely similar. The important overlapping part between two consecutive frames allows the estimation of motion parameters between them. The motion between consecutive frames can be efficiently represented by a rigid transformation (rotation,

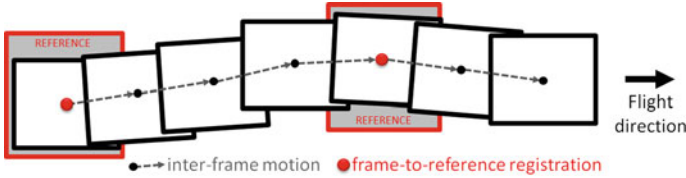


Fig. 2 Schematic representation of frame-to-frame and frame-to-reference registrations

scaling and shifts). Therefore PC can be used between video frames to estimate the inter-frame motion parameters. This is often used for video blocks motion in video compression applications (Argyriou and Vlachos 2004), however here we consider the entire image and not several blocks. The motion between video frames allows to extrapolate the parameters of previous frame-to-reference image registrations. Therefore the registration between a video frame and the reference image gets an accurate initialization from the extrapolated registration parameters of the previous frame. If important changes occurred between a frame and the reference image the registration between them could rely only on the extrapolated registration parameters. The next section presents a PC score allowing to decide whether the registration parameters should only be extrapolated or not (Fig. 2).

2.5 A Phase-Correlation Score for Registration Quality Assessment

The algorithm works at two levels: inter-frame registration (motion estimation between consecutive video frames) and frame-to-reference registration (registration of a video frame with a reference image). The inter-frame registration is used to initialize the frame-to-reference image registration. The maximum peak search in the PC plane is restricted to a certain window, avoiding to get extreme values corresponding to incoherent parameters. The search window is defined by the maximum angle, scaling and translation values estimated possible between two consecutive frames.

The introduction of a PC score at the frame-to-reference translation level allows to regulate the registration. The PC score is derived from the normalized PC peak power. The peak can be spread over several pixels around the maximum (Dirichlet kernel), as seen in Sect. 2.2 therefore the PC score is defined as

$$score = \frac{\sum_{i=-1}^1 \sum_{j=-1, |j| \neq |i|}^1 PC(x_0 + i, y_0 + j)^2}{\sum_{i=1}^N \sum_{j=1}^M PC(i, j)^2} \quad (5)$$

The score reflects directly the quality of the registration. In case of a low score, under a user-defined limit, the frame-to-reference parameters are discarded and

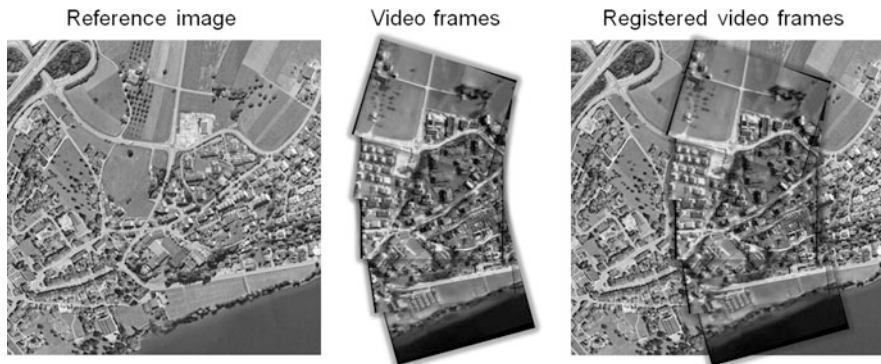


Fig. 3 Reference image (*left*), video $n^{\circ}2$ frames (*middle*) and overlay of the registration (*right*)

only the initial parameters, obtained by extrapolation from the previous frame, are used for registration.

3 Experimental Setup

Performances of our Robust Phase-Correlation algorithm using Motion Estimation (RPCME) have been assessed and compared in an experimental setup made of two airborne videos and airborne reference images from GeoEye at high-resolution (~ 50 cm). One reference image and the corresponding video frames registered are represented and overlaid in Fig. 3.

The airborne videos were taken at constant altitude with a fixed field of view. They have an angle of view close-to-vertical and have been corrected for perspective distortion (some residual distortions are remaining in video $n^{\circ}2$ from inaccurate angle of view information). The video camera has a single panchromatic band ranging up to near-infrared wavelengths with a resolution of 567×720 pixels resulting in pixels with a resolution around 50 cm and a monochrome depth of 8 bits. The sensitivity to near-infrared gives an appearance not similar to the Geoeeye products, which is important to assess the robustness of our algorithm to different sources.

The video $n^{\circ}1$ is a straight flight over agricultural fields and roads with a cloud partially covering frames 20–50. The video $n^{\circ}2$ is a right-turning flight over urban and agricultural areas with important urban changes between frame 15–25: (Fig. 3).

The RPCME algorithm is compared with the Mutual Information maximization algorithm (MI), having an important robustness to multi-sources images (Lu et al. 2008).

Both RPCME and MI have the same initializations from registration parameters extrapolated using the inter-frame motion estimated by phase-correlation. Therefore, only the frame-to-reference registration changes between the two algorithms

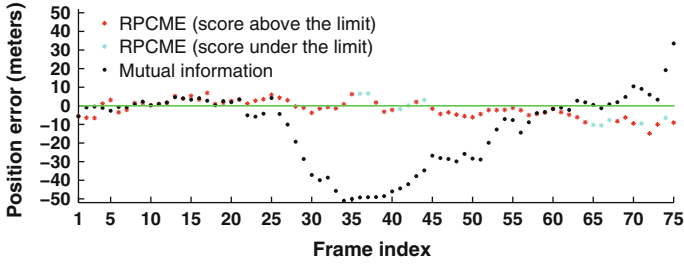


Fig. 4 Video n°1 registration results: position of the registered frames with RPCME and MI

and the image representation (pixel intensities for MI). The MI algorithm always perform a frame-to-reference registration since the regulation is done using the PC score.

The registration positions and orientations are compared with a flight track obtained from coarse GPS flight positions. Hence the position accuracy assessment can only be done on the across-track position.

4 Experimental Results

In Fig. 4, the position of the registered frames from video n°1 are compared relatively to the flight track. RPCME is extremely robust to the cloud partial occlusion between frames 25 and 50 compared to the MI algorithm. The frames having a PC score lower than the score limit, can be observed in light blue dots. Already between the frames 25 and 35 the RPCME is more accurate than MI, without resorting to regulation (score above the score limit). This robustness is obtained from the gradient norm representation of the images. In Fig. 5, the influence of the score regulation on the position error can be well observed. The score regulation avoids the position error to increase when the cloud is present by stopping the frame-to-reference registration.

The position accuracy in video n°2 is equivalent for both algorithms, except the position error increasing for the RPCME algorithm from frame 15–20 (Fig. 6). Some perspective distortions that could not be corrected between these frames led to inaccurate registration positions. The errors common to both algorithms observed between frames 20–40 are due to the flight track interpolation (Figs. 6 and 7).

Concerning the orientation accuracy, the RPCME algorithm is more accurate along the flight than MI (Fig. 8). After the frame 20 the MI algorithm cannot get back the correct orientation and propagates an orientation error through the frames. It is important to notice that the reference orientation can be locally less accurate, since derived from the interpolated flight track. The end of the flight is over agricultural land having less distinct structures causing trouble to the algorithms to

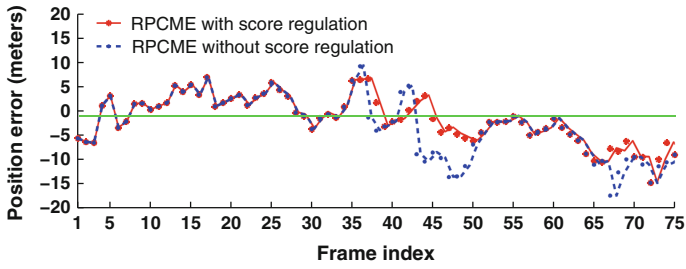


Fig. 5 Video n°1 registration results: RPCME position errors with and without score regulation

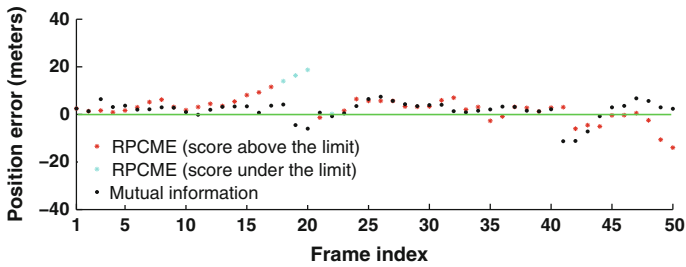


Fig. 6 Video n°2 registration results: position of the registered frames with RPCME and MI

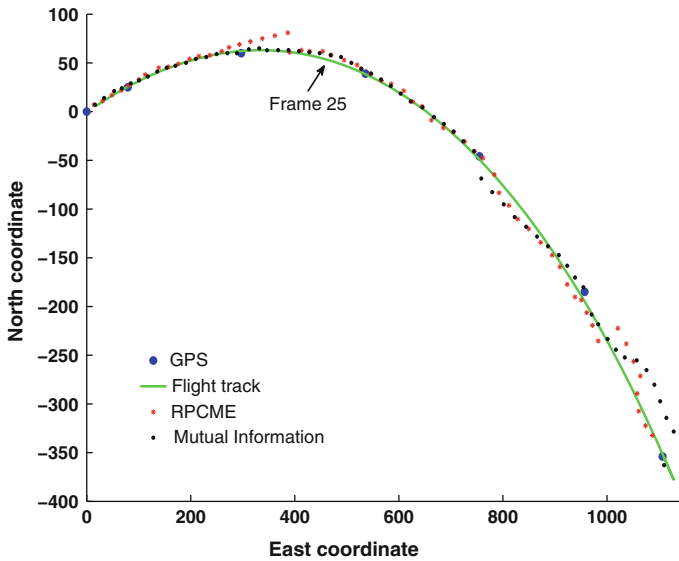


Fig. 7 Video n°2 registration results: 2D view of the position of the registered frames

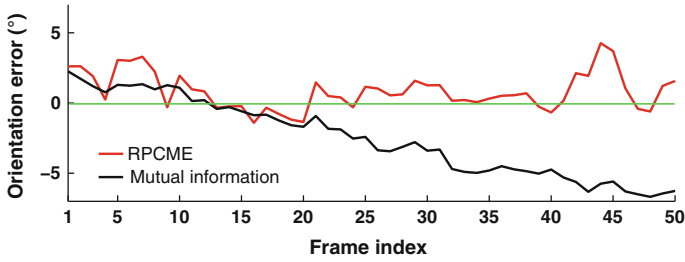


Fig. 8 Video n°2 registration results: orientation errors of the registered frames

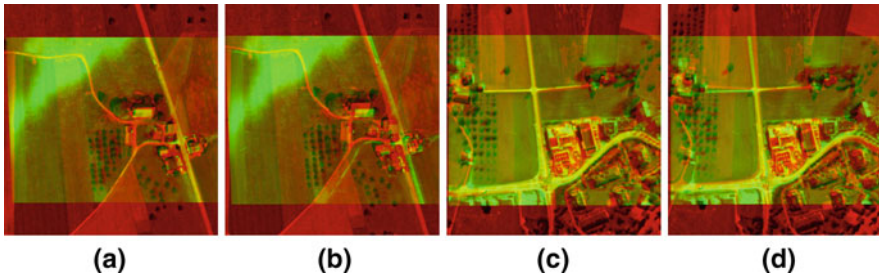


Fig. 9 Video n°1 frame 26: registration results for **a** RPCME, **b** MI. Video n°2 frame 25: registration results for **c** RPCME, **d** MI

get an accurate position and orientation. A visual assessment of the accuracy between RPCME and MI on Red-Green overlays is presented in Fig. 9.

The main limitation are still coming from the presence of structures in the data. Our RPCME method is pixel-based but exploit the gradient norm of the images which keeps only structure in the data. This is a tradeoff between multisource robustness and the required amount of distinct features in the images.

In terms of computation time, RPCME is performing more than five time faster than MI. RPCME requires $8 \text{ FFT} + 4 \text{ FFT}^{-1} + 2 \text{ rotations} + 4 \text{ log-polar inter-polarizations}$. This shows the potential for implementing this algorithm in hardware for near-real time registration. The RPCME algorithm shows accurate position for most of the registered frames and is very accurate on frame orientation compared to the MI algorithm.

5 Conclusion

We have presented how Phase-Correlation (PC) can be used for the registration of airborne video frames with a reference image from a different sensor in the presence of temporal changes. The robustness is obtained from an enhanced image representation based on the gradient norm, from a low-pass filtering of the phase-

correlation and from the extrapolation of the registration parameters between video frames to initialize or replace the video-to-reference registration when not reliable (indicated by a low Phase-Correlation score).

Globally, the Robust Phase-Correlation registration using Motion Estimation (RPCME) algorithm shows robustness and good accuracy to multi-sources images with the advantage of being a fast, non-iterative method. This combination of rapidity and robustness can be particularly interesting for the registration of videos from unmanned aerial vehicles (UAV), allowing some further on-board post processing of the images. It would be particularly interesting to test the algorithm on more airborne videos having higher groundtruth accuracy for a better assessment.

Further perspectives are on the automatic tuning of the score threshold limit, based on the images content and the extension to other types of image modalities.

Acknowledgments Thanks to the EPFL Swiss Space Center and RUAG Schweiz AG for supporting this research project.

References

- Argyriou V, Vlachos T (2004) Using gradient correlation for subpixel motion estimation of video sequences. In: Proceedings of IEEE international conference on acoustics, speech, and signal processing, Citeseer, pp 329–332
- Cannata RW, Shah M, Blask SG, Van Workum JA (2000) Autonomous video registration using sensor model parameter adjustments. In: Proceedings of the 29th applied imagery pattern recognition workshop, pp 215–222
- Foroosh H, Zerubia J, Berthod M (2002) Extension of phase correlation to subpixel registration. *IEEE Trans Image Process* 11(3):188–200
- Gibson S, Kreinovich V, Longpre L, Penn B, Starks S (2001) Intelligent mining in image databases, with applications to satellite imaging and to web search. In: Data mining and computational intelligence, pp 309–336
- Hirvonen D, Matei B, Wildes R, Hsu S (2001) Video to reference image alignment in the presence of sparse features and appearance change. In: Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition, vol 2. pp 366–373
- Kuglin C, Hines D (1975) The phase correlation image alignment method. In: IEEE 1975 conference on cybernetics and society, pp 163–165
- Kumar R, Samarasekera S, Hsu S, Hanna K (2000) Registration of highly-oblique and zoomed in aerial video to reference imagery. In: International conference on pattern recognition, pp 303–307
- Lu X, Zhang S, Su H, Chen Y (2008) Mutual information-based multimodal image registration using a novel joint histogram estimation. *Comput Med Imaging Graph* 32(3):202–209
- Marcel B, Briot M, Murrieta R (1997) Estimation of translation and rotation by fourier transform. *Traitement du Signal* 14(2):135–149
- Reddy BS, Chatterji BN (1996) An FFT-based technique for translation, rotation, and scale-invariant image registration. *IEEE Trans Image Process* 5(8):1266–1271
- Shastry A, Schowengerdt R, Res G, Bangalore I (2005) Airborne video registration and traffic-flow parameter estimation. *IEEE Trans Intell Transp Syst* 6(4):391–405

- Vandewalle P, Susstrunk S, Vetterli M (2006) A frequency domain approach to registration of aliased images with application to superresolution. *EURASIP J Appl Signal Process*, Hindawi Publishing Corporation, pp 233–233
- Wu Y, Luo X (2008) A robust method for airborne video registration using prediction model. In: *International conference on computer science and information technology, ICCSIT'08*, pp 518–523
- Zitova B, Flusser J (2003) Image registration methods: a survey. *Image Vis Comput* 21(11): 977–1000