ENERGY FUNCTION BEHAVIOR IN OPTIMIZATION BASED IMAGE SEQUENCE STABILIZATION IN PRESENCE OF MOVING OBJECTS

F. Karimi Nejadasl[a, *], B. G. H. Gorte[a], M. M. Snellen[a], S. P. Hoogendoorn[b]

a Delft Institute of Earth Observation and Space Systems, Delft University of Technology, Kluyverweg 1, 2629 HS, Delft, The Netherlands - (F.KarimiNejadasl, B.G.H.Gorte, M. Snellen) @tudelft.nl
b Transport & Planning Department, Delft University of Technology, Stevinweg 1, 2628 CN, Delft, The Netherlands - S.P.Hoogendoorn @tudelft.nl

Commission III, ICWG III/V

KEY WORDS: Registration, Transformation, Visualization, Orientation, Correlation, Image Sequences, Aerial

ABSTRACT:

In this paper, we address the registration of two images as an optimization problem within indicated bounds. Our contribution is to identify such situations where the optimum value represents the real transformation parameters between the two images. Consider for example Mean Square Error (MSE) as the energy function: Ideally, a minimum in MSE corresponds to transformation parameters that represent the real transformation between two images. In this paper we demonstrate in which situations the optimum value represents the real transformation parameters between the two images. To quantify the amount of disturbances allowed, these disturbances are simulated for two separate cases: moving objects and illumination variation. The results of the simulation demonstrate the robustness of stabilizing image sequences by means of MSE optimization. Indeed, it is shown that even a large amount of disturbances will not cause the optimization method to fail to find the real solution. Fortunately, the maximal amount of disturbances allowed is larger than the amount of signal disturbances that is typically met in practice.

1. INTRODUCTION

Collection of vehicle dynamics data from airborne image sequences is required for setting up and calibrating traffic flow models (Ossen and Hoogendoorn, 2005). The image sequence is collected by a camera mounted below a helicopter hovering over a highway. The images are not stable because of the helicopter drift. Therefore the camera motion should be separated from vehicle motion. Toth (Toth and Grejner-Brzezinska, 2006) used GPS/INS for camera position estimation but only for image sequences at low frame rate. Feature based solutions have to deal with considerable amount of errors caused by mismatching and moving objects. Kirchhof (Kirchhof and Stilla, 2006) and Medioni (Yuan et al., 2006) have used RANSAC as a robust estimator to remove outliers. Although this method could handle considerable amount of outliers robustly it fails for images with low frequency content due to the lack of availability of enough matched points. This contradicts with the main requirements of our application which are automation and robustness.

Consequently we have proposed a method (Karimi Nejadasl et al., 2008) to use explicit radiometric and implicit geometric information even for pixels with a very low gray value change with respect to their neighbors. The main idea is based on having one dominant motion between two images which can be formulated as one transformation matrix that transforms the whole image geometrically to achieve the second image. As a result, the transformation parameters are the one that provide the best match between two images: reference and candidate image that should be registered to the reference image.

Between consecutive images moving objects and illumination variations cause only small difference. But between an arbitrary image and the reference image these disturbances are more than in the consecutive case. The amount of disturbances is influenced by ambient conditions that can be subdivided into environmental, traffic and scene circumstances. A large amount of these disturbances could cause a failure of our optimization method.

Before being able to apply our method on large data sets it is necessary to find out how robust our method is by determining which disturbances are manageable.

We simulate two types of disturbances: illumination variations and moving objects. Then the transformation parameters are estimated for each disturbed data set. Later on the amount of errors on the estimated parameters and the image coordinates is calculated. The amount of disturbances is increased until the energy value of the estimated parameters with a high geometric error is lower than the energy value of the real result. This situation corresponds to the real failure of the method. The amount of disturbances is then lowered until a correct result is obtained. The amount of disturbance related to this result indicates the acceptance boundary.

In Section 2, the image-sequence-stabilization framework is introduced. The procedure of finding the boundary of our method is described in Section 3. Results and conclusions are presented in Section 4 and 5 respectively.

* Corresponding author.

2. IMAGE-SEQUENCE REGISTRATION

Movement of the camera results in recording different images. In principle, reconstructing an image in the new camera position is possible from the previous image by knowing the movement of the camera and the distance of an object in the scene to the camera.

Using wrong transformation parameters between two images, results in a transformed image that is not oriented in the same way as the reference image. The first image is the reference image and the second one the candidate image which should be registered to the reference one. The mismatch can be visualized by differences between the reference image and the transformed candidate image. The Mean Square Error (MSE), is used to express the misalignment between the transformed image and the reference one. The optimized transformation parameters are those that provide the maximum agreement between the reference and transformed candidate image.

Consequently, the transformation parameters are the ones where the difference between the transformed image and the reference image is minimal. In other words, the transformation parameters are obtained by minimization of the MSE between the transformed image and the reference image.

2.1 Transformation Parameters

In this paper a projective model without shearing and different scale parameters is used as a transformation model on the calibrated images (Heikkila, 1997; Zhang, 1999). This model can be described by:

$$x_1 = \frac{s\cos(\theta)x_2 + s\sin(\theta)y_2 + t_1}{v_1x_2 + v_2y_2 + 1}$$
$$y_1 = \frac{-s\sin(\theta)x_2 + s\cos(\theta)y_2 + t_2}{v_1x_2 + v_2y_2 + 1} \qquad (1)$$

S, θ, t_1, t_2, v_1, and v_2 are respectively scale, rotation, translational and special projective parameters. x_1 and y_1 are image coordinates of the first image and x_2 and y_2 are the image coordinates for the second image. All the image coordinates are given w.r.t. the center of the image at hand. As a consequence our parameter space is six dimensional. Each point in parameter space is a parameters' combination which corresponds to a transformed image and therefore to an energy value.

The Mean Square Error (MSE) is used as an energy function:

$$F(p_1, p_2, \ldots, p_6) = \min \sum_x \sum_y (A - B)^2 / n \qquad (2)$$

with $A = \tilde{I}_{i+1}^{T(p_1, p_2, \ldots, p_6)}(y, x)$ and $B = \tilde{I}_i(y, x)$

Here $\tilde{I}_i$ and $\tilde{I}_{i+1}^T$ are respectively the reference normalized image intensity, the transformed target normalized image intensity while *n* is the number of pixels in the common area after transformation. Note that $I_{i+1}^T = I_{i+1}(T_{i+1,i}X_{i+1})$ with $T_{i+1,i}$, the transformation matrix and X_i, the image coordinate system for the i-th image.

Searching the whole parameter space for finding the optimum value is computationally very expensive. The complexity is $O(\prod_{i=1:no} n_{p_i})$ with n_{pi} the number of all possible values for each parameter, p_i, and *no* the number of parameters. In our case the search space is 6-dimensional. One could imagine the real number, $\mathbb{R}$, as the search range for each parameter. However, not every combination of parameters is allowed. Each parameter has a certain range beyond which the transformed image is meaningless. Moreover, for each parameter there is a resolution value such that within the resolution value the transformed images are equal. Although incorporating range and resolution of parameters reduces the search space, still the number of potential parameters is quite high.

2.2 Differential Evolution

Therefore, we have applied a global optimization technique. Here Differential Evolution (DE) (Price et al., 2005) is used to find the global optimum.

DE starts with an initial population of *q* randomly (McKay et al. 1979) chosen parameter value combinations **m**. These **m**'s are improved during successive generations of constant size *q*, in the sense that a descendant replaces an **m**, becoming its successor, if it has a lower energy value. The distinctive feature of DE is the way in which these descendants are created. Various ways to generate new **m**'s exist, but here only the following procedure is considered. At the start of generation *k* the parameter vectors $\mathbf{m}_{k,1},\ldots,\mathbf{m}_{k,q}$ are given and for each of them a descendant is created. To create a descendant $d_{k,i}$ a partner $p_{k,i}$ is constructed as follows:

$$\mathrm{p_{k,i}} = \mathbf{m}_{k,j_1} + \mathrm{F}(\mathbf{m}_{k,j_2} - \mathbf{m}_{k,j_3}) \qquad (3)$$

with the three different **m**-vectors chosen at random from the population and F a scalar multiplication factor between 0 and 1. The descendant $d_{k,i}$ of $\mathbf{m}_{k,I}$ results from applying crossover to $\mathbf{m}_{k,i}$ and $p_{k,i}$ with crossover probability *pc*. A higher value of *pc* leads (on average) to more dimensions of $p_{k,i}$ being copied into $\mathbf{m}_{k,i}$. Descendant $d_{k,i}$ only replaces $\mathbf{m}_{k,i}$, becoming its successor, if its energy is lower. The setting parameters of DE are population size q, multiplication factor F, crossover probability *pc* and the number of generations *NG*. The values chosen for the setting parameters are used according to (Snellen and Simons, 2007).

Two types of image registration occur in our data sets: registration between consecutive images and registration between an arbitrary image to the reference image. There is a high correlation between image frames because of the helicopter hovering to keep the viewing area fixed. However, shaking of the helicopter causes a drift. This movement can be enhanced by increasing temporal differences. The small movement between consecutive frames and the high correlation between image frames direct us to the design a framework for the registration of two arbitrary images to avoid excessive computations. A final result of this framework, after applying it to all available frames is a stabilized image sequence. The framework is summarized as follows:

1. compute $T_{i+1,i}$, the transformation between I_{i+1} and I_i
2. compute $\tilde{T}_{i+1,1} = T_{i,1}T_{i+1,i}$ the estimated transformation between I_{i+1} and I_1

3. use $\tilde{T}_{i+1,1}$ as the initial value for computing $T_{i+1,1}$

In this solution the image is processed frame by frame, starting from a reference frame, which (for simplicity) we will assume to be frame *1* in the sequence. Assume frames *2...i* are already registered to frame *1*, which means that the transformation $T_{i,1}$ between frames *i* and frame *1* is known. Within the framework the image *i+1* is registered to the first image.

This strategy also prevents registration errors to accumulate. Matching consecutive images (step 1) is easier (i.e. less error-prone) than matching arbitrary images, since the misalignment is limited. In step 3, this problem is avoided by providing an accurate approximate value to the matching process.

3. SIMULATION OF THE DISTURBANCES

In this section, the disturbances of our method which are the illumination variation and the moving objects are simulated. The amount of permitted disturbances will give a quantitative indication of the robustness.

3.1 Type of the Disturbances

Within our stabilization framework as sketched in Section 2 any arbitrary image registration is treated as a consecutive image registration. But in fact, the registration problem becomes different due to the disturbances: moving objects and illumination variations. The disturbances are increasing with increasing temporal differences (type small and gradual to type large and sudden). The number of pixels changing due to moving objects is in general lower than the total number of pixels that represent the moving objects duet to overlap of a moving object in different images. The illumination values in the overlapping area are almost the same. The illumination variation is small in consecutive images. Therefore the effect of theses disturbances is very small in the process. This effect results in a small MSE value. By increasing the temporal distance the amount of these disturbances is increasing. Decrease of the overlapping area increases the number of pixels in moving objects. Although after a while when there is no overlap, the amount of moving pixels stabilize. On the other hand, the number of moving objects may increase by changing traffic situation, e.g. from a moving type to a congested type. Also many object outside will influence the number of moving pixels. The effect of local illumination variation is increased for example by the appearance of clouds in one part of the image. Global illumination variations are not problematic as they can be removed by using a normalized form, a difference of the image gray values from their mean.

The change of illumination depends on the source of the light, object characteristics, viewing angle, and influence of other objects. Examples of these changes are shadows of fixed and moving objects; a reflection of vehicle lights from the road surface; changing the viewing angle caused by shaking of the helicopter results in illumination variation of road lines and vehicles especially because of specular effects.

In fact, moving objects can be interpreted as the local illumination variations which destruct the image structure of an occupied area. The energy function, which explicitly depends only on illumination values, cannot distinguish between these two types of disturbances. As a result, in our simulation, moving objects and small region illumination variations are treated the same.

3.2 Used Simulation

All simulated moving objects are rectangular, consisting of 100 × 22 pixels. The image size is 1392 × 1040 pixels. The position of these objects is randomly distributed over the whole image area in the reference image. To have maximum variation, the gray value is specified as the maximum value in an intensity range, here 255, because of having mainly darker background in our data sets. All these white simulated objects are moved with object width, 100 pixels, in x-direction and object height, 22 pixels, in y-direction to have a higher amount of disturbances with very high correlation in object motion. The disturbances, in this case, are destructing image content as if there was a destructive structure occurred such as a moving object or a specular reflection in water or windows. This is the worst case of moving object simulation because of high correlation motion. If the objects move differently or the objects are different in two images, the disturbance of this type is less problematic than having moving objects which move the same.

To generate the illumination changes, the reference image is subdivided to in four non equal regions. In each region all gray values are disturbed by a fixed amount. The worst case of illumination variation is when the structure of an image is destructed by the disturbances. For example reducing the gray value in the dark image can cause more severe problem than increasing the gray value as in the later case the image structure is not essentially affected. Although in preserving case the amount of the disturbance is more than the constructive case.

After simulation of disturbances, a camera motion is simulated. The reference image is transformed by applying the simulated camera motion parameters. Ideally, the estimated transformation parameter values should be the same as the parameter values applied to simulate the camera motion. The reason of simulating a transformation is to have real parameter values for validation. Although the transformation parameters are obtained by manual corresponding point selection and then parameter estimation, exact positioning of correspondence points manually is erroneous due to image resolution.

The total amount of disturbances should be calculated after removing the camera movement. Therefore the intentionally moved object and illumination variations are introduced before inserting motion. The advantage of this order is that additional radiometric errors are avoided. Consequently, two images are the same before inserting disturbances in both of them and transforming the reference one.

3.3 Boundary Calculation

The percentage of the amount of disturbances is the total amount of absolute disturbances relative to the maximum total amount of possible disturbances, i.e. the number of pixels multiplied by the maximum grayscale of the pixel depth. For example for a 8 bit image, the pixel depth equals 256. Accuracy of the calculated parameters is quantified as normalized parameters' error and geometric error.
The parameters are normalized by dividing for each parameter its absolute error by its resolution. This value indicates how many times each parameter value error deviates from its resolution. The resolution of each parameter is calculated by discarding the other parameters and obtaining maximum one

pixel displacement. In principle, this way is not measuring the main resolution. But to make a procedure tractable, resolution is measured without considering the effects of other parameters.

The geometric error, a positional error in correspondents, is calculated by making a grid over the whole image area for both images. This grid is transformed by the real transformation parameter values and by the estimated transformation parameter values for the reference and the candidate image respectively. The maximum, minimum and the number of displacements larger than one pixel are also recorded. Our final decision is based on having no displacement larger than one pixel.

For each percentage of the amount of the disturbance, parameters are estimated and geometric and parameter errors are calculated. To find the boundaries, the amount of errors is increased until the estimated parameters result in a wrong transformation.

3.4 Failure Mechanism

One obvious reason why the DE method may result in a wrong transformation is when the obtained transformation has a lower energy value, given the simulated errors, then the energy value that corresponds to the real transformation. I.e. in this case the minimum of the energy function is no longer corresponding to the real solution.

Another case occurs when the estimated parameter values are wrong although their energy is higher than the energy of the real parameter values. This case corresponds to the failure of the optimizer in finding the global minima with our settings even if the minimum is shifted. If the DE setting parameters are not sensitive enough, if may be necessary to increase the number of generations, *NG*, in combination with using a smaller multiplication factor, F, and a small cross over probability, *pc*, to find the global minimum. In our parameter space, the special projective parameters v_1 and v_2 (i.e. the fifth and sixth parameters) are less sensitive than the other transformation parameter of Equation 1 in changing energy value especially by increasing amount of disturbances. This sensitivity is reduced by increasing the amount of disturbances. Changing the optimizer settings in these cases is likely to succeed, of course at the cost of increasing the computational effort. The added value seems not high which results either in very little increasing acceptance boundary or very little increasing the rejection boundary. We consider this case also as a failure.
The amount of the disturbances is increased until the global minimum no longer corresponds to the real parameter values. Then the amount of errors is reduced and the optimizer is run a few times till the result of all runs are correct. Otherwise the errors are reduced and run again.

The above-mentioned procedure is done for both simulation types to find the minimum amount of disturbances cause failure. The method therefore can handle disturbances lower than this amount.

4. RESULTS

Our image sequences are recorded from a non-stable platform, in this case a helicopter hovering above a highway. These image sequences are used to collect statistics concerning the behavior of drivers of all vehicles on a highway stretch in busy (nearly congested) traffic during an elongated period of time. Typically, we record highway stretches with a length of 300−500m during one hour or more. We use a b/w camera with 1392 × 1040 pixels which gives a ground resolution of approx. 25−40cm, at a frame rate of 15 fps. The transformation parameters ($S, \theta, t_1, t_2, v_1, v_2$) used for the simulation in this paper are:

[0.9942 -0.7184 6.3931 8.1876 1.1395e-5 -2.4079e-5]

The number of generation, *NG*, population size, q, multiplication factor, F, and cross over probability, *pc*, are respectively 50, 16, 0.6, and 0.55. All the calculations are done in a second fine image scale of an image pyramid and the results are scaled up. The range of the parameters for the maximum 10 pixel movement is:

[1- 0.0912 -0.8232 -10 -10 -2e-5 -3.7e-5]
[1+0.0912 0.8232 10 10 2e-5 3.7e-5]

respectively for lower and higher band. The resolution of the parameters for one pixel movement is:

[1+0.0091 0.0823 1 1 2e-6 3.7e-6]

Figure 1 and Figure 2 demonstrate the maximum amount of allowed moving objects and illumination variations respectively for two different data sets. The result represents the fact that the second data set (highway crossing) can handle both a large amount of moving objects and a larger amount of illumination variations. The boundaries of the acceptance of the method are represented in Figure 3 and Figure 4 as the amount before the star on the x-axis for the moving objects and illumination variations respectively for two the different data sets. The amount after the star indicates the rejection boundaries either because of the failure of the optimizer within our settings, the amount between the star and the rectangle, or because of the real failure, the amount after the rectangle. The star is an example of the optimizer failure. The y-axis demonstrates the absolute parameter error divided by the resolution. This error is visualized for each parameter.

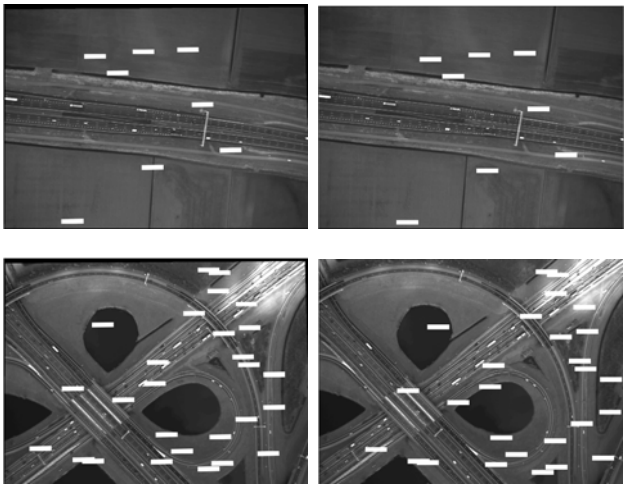

Figure 1: Moving objects before failure in data set 1 (up) and data set 2 (down). The left figures are the reference images and the right ones are the candidate images. The difference between the images is the transformation of the whole image and object motion.

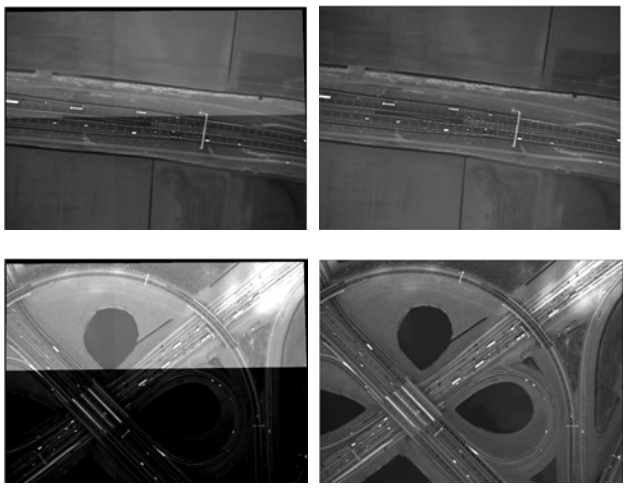

Figure 2: Illumination variation before failure in data set 1 (up) and data set 2 (down). The left figures are the reference images and the right ones are the candidate images. The difference between the images is the transformation of the whole image and illumination differences.

5. CONCLUSION

We have evaluated and quantified the robustness of our stabilization method with respect to the amount of disturbances. The disturbances were simulated as either moving objects or illumination variation. The acceptance of our method is decided based on acceptable parameter and geometric errors.

The simulation is done based on extreme case of illumination variation and moving objects. The presented percentage of amount of the disturbances can be increased in the case of having illumination variations with non destructive nature or moving object with low motion correlation.

More disturbances would be handled by having more structure in the image. The results demonstrate the low percentage of disturbances in the acceptance boundaries in an image with almost no structure outside the road area and having a road in the middle of the image. However the stabilization of this case can be done without having a very high amount of disturbances. This data set is an extreme case. In the other data set with highway crossing, even a very high amount of the disturbances in both moving objects and illumination variations provide acceptable results. This case shows the robustness of our method in handling very large disturbances which in reality would not occur.

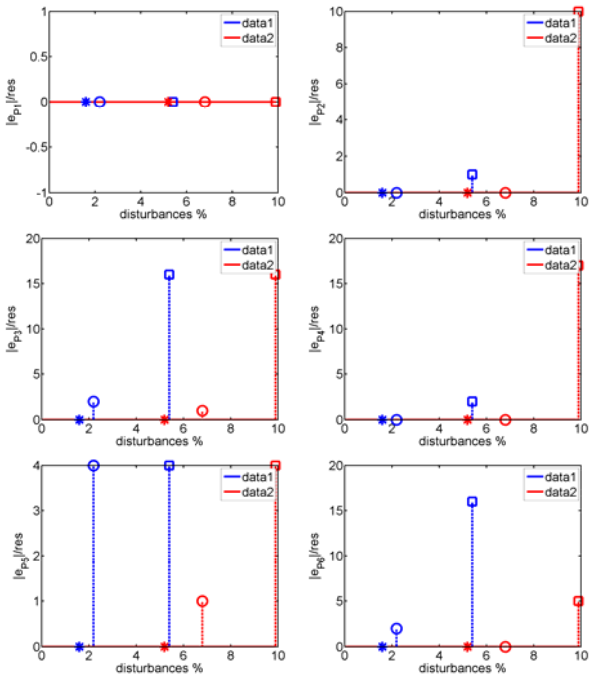

Figure 3: Moving object boundaries

Data set 1 is represented in blue and data set 2 in red. The acceptable percentage of the amount of the illumination variation is represented by the star. The region between the star and rectangle is the optimizer failure within our settings. The circle shows one of the failure cases from this type. The area after rectangle shows the real rejection. The method is robust before the specific amount of moving objects which is indicated by star. All the normalized parameter errors are also zero.

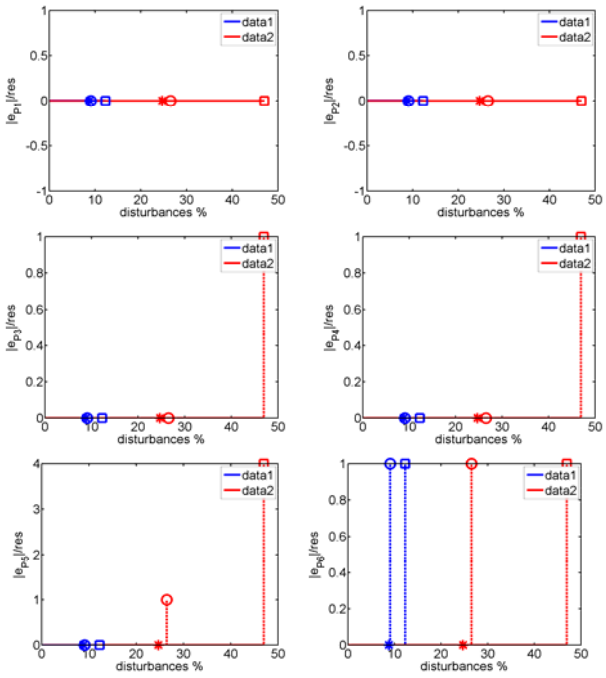

Figure 4: Illumination variation boundaries

Data set 1 is represented in blue and data set 2 in red. The acceptable percentage of the amount of the illumination variation is represented by the star. The region between the star

and rectangle is the optimizer failure within our settings. The circle shows one of the failure cases from this type. The area after rectangle shows the rejection boundaries. The method is robust before the specific amount of illumination variation which is indicated by star. All the normalized parameter errors are also zero.

ACKNOWLEDGEMENTS

The research presented in this paper is part of the research program "Tracing Congestion Dynamics with Innovative Traffic Data to a better Theory", sponsored by the Dutch Foundation of Scientific Research MaGW-NWO.

REFERENCES

Heikkila, J., O. Silven, 1997. Four-step camera calibration procedure with implicit image correction. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1106-1112.

Karimi Nejadasl, F., Gorte, B.G.H., Hoogendoorn, S.P. and Snellen, M., 2008. Optimization Based Image Registration in Presence of Moving Objects. In: *Proceeding of International Calibration and Orientation Workshop.* Castelldefels, Spain

Kirchhof, M. and Stilla, U., 2006. Detection of moving objects in airborne thermal videos. *ISPRS Journal of Photogrammetry and Remote Sensing*, 61(3-4), pp. 187-196.

McKay, M. D., W. J. Conover, and R. J. Beckman, 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics,* 21, pp. 239–245.

Price, K.V., Storn, R.M. and Lampinen, J.A., 2005. *Differential Evolution: A Practical Approach to Global Optimization.* Springer, 1 edition.

Snellen, M. and D. G. Simons, 2007. An assessment of the performance of global optimisation methods for geo-acoustic inversion. *Accepted for publication in the Journal of Computational Acoustics.*

Toth, C.K. and Grejner-Brzezinska, D., 2006. Extracting dynamic spatial data from airborne imaging sensors to support traffic flow estimation. *ISPRS Journal of Photogrammetry and Remote Sensing,* 61(3-4), pp. 137-148.

Yuan, C., G. Medioni, J. Kang, and I. Cohen, 2007. Detecting Motion Regions in the Presence of a Strong Parallax from a Moving Camera by Multi-view Geometric Constraints. *IEEE Trans. on Pattern Analysis and Machine Intelligence,* 29(9), pp. 1627-1641.

Zhang, Z. 1999. Flexible Camera Calibration by Viewing a Plane from Unknown Orientations. In: *Proceedings of the IEEE International Conference on Computer Vision*, 1, pp. 666-673.