7-th IASTED International Conference on SIGNAL AND IMAGE PROCESSING, August 15 – 17, 2005, Honolulu, Hawaii, USA
Paper 479-213

PERFORMANCES OF IMAGE MOTION TRACKING WITH OPTICAL CORRELATION

K. Janschek, V. Tchernykh, S. Dyblenko

Technische Universität Dresden, Institute of Automation
Department of Electrical Engineering and Information Technology
D-01062 Dresden, Germany
e-mail: klaus.janschek@tu-dresden.de,
chernykh@ifa.et.tu-dresden.de, dyblenko@ifa.et.tu-dresden.de

ABSTRACT
The paper tackles the problem of real-time image motion tracking as a baseline function for visual motion estimation. Motion tracking is performed using the principles of 2D area correlation and joint transform correlation. To cope with real-time requirements a compact embedded optical Joint Transform Correlator design is presented, which allows using image motion tracking in mobile applications, such as airborne and spaceborne remote sensing as well as mobile robot and satellite navigation. The paper presents the general principles of motion tracking and gives detailed simulation performance results showing the accuracy and robustness under different operational conditions. Airborne flight test results show the achievable performances for an optical correlator hardware model.

KEY WORDS
image motion tracking, image motion estimation, joint transform correlation, optical signal processing, optical correlator.

1. Introduction

Motion estimation from image data is attractive for many applications. In particular mobile systems can benefit a lot from image sensors, because these sensors are in general passive and contact-less and the obtained images contain a lot of information. However the extraction of motion information out of an image sequence needs some specific baseline functions, such as image motion tracking.

Image motion is considered as a time discrete motion of small image parts in a camera image plane. This is caused by the mutual motion of the camera and a scene in the camera field of view. This discrete motion trajectory generally carries information about the motion of the camera with respect to the observed scene as well as information about the 3D structure of the scene. This is similar to the well known principle of stereo-vision, where the analysed scene is imaged by a set of fixed cameras from two or more view points.

The more general problem of image motion estimation can be formulated as the extraction of the two-dimensional (2D) projection of the 3D relative motion into the image plane in form of a field of correspondences (motion vectors) between points in consecutive frames. For practical applications the so called block or window based approach has been proved to be most appropriate. The spatial dynamics of image windows can be analyzed by feature or area based methods, to derive image motion information. Feature based methods basically use computationally efficient edge detection techniques, but they rely very much on structured environment with specific patterns [1].
Area based methods have been proved to be much more robust in particular for image data resulting from unstructured environment. These methods exploit the temporal consistency over a series of images, i.e. they assume the appearance of a small region in an image sequence changes little. Block matching algorithms measure the motion of a block of pixels in consecutive images, such as the sum of squared differences (SSD) algorithm, which incorporates a minimization problem with the desired image shift vector as optimization variable [1, 2] or the method of direction of minimum distortion (DMD) [3, 4].
The classical and most widely used approach however is the area correlation, used originally for image registration [5]. Area correlation uses the fundamental property of the cross-correlation function of two images, which gives the location of the correlation peak directly proportional to the displacement vector of the original image shift. Different correlation schemes are known beside the standard cross correlation, e.g. phase correlation [6] or the Joint Transform Correlation [7].

A more recent and evolving application area for image motion estimation is the video coding (e.g. MPEG-standard) where sub-pixel accuracy is required and multiple moving area blocks have to be tracked using adaptive correlation techniques [8] as well as hierarchical multi-resolution algorithms based on complex wavelets [9].
The limitation of the applicability of area based methods results from the trade-off between computational effort, robustness to non structured image texture and a sufficient signal-to-noise ratio.
The high robustness of area based methods to weakly structured image texture and small signal-to-noise ratio has to be paid by a considerable high computational effort. As the complete image area content has to be processed pixel-wise, the real-time application is restricted to rather small image blocks in the range 8x8 to 32x32 pixels. This limits the accuracy, which is known to be poor when the block size gets too small.

This paper addresses a real-time solution for high precision image motion tracking on the basis of an optical correlator. It exploits the principle of Joint Transform Correlation (JTC) and uses up-to-date opto-electronics technology to derive a compact and robust embedded optical correlation processor. This solution is in particular qualified for mobile applications, such as airborne and spaceborne remote sensing [10] as well as mobile robot navigation [11] and satellite navigation [12].

2. Image Motion Tracking Principle using 2D-Correlation

The basic step for image motion tracking is the measurement of the shift between two overlapped images. Normally they are taken at two close time moments $t_k, t_{k+1} = t_k + \Delta t$ during camera motion with a single image sensor. The second image will be shifted with respect to the first by a shift vector $\vec{D}_0$. This vector is defined by the size and position of overlapped parts for both images. The overlapping can be effectively determined by two dimensional (2D) correlation of the images. The 2D correlation function represents the mutual shift of the second image w.r.t. the first image. This location is normally given by a distinctive peak in the correlation function. Detection of the correlation peak and measurement of its coordinate in the correlation plane allow determining the shift between both images (Figure 1).

For a large class of applications it is necessary to measure the motion of different parts of the camera image plane. In this case the size of the image sensor is much larger than the size of tracked image parts or blocks. This allows measuring the motion of the blocks on rather long distances and increases the operational range of the image motion tracking system.

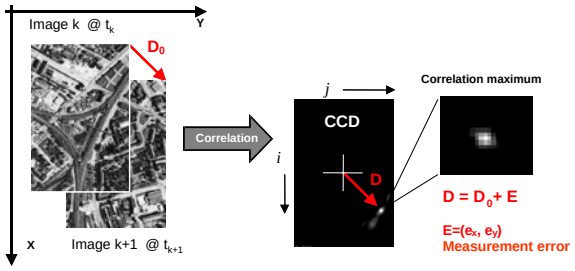

Fig. 1 Image shift vector determination by 2D-correlation

This long range tracking procedure is implemented by a prediction of motion for each tracked block and a subsequent correction of the predicted shift with the correlation data. The correlation is performed for the image from the initial position (reference image) and the image from the predicted position (current image). The prediction can use some external information about camera motion or previous results of image motion measurements.

The so-called *Joint Transform Correlation* (JTC) scheme is used to minimise the overall computational effort [7]. It makes use of two subsequent 2D-Fourier transforms without using phase information (this is highly beneficial for a hardware realisation by optical Fourier processors, see next paragraph).

The two images $f_1(x,y)$ and $f_2(x,y)$ to be compared are being combined to an overall image $I(x,y)$ as shown in Figure 2 (top). A first Fourier transform results in the joint power spectrum $S(u,v) = \mathbf{F}\{I(x,y)\}$. Its magnitude contains the spectrum $F(u,v)$ of the common image contents augmented by some periodic components which are originating from the spatial shift $\vec{G}$ of f_1 and f_2 in the overall image I (Figure 2 centre).

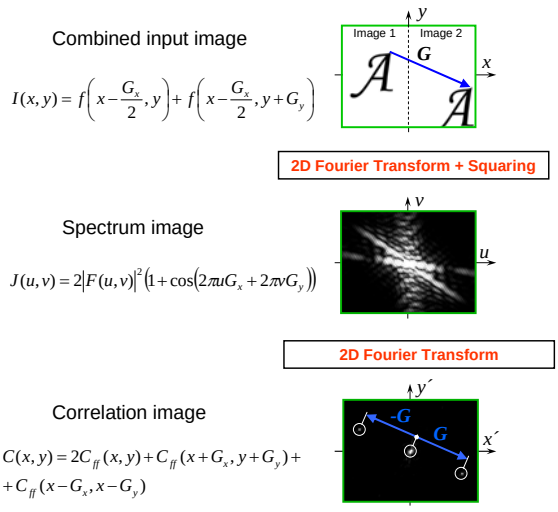

Fig. 2 Principle of joint transform correlation

A second Fourier transform of the squared joint spectrum $J(u,v) = S(u,v)^2$ results in four correlation functions (Figure 2 bottom). The centred correlation function $C_{ff}(x,y)$ represents the auto-correlation function of each input image, whereas the two spatially shifted correlation functions $C_{ff}(x \pm G_x, y \pm G_y)$ represent the cross-correlation functions of the input images. The shift vector $\vec{G}$ contains both the technological shift according to the construction of the overall image $I(x,y)$ and the shift of the image contents according to the image motion. If the two input images f_1 and f_2 contain identical (but shifted) image contents, the cross-correlation peaks will be present and their mutual spatial shift $\vec{\Delta} = \vec{G} - \left(-\vec{G}\right)$ allows determining the original image shift in a straightforward way.

Several advantages of the correlation approach compared to feature based image motion tracking are evident. The correlation approach is much more robust against image noise. Moreover it requires no a-priori information on the image contents and it works well as long as "enough" image texture is existent. The determination of the actual image shift is reduced to the comparable simple task of determining the bright spot of the correlation peak within a certain region of the correlation image.

The main drawback of the correlation approach however is the high computational effort for the two 2D-Fourier transforms, which limits its applicability for real-time solutions. A very promising solution to this problem is given by using optical processors, as outlined in the next paragraph.

3. Embedded Joint Transform Optical Correlator

A *Joint Transform Optical Correlator* (JTOC) implements the JTC-principle by using two identical optoelectronic modules – Optical Fourier Processors (OFP), as sketched in Figure 3.
Each of these Optical Fourier Processors delivers the magnitude of the spatial power spectrum of an input image, which is entered in the optical path by a *spatial light modulator* (SLM). The power spectrum can be read by a CCD or CMOS image sensor located in the focal plane of the Fourier lens of the OFP [13, 14].
For a JTOC the two digital input images $f_1(x,y)$ and $f_2(x,y)$ to be compared are entered into the optical system of the first OFP by a *spatial light modulator*. After the first optical Fourier transformation, the joint power spectrum is read by the CCD image sensor and loaded to the SLM of the second OFP. A second optical Fourier transformation forms the correlation image as sketched in Figure 2 – bottom.

The position of peaks in the correlation image and the shift value can be measured with sub-pixel accuracy using standard algorithms for centre of mass calculation. Optical processing thus allows a unique real time processing of high frame rate video streams.

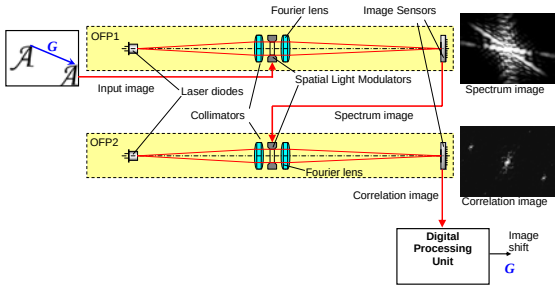

Fig. 3 Joint Transform Optical Correlator (JTOC)

This advanced technology (which is not yet commercially available today) and its applications have been studied during last years at the Institute of Automation of the Technische Universität Dresden. Different hardware models have been manufactured, e.g. under European Space Agency (ESA) contracts.

Due to special design solutions the devices are very robust to mechanical loads and do not require precise assembling and adjustment [15, 16].

The typical optical correlator accuracy of shift determination errors is below 0.2 pixels (1σ) even for extremely noisy images with a SNR less then 0 dB. Processing rates up to 10.000 correlations per second with 128x128 pixel images are possible with up-to-date opto-electronic components [17]. This makes an optical correlator particularly suitable for the determination of the motion of dark and fast moving images in mobile vehicle applications (space, robotics) under weak illumination conditions.

A compact embedded optical correlator model has been developed under ESA contract (see Figure 4), the main performance figures are given in Table 1. To cope with the limited project funding and to save development time, the design uses standard video cameras as image sensors. This limits the image processing rate to 30 optical Fourier transforms per second or 15 correlations per second per one optical Fourier processors (one correlation requires two Fourier transforms).

To provide a net throughput of 60 correlations per second, two optical Fourier processors have been cascaded and the image processing rate for each of them has been doubled by simultaneous processing of two image pairs (Figure 5). Each pair of input images produce a pair of correlation peaks, which are then processed separately and two image shifts are determined.

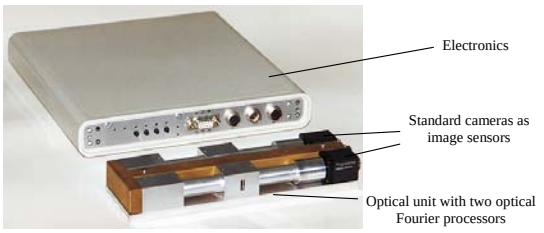

Fig. 4 Optical correlator model

Table 1. Main performances of the optical correlator model

Input	Two standard video inputs 2 x 30 frames/s
Output	Image motion record via RS 232 interface
Image processing rate	60 correlations per second
Dimensions /Mass (optical unit)	210 x 62 x 30 mm / 500 g

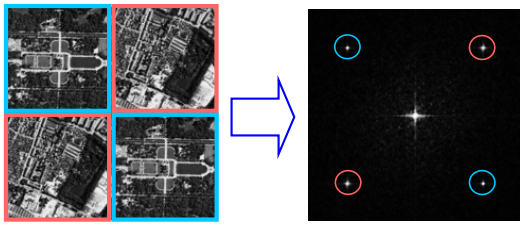

Fig. 5 Simultaneous processing of two image pairs

4. Simulation Performance Analysis

The image motion estimation accuracy depends fundamentally on the accuracy of the determination of the correlation peak. This accuracy has been analyzed with respect to several operational parameters on the basis of simulation experiments with a software model of the JTOC.

Influence of sensor noise

Sensor noise defined as the deviation of measured pixel brightness from the actual value, can disturb the image shift measurement by a distortion of the content of both reference and current image. For real photosensitive image sensors like CCD this noise consists of photon noise, read-out noise and dark current noise. These types of noise are additive and independent, thus sensor noise can be presented as a spatially uncorrelated random process.

It is well known that correlation algorithms are very robust to such type of white noise. Figure 6-a represents the robustness of images shift determination to the noise level of the input images [12]. The signal to noise ratio (SNR) is defined here as the relation of the average image value to the standard deviation of the additive noise.
The test results show, that a sufficiently accurate shift determination (RMS errors below 0.2 pixel) can be obtained even for very noisy images with a SNR about 4 dB (SNR= ~1.6).

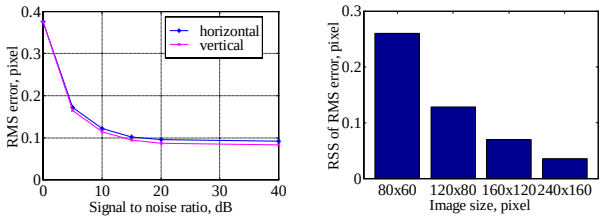

Fig. 6 Dependency of the image shift error on image noise and image size

Influence of image size

Decreasing the size of the image blocks can reduce the calculation time but results in reduced accuracy. The measurement error is increased due to larger non-symmetric distortions of the correlation peak. Figure 6-b shows estimated results for different image size. The range of image shift is ±4 pixels.

It can be concluded, that for different image size the residual error changes approximately inverse to the image area, whereas the time for calculation of the correlation function is proportionally to the image area. A more detailed analysis can be found in [12].

Robustness to image scenes

To assess the possible performances for airborne and spaceborne application, a series of IKONOS demo images with different image scenes has been analyzed (Figure 7). The test images have been normalized to have an equal average brightness of 100. From each large image, 36 reference images (256x256 pixels each) have been cut (uniformly distributed within the test image). For each reference image 81 current images (also 256x256 pixels) have been taken with the shift within ± 144 pixels in both vertical and horizontal directions with respect to the reference image (mutual overlapping of the images was kept at least 40%). This resulted in 2916 correlation pairs for each scene type.

To simulate the camera noise, the random noise with standard deviation σ = 10 has been added separately to each reference and current image (SNR = 20 dB with respect to the average brightness of the image). After correlation the mutual shift of input images has been determined by detection of the correlation peaks within the correlation image and determination of their positions. The detection of peaks has been made by a threshold detector. If no peaks had been found in the correlation

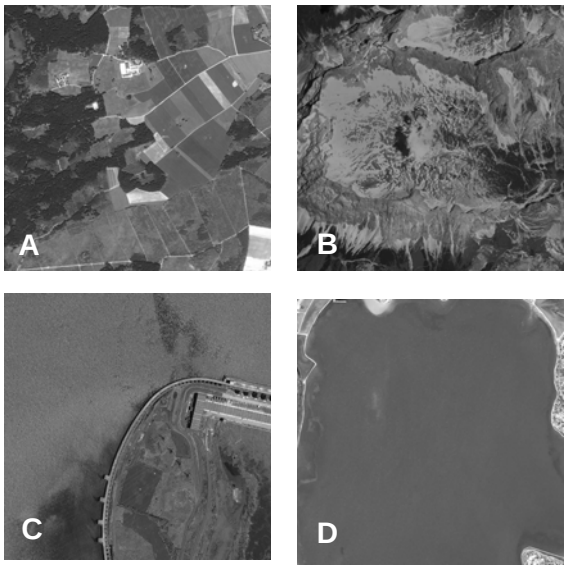

Fig. 7 Test images for scene robustness analysis (source: IKONOS demo images)

Table 2 Results of scene robustness analysis (input images of 256x256 pixels)

Test image	Successful correlations	Standard deviation of error [pel]
Urban scene (similar to Fig. 9)	100.0%	0.0127
Rural scene (Fig. 7-A)	91.9%	0.0166
Mountain with snow and clouds (Fig. 7-B)	93.1%	0.0332
Water surface with waves (Fig. 7-C)	98.2%	0.0062
Water surface w/o waves (Fig. 7-D)	28.6%	0.1061

image exceeding the threshold value, then the correlation was considered unsuccessful. The results are summarized in Table 2.

It can be seen that a high contrast and evenly distributed textures, such as in *urban scenes*, shows perfect correlation results with 100% of successful correlations and standard deviation of errors about 0.01 pixels (for a typical urban scene see Figure 9 left).
If more smooth texture areas occur in the image, the number of successful correlations decreases and the position error increases.

For the *rural scene* (Figure 7-A) the main difficulties for correlation were associated with the field areas with low and regular texture, especially in case, when the correlated images contained also some areas with strong texture, but these areas were not overlapped.

The *mountain scene* (Figure 7-B) showed significant accuracy degradation. A possible reason can be the combination of high general contrast (presence of very bright and dark areas in the image) and low local contrast (within the snow-covered and clouds areas). The images of clouds can be in many cases successfully correlated, but can cause significant errors because of their unknown altitude. The errors, caused by large relief variations within the image, can be corrected using the relief map of the area. Therefore the clouds areas should be excluded from the correlated images (e.g. by a threshold detector), in order to keep the high accuracy.

Very good results were obtained for *water surface with waves* (Figure 7-C): the accuracy was even better, than for the urban scene (standard deviation of errors within 0.01 pixels). Obviously, distinct wave pattern makes a rather good target for correlation. Additional errors caused by the intrinsic motion of the wave field can be neglected for small observation intervals.

Water surface without waves (Figure 7-D) gives an extremely difficult object for correlation: it contains practically no texture. With the same threshold level, as for other test images ($8x10^6$), there were no successful correlations detected. A reduction of the threshold to $0.3x10^6$, showed a result of 28.6% successful correlations (with however considerably large errors).

Robustness to image texture

The robustness of the shift vector determination with respect to different image texture has been analyzed by a specific experiment (Figure 8). Figure 9-left shows a test image with a ground resolution of 0.25 m per pixel from an aerial test campaign with the German DLR High Resolution Stereo Camera HRSC-AX and processed at DLR, Institut für Planetenforschung. The image contains areas with rather distinct texture, which allows testing the system operation with different image content. For the 2D correlation all fragments of the base image were re-sampled to a resolution of 0.75 m per pixel. Subpixel image motion has been simulated by shifting (also at subpixel level) of the base image fragments before re-sampling. The motion blur has been simulated in Fourier domain by multiplication with the Fourier transform of the motion vector.
The image block size for image motion tracking was chosen 128x128 pixel with greyscale 0…255 and random additive image noise 20 greyscale bit (1σ).
Figure 9-right shows the 2D-distribution of the image shift estimation error (sum of bias and random errors 1σ). The error varies from 0.013 pixels for areas with strong texture up to 0.104 pixels for low texture areas. The average error value for the whole test image was 0.027 pixels.

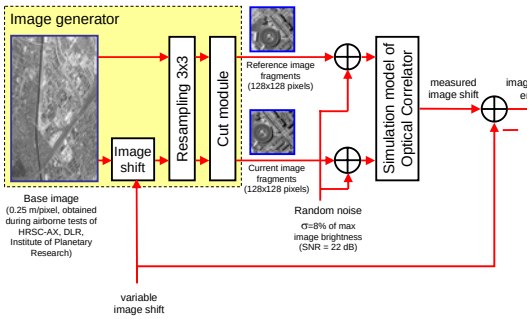

Fig. 8 Simulation configuration of texture

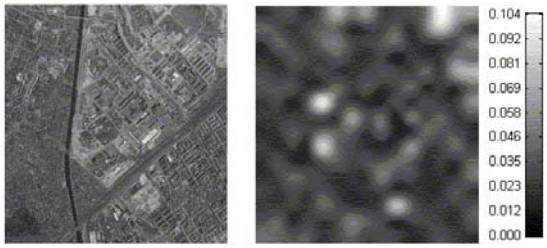

Fig. 9 Results of texture robustness analysis

5. Flight Test Results

A proof of real-time image motion determination performances has been performed with the optical correlator hardware, as presented in para.3, onboard a turboprop aircraft in the frame of a test campaign of a smart pushbroom imaging system [18].

The image acquisition has been performed with a camera with a focal length of 110 mm and an aperture ratio of 1:8. The flight altitude was approximately 2400 m with a velocity 240 km/h (67 m/s), which resulted in a ground resolution of the camera model of 0.45 m per pixel. During imaging, the plane produced considerable attitude disturbances and vibrations, which were partially filtered by a flexible camera suspension (high frequency components were suppressed).

A direct determination of errors of the image motion determination was not possible for these airborne tests, because no reference attitude and position data with required accuracy and bandwidth were available. To estimate the errors under these conditions, the following procedure has been applied. All matrix sensor images (60 frames per second – totally 512 frames for two matrix sensors) have been recorded in flight by high rate digital data recorder. After flight the sequence of images was played back to the optical correlator, simulating the in-flight output of the matrix sensors. The playback was performed two times and the correlator produced two image motion records with *the same set of matrix sensors images*. For the second record, however, the optical correlator was forced to use different image fragments for image motion tracking (Figure 10).

In the ideal case the records, derived by such a procedure, should coincide after subtraction of the initial shift Δ. In

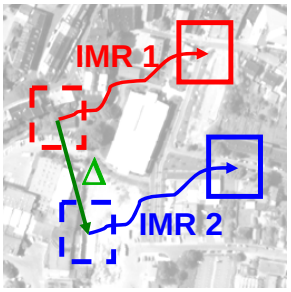

Fig. 10 Two image motion records producing by tracking of two different fragments of the same set of images

our case some differences were observed (Figure 11).

The errors of two records due to tracking of two different image fragments can be considered generally as statistically independent (at least for the high frequency component). In this case the difference between the records represents the vector sum of errors of both records. The mean square value of the difference for all 9 imaging sessions was generally within $\sigma \leq 0.25$ pixels, what allows the conclusion that the error of the image motion record is also within $\sigma \leq 0.25$ pixels.

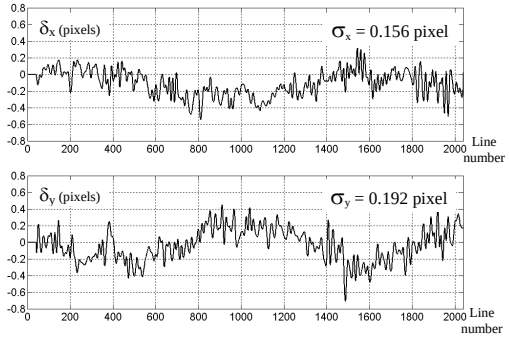

Fig. 11 Difference between the image motion records, δ_y – in the flight direction, δ_x – perpendicular to the flight direction

6. Conclusions

The paper has shown a real-time solution for image motion tracking using a Joint Transform Optical Correlator. The performance analysis based on simulation studies with a software model of the correlator and flight tests with an embedded correlator hardware model has shown that subpixel accuracy for image motion

determination is possible even under weak imaging conditions.

The outstanding computational performance of the optical processor makes it a favourable candidate for a key component in visual navigation and control applications. Currently the main research emphasis of our group is directed towards the use of this correlator principle for the real-time determination of optical flow in the context of visual navigation applications.

7. Acknowledgements

The work presented in this paper was partially funded by different contracts from European Space Agency (ESA).

References

[1] Hutchinson, S., Hager, G.D., & Corke, P.I., A tutorial on visual servo control. *IEEE Transactions on Robotics and Automation*, Volume 12, Issue 5, Oct. 1996, 651 – 670.

[2] Oh, P.Y., & Allen, K. Visual servoing by partitioning degrees of freedom. *IEEE Transactions on Robotics and Automation*, Volume 17, Issue 1, Feb 2001, 1 – 17.

[3] Jain, J., & Jain, A., Displacement Measurement and Its Application in Interframe Image Coding. *IEEE Transactions on Communications*, Volume 29, Issue 12, Dec 1981, 1799 – 1808.

[4] Haworth, C., Peacock, A.M., & Renshaw, D., Performance of reference block updating techniques when tracking with the block matching algorithm. *Proceedings International Conference on Image Processing*, 2001. Volume 1, 7-10 Oct. 2001 (pp 365 – 368, vol.1).

[5] Pratt, W.K., Correlation techniques of image registration. *IEEE Transactions on Aerospace Electronic Systems*, Volume 10, May 1974, 353-358.

[6] Weber, T.E., & Hollis, R.L., A vision based correlator to actively damp vibrations of a coarse-fine manipulator. *Proceedings IEEE International Conference on Robotics and Automation*, 14-19 May 1989 (pp 818 – 825, vol.2).

[7] Jutamulia, S., Joint transform correlators and their applications. *Proceedings SPIE*, 1812, p. 233-243.

[8] Xu, J.B., Po, L.M., & Cheung, C.K., Adaptive motion tracking block matching algorithms for video coding. *IEEE Transactions on Circuits and Systems for Video Technology*, Volume 9, Issue 7, Oct. 1999, 1025 – 1029.

[9] Magarey, J., & Kingsbury, N., Motion estimation using a complex-valued wavelet transform. *IEEE Transactions on Signal Processing*, Volume 46, Issue 4, April 1998, 1069 – 1084.

[10] Janschek , K., & Tchernykh, V., Optical Correlator for Image Motion Compensation in the Focal Plane of a Satellite Camera. *Space Technology* , Volume 21 (2001), Issue 4, p. 127-132.

[11] Janschek , K., Tchernykh, V., Beck, M., Optical Flow based Navigation for Mobile Robots using an Embedded Optical Correlator. *Preprints of the 3rd IFAC Conference on Mechatronic Systems - Mechatronics 2004*, 6-8 September 2004, Sydney, Australia, pp.793-798.

[12] Dyblenko, S., Autonomous Satellite Navigation with Image Motion Analysis Using Two-Dimensional Correlation. PhD Thesis. Technische Universität Dresden. 2004.

[13] Goodman, J.W., Introduction to Fourier optics. *McGraw-Hill*, New York 1968.

[14] Stark, H. (Editor), Application of optical Fourier transform. *Academic Press*, New York 1982.

[15] Tchernykh, V., Janschek, K., & Dyblenko, S., Space application of a self-calibrating optical processor or harsh mechanical environment. *Proceedings of 1st IFAC Conference on Mechatronic Systems - Mechatronics 2000, September 18-20, 2000, Darmstadt, Germany.* Pergamon-Elsevier Science. Volume 3, p. 309-314.

[16] Janschek, K., Tchernykh, V., Dyblenko, S., Verfahren zur automatischen Korrektur von durch Verformungen hervorgerufenen Fehlern Optischer Korrelatoren und Selbstkorrigierender Optischer Korrelator vom Typ JTC. *Deutsches Patent Nr. 100 47 504 B4*, Erteilt: 03.03.2005.

[17] Tchernykh, V., Janschek, K., Application Study for an Optical Correlator. Final Report, ESTEC/Contract No. 17572/03/NL/SFe, Sep. 2004

[18] Tchernykh, V., Dyblenko, S., Janschek, K., Göhler, W., & Harnisch, B., SmartScan - Hardware Test Results for Smart Opto-Electronic Image Correction for Pushbroom Cameras. *Proceedings of the SPIE, Earth Observing Systems VII.* Edited by Barnes, William L., Vol. 4814, p. 264-272.