Optical Flow from Warping by the Fourier Transform.

L. Streeter*, M. J. Cree*
*School of Engineering, The University of Waikato, Private Bag 3105, Hamilton 3240, New Zealand.
Email: lvs2@phys.waikato.ac.nz

***Abstract*—The optical flow problem is posed as a nonequispaced inverse Fourier transform. An iterative minimisation scheme is devised using least squares and is assisted by a duality paradigm. The implementation targets the optical flow of simple low texture images. The method is tested on an automatically generated test set. The flow estimation finds a nearby pixel that matches the brightness, so the flow magnitude is prone to decreasing linearly in the direction of the flow vector. As such the proportion of the endpoint error in the flow estimates decreases linearly with an error threshold. The proposed method is successful in a restricted case that is difficult for some of the prior art, thus is deemed to have merit and should be investigated further.**

***Index Terms*—Optical flow, Fourier transform**

I. Introduction

The optical flow problem is the estimation of motion though a video sequence. In video, when objects of different brightness/colour move relative to the image sensor then the position of bright and dark pixels in the video frames also move. Optical flow is the two-dimensional vector estimate that describes the displacement of pixel brightness from one frame to the next. Estimation of optical flow provides useful information about object motion for machine vision and robotics. For example, optical flow vectors indicate the direction in which an automaton is moving, when a backing car might hit an object behind it, or the magnitude and direction of moving objects required to correct or compensate for that motion.

The seminal approaches on optical flow posed the estimation of an inverse fluid flow-alike problem [1], [2]. A constraint for the data was obtained by considering the expansion of a brightness change model in time. Smoothness based on the L2 norm (sum of squares) of spatial differences of the flow vector estimates provided a second constraint for the two variable problem, making the solution tractable.

The estimation of optical flow has progressed by reconsideration of the constraints. The literature is extensive, so we only consider particular references. The reader is referred to the Middlebury website[1] on optical flow for an overview of how the different methods perform [3]. The use of a theory of warping [4], [5], has targeted the difficult problem of large displacements, including when small objects move large distances. Total variation L1-norm (TVL1) optical flow [6] constrained the problem using absolute values (L1-norms) instead of the L2 norms frequently used. The L1-norm does not cause large errors to excessively influence the solution like the L2-norm does, leading to a desirable solution. The mathematical framework of duality separated the optimisation procedure into several subproblems, where each subproblem is simple and computationally inexpensive. The smoothness subproblem was solved via TVL1 denoising [7].

A common theme throughout the literature is the estimation of optical flow via direct inspection and manipulation of the data. An alternative approach not thoroughly explored is the use of a warping operator (or transformation [8]), where one can manipulate and investigate the operator rather than directly examine the data itself. Phase correlation is a well known application of the Fourier shift theorem [9] to determine global translation in a digital signal. The Fourier shift theorem is an example of an operator to perform global operations of pixel shifting on signal and image data. The phase correlation method was localised by considering local Gaussian windows [10]. The nonequispaced Fourier transform [11] can be used to resample the location of data points, effectively warping a signal. The nonequispaced inverse Fourier transform then defines a matrix operator generalisation of the Fourier shift suitable for warping a signal or image.

In this paper the use of the nonequispaced inverse Fourier transform as a basis for estimating optical flow is investigated. The work herein constitutes a feasibility study to determine if the techniques described have merit. The target problem is to estimate the optical flow in large objects of near homogenous brightness that travel a moderate distance. In Sec. II the basic theory is developed from the concept of warping by convolution to the nonequispaced inverse Fourier transform. Algorithms for estimating optical flow as a warping function are described. In Sec. III experiments are described to probe the merit of the technique, the results of which are given and discussed in Sec. IV. In Sec. V conclusions are drawn, including possible directions for improving both speed and accuracy of the optical flow method.

II. Theory

Throughout, both continuous and discrete notation are used, where the choice of notation is a matter of convenience. It should be noted that all the development herein implies a discrete setting, whether directly or indirectly, and that we are restricted to the class of functions of finite support and range. To avoid ambiguity square brackets denote factorisation and curved brackets enclose arguments to a function.

[1] http://vision.middlebury.edu/flow/, date accessed 15/08/2011

Consider the warping of a signal $f(x)$ in one dimension via the convolution with a warping function $w(x+u(x))$, with $u(x)$ the warping signal, namely,

$$f(x+u(x)) = \int_{-\infty}^{\infty} \mathrm{d}x'\, f(x')w(x+u(x)-x'). \quad (1)$$

The action of the convolution is to pick up the value f at position $x+u(x)$ and to place that value at the position x. To perform warping in the conventional sense, w would take the form of a delta function, so we write Eq. 1 using the Fourier kernel delta sequence [9] as,

$$\begin{aligned} &f(x+u(x)) \\ &= \int_{-\infty}^{\infty} \mathrm{d}x'\, f(x') \int_{-\infty}^{\infty} \mathrm{d}k\, e^{i2\pi k[x+u(x)-x']}, \\ &= \int_{-\infty}^{\infty} \mathrm{d}k\, \hat{f}(k) e^{i2\pi k[x+u(x)]}, \end{aligned} \quad (2)$$

where $\hat{f}(k)$ is the Fourier transform of $f(x)$.

Eqn. 2 is a nonequispaced inverse Fourier transform [11]. In explicitly discrete form, let $W_{\mathbf{x}}$ be the equispaced Fourier transform matrix, and let $W_{\mathbf{x}+\mathbf{u}(\mathbf{x})}$ be the nonequispaced equivalent, then the discrete signal in vector form is $\mathbf{f}(\mathbf{x})$, and transforms as

$$\mathbf{f}(\mathbf{x}+\mathbf{u}(\mathbf{x})) = \frac{1}{N} W^{*}_{\mathbf{x}+\mathbf{u}(\mathbf{x})} W_{\mathbf{x}} \mathbf{f}(\mathbf{x}), \quad (3)$$

where the in W^* denotes a conjugate transpose operation on the matrix W. The matrix T such that,

$$T_{\mathbf{x}+\mathbf{u}(\mathbf{x})} = \frac{1}{N} W^{*}_{\mathbf{x}+\mathbf{u}(\mathbf{x})} W_{\mathbf{x}}, \quad (4)$$

then defines a warping transformation *operator*. Manipulation of the data in $\mathbf{f}$ is performed by manipulation of $\mathbf{u}(\mathbf{x})$ in T and then application of the corresponding operator T on $\mathbf{f}$.

A. Optimisation: The One Dimensional Case

Consider a second signal $g(x)$ which is related to $f(x)$ as

$$g(x) = f(x+u(x)). \quad (5)$$

Given $g(x)$ and $f(x)$ we want to determine $u(x)$ subject to the constraint that $u(x)$ is smooth. The Fourier warping inversion tends to produce spikey errors, so for smoothness in one dimension a local median filter [12] is used.

Define duality as an external constraint that two variables are equal to each other in a least squares sense. Let v(x) be an independent dual function of $u(x)$. Define a set of simultaneous minimisation problems via a duality relationship [6] to find $u(x)$:

$$\begin{aligned} \epsilon_1(u(x)) &= \frac{1}{2}\left[f(x+u(x)) - g(x)\right]^2 \\ &\quad + \frac{1}{2}\lambda_d \left[v(x)-u(x)\right]^2 \end{aligned} \quad (6a)$$

$$\epsilon_2(v(x)) = \lambda_s \mathrm{med}(v(x)) + [1-\lambda_s]\left[u(x)-v(x)\right]^2. \quad (6b)$$

where $\mathrm{med}(\cdot)$ denotes the median filter, and λ_s and λ_d weight the relative importance of each respective constraint against the duality. Of the optimisation objective equations ϵ_1 is the data constraint and ϵ_2 the smoothness.

In minimising ϵ_1, iterative least squares is employed with $v(x)$ held constant, that is,

$$u_{\text{new}}(x) = u(x) + U_d \Delta \quad (7)$$

where Δ is a user defined update rate parameter and

$$\begin{aligned} U_d &= \lambda_d \left[v(x)-u(x)\right] + \\ &\left[f(x) - g(x+u(x))\right] \int_{\infty}^{\infty} \mathrm{d}k\, i2\pi k \hat{g}(k) e^{i2\pi k[x+u(x)]}. \end{aligned} \quad (8)$$

The warping of the second frame onto the first avoids the possiblility of multiple points in the first frame mapping onto the same position in the second. For ϵ_2, $u(x)$ is held constant and iterations inspired by projection onto convex sets are employed: by simple interpolation between the current prime iteration and the filtered dual, that is,

$$v_{\text{new}}(x) = (1-\lambda_s)u(x) + \lambda_s \mathrm{med}\left(v(x)\right). \quad (9)$$

B. Optimisation: The Two Dimensional Case

Application of the Fourier transform warping to the optical flow problem in two dimensions requires special consideration in order to obtain a reasonable solution. In one dimension there are only two possible directions that the warping can operate in, either the positive or negative direction. In two dimensions there are effectively an infinite number of directions.

The issue arises when one extends equation 2 to the two dimensional case:

$$\begin{aligned} &f\left((x+u_x(x,y), y+u_y(x,y)\right) \\ &= \iint_{-\infty}^{\infty} \mathrm{d}l\, \mathrm{d}k\, \hat{f}(k,l) e^{i2\pi l[y+u_y(x,y)]} e^{i2\pi k[x+u_x(x,y)]}, \end{aligned} \quad (10)$$

where both integrals are improper integrals of the first kind. Immediately obvious is the fact that the flow vector terns u_x and u_y are inside the complex exponential, so there is no obvious weighting in terms of direction, other than to search in either the positive or negative direction in each of the x and y dimension.

A two stage process is devised, primarily to overcome the issue of search direction. The objective of the procedure is to gently obtain a y-direction estimate that is accurate enough that the x-direction estimate may find an accurate solution. For a set number of repetitions the following procedure is performed:

1) One iteration of the one-dimensional algorithm is performed in the y-direction, independently of x.
2) The output of the single y-direction iteration is accumulated to a total y-direction result and then smoothed via total variation minimisation [7], [6].
3) The original second frame is warped in the y-direction according to the current u_y, back towards the first frame.
4) The one-dimensional algorithm is run on the x-direction for a set number of iterations, and the result smoothed using total variation minimisation.

5) The original second frame is warped according to the current u_x and u_y, back towards the first frame.

The final estimate of u_x from the final repetition is kept. The frames are transposed and the procedure is repeated to obtain u_y.

III. Experimentation

The target problem for the optical flow method currently under discussion is objects with little texture that move a moderate distance between frames. For quantitave analysis a dataset that fits the target model was generated and examined.

A. Testing the Optical Flow Method

For each of one to five objects, one hundred frame pairs were generated with the corresponding ground truth optical flow. The frame size is 150×151 pixels, and each moving object is a square with random width between ten and fourty pixels. Each object moved in a random direction a distance equal to its width. For examples see Figs. 1–3 below. Random Gaussian noise with standard deviation 0.01 ($1/100$ of the brightness range) was added to each generated frame.

Objects entering and leaving the scene were strictly prevented. Object overlap was minimised, however an average overlap of objects, over each frame pair, of less than five percent was permitted. In some cases the limit of five percent was adequate for significant overlap, for example a small object might be completely lost behind a large object.

The optical flow estimation was run with the following parameters: $\Delta = \lambda_d = \lambda_s = 0.5$; the y-direction flow updates were performed with an update parameter 1; the TV smoothing had a smoothing coefficient of 50; thirty y-direction updates were performed, at each of which, thirty x-direction iterations were used; and a window size of 3×5 was used in the median filtering.

B. Error Metric and Reporting

For assessment of the error of the optical estimates the endpoint error (EE) and angle error (AE) were used [3, and references therein]. The EE is computed as

$$EE = \sqrt{(u_x - u_{x,t})^2 + (u_y - u_{y,t})^2}, \tag{11}$$

where $u_{x,t}$ and $u_{y,t}$ are the true flow, and the AE is computed as

$$AE = \cos^{-1}\left(\frac{1 + u_{x,t}u_x + u_{y,t}u_y}{\sqrt{1 + u_{x,t}^2 + u_{y,t}^2}\sqrt{1 + u_x^2 + u_y^2}}\right). \tag{12}$$

The error metrics are anaylsed by choosing a set of threshold values from zero to some maximum and calculating the fraction of error values that are above that threshold. For the EE we use thresholds from zero to forty pixels (the maximum true motion in the generated test set) in steps of one pixel. For the AE it is natural to use thresholds in from zero to π radians. The result are error curves, where the closer each curve is to zero along the y-axis for a given threshold value, the better the performance up to that error magnitude.

The ability of an optical flow estimator to identify homogenous static regions as zero flow is important, so the entire frame is incorperated in the computation of the error fraction.

IV. Results and Discussion

The Fourier transform optical flow was tested over one hundred frames for each of varying number of moving objects. For visual qualitative comparison the optical flow from a theory of warping with descriptor matching [4] and the TV-L1 optical flow [6] were also performed. A different parameter set was necessary for each frame pair to produce an optimal result for the other optical flow methods. A single parameter set misrepresents the power of those algorithms, so quantitative analysis was not performed. For the proposed method the same parameter set was used for each frame pair.

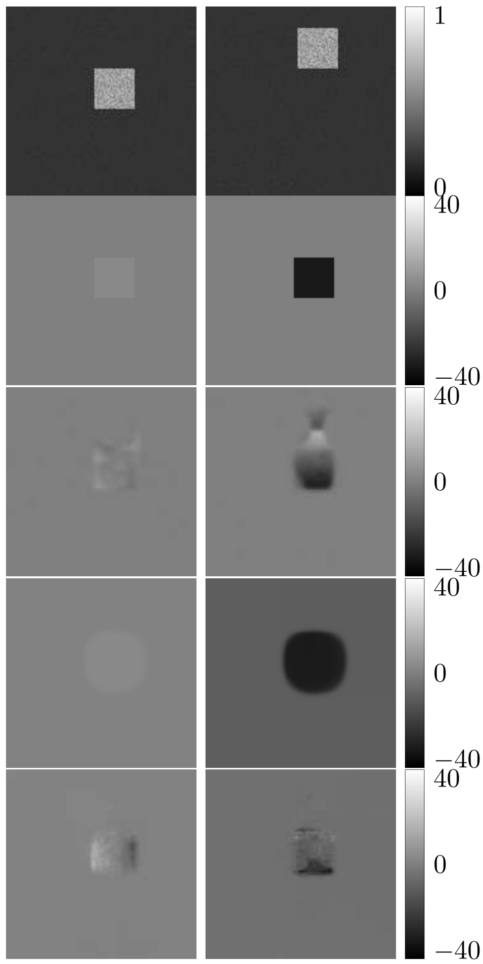

Fig. 1. An example optical flow result for one moving foreground object. From top to bottom: the two frames; the true flow vector x and y componenets; the proposed method; the theory of warping method; and the TVL1 method.

Some example frame pairs and flow estimates are shown in Figs. 1, 2 and 3. In each the frame pair, true flow and estimates from the proposed method from, the warping with descriptor matching, and the TVL1 optical flow are presented. The warping with descriptor matching produces remarkably accurate flow vector estimates within the region of the moving

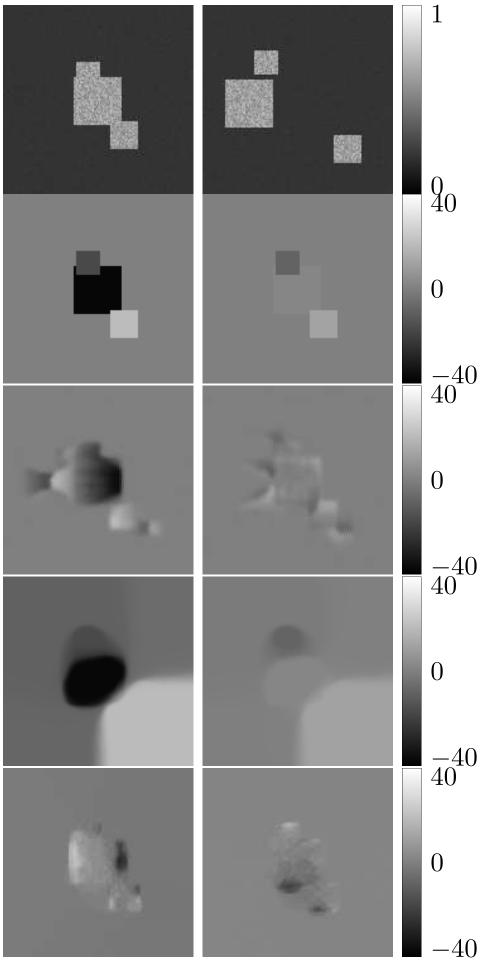

Fig. 2. An example optical flow result for three moving foreground objects. From top to bottom: the two frames; the true flow vector x and y components; the proposed method; the theory of warping method; and the TVL1 method.

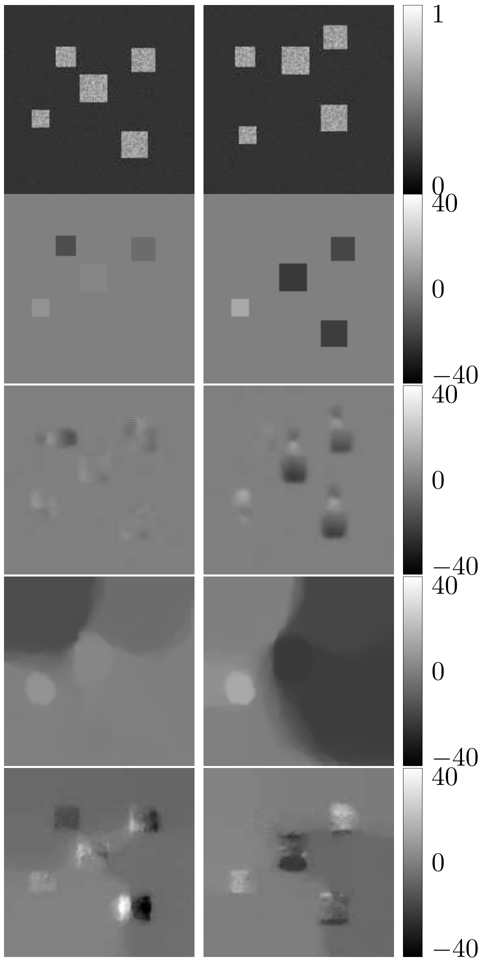

Fig. 3. An example optical flow result for five moving foreground objects. From top to bottom: the two frames; the true flow vector x and y components; the proposed method; the theory of warping method; and the TVL1 method.

objects, but does not keep within the boundaries of the moving objects. The TVL1 flow has not performed well in the examples shown.

The proposed method yields reasonable flow estimates, but with the outstanding issue that non zero flow estimates that clear the foreground object from the second frame are technically erroneous. Also the Fourier transform based warping optical flow simply finds the nearest pixel that matches the target value and then stops, as such it is prone to finding local minima. For the example frame pairs generated, the local minimum is frequently the nearest pixel of the same object in the second frame. The flow estimate tends to be accurate at the furtherest edge from the direction of motion, and linearly decreases in magnitude in the direction of motion.

These results do not minimise the power of the prior art methods considered, rather they highlight the difficulty of the target problem. The Middlebury data, which the prior methods perform well on, are generally too highly textured for the proposed method to obtain a good flow estimate. The higher spatial frequencies confuse the current incarnation of the Fourier transform optical flow method.

From the Average endpoint error (EE), Fig. 4, immediately we see that the error increases with the number of objects. The average area covered by the moving objects increases with the number of objects, indicating that the flow estimates in the homogenous static background regions are zero, as they should be.

It is instructive to consider the rate at which the error decays as the EE threshold is increased. In no case does the error come close to zero before the maximum threshold, indicating that there were some cases where the error was larger than the maximum flow considered of fourty pixels.

For the threshold range 0–12 pixels, the average error fraction decreases in a roughly parabolic manner, and above 12 pixels the error fraction decreases in a near linear fashion. The parabolic region is attibuted to the need for the warping to, not only copy the foreground objects in the second frame to the corresponding position in the first frame, but to also clear the forground objects from their positions in the second frame as well. To clear the foreground objects from the first frame the warping finds the nearest background pixel outside of the foreground region. It is no coincidence that the average

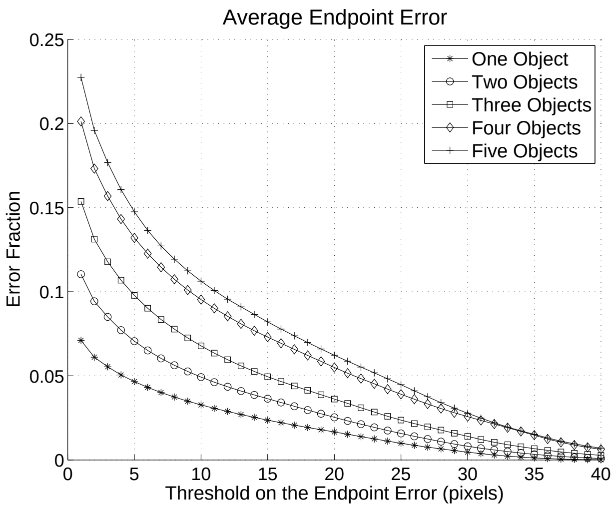

Fig. 4. Endpoint error curves with relation to an error threshold.

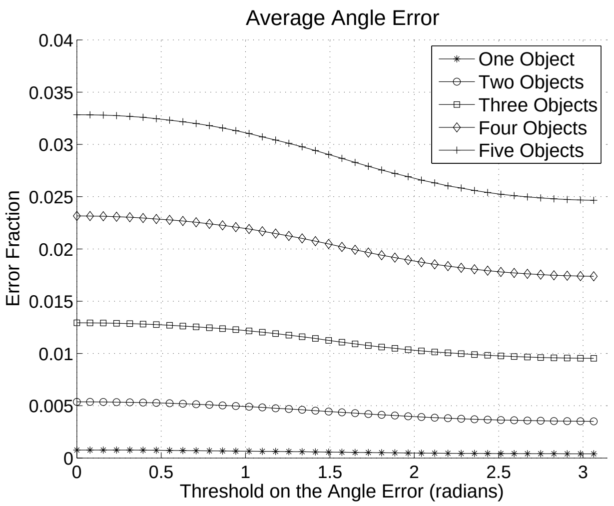

Fig. 5. Angle error curves with relation to an error threshold.

maximum distance from the center of each object to the nearest background pixel is 12.5 pixels. If the effect of clearing the foreground object from the first frame is neglected, then one deduces that the error fraction of the flow estimate is appoximately linear with the endpoint error threshold. The linear decrease in error is attributed to the local minima problem described above, causing a linear decrease in flow magnitude along the direction of motion.

The angle error (AE), Fig. 5, is consistently low, but naturally rises with the number of objects corresponding to the increase in total pixels that represent moving foreground objects. There is a marked difference in error fraction between the EE and the AE. So that we may see the shape of the error curves the y-axis scale used in Fig.5 is smaller that that of Fig. 4. The definition of the AE (Eq.12) causes regions of zero true flow but nonzero estimated flow (and vice versa) to to be ignored. The AE has underestimated the overall level of error and the EE is arguably more representative. Regardless there is still sufficient data where the AE could have been large, but is not, so the fact that the AE error fraction is consistently below 0.05 is encouraging.

Overall a principle difficulty encountered was that the data constraint, namely the warping by the Fourier transform itself, is different enough to the prior art that much of the wisdom [13] in the implementation of optical flow requires careful consideration. This is the reason behind using the simple optimisation method of iterative least squares. Median filtering, for example has been exposed as helpful in optimising the total energy of an estimated flow field, but here median filtering is essential to mitigate spiky errors generated by the data constraint itself. The mathematical formulation of the data constraint yields a procedure that searches for a minimum to match the local image data, where the primary cue for scanning is the magnitude of the difference between the frames. The procedure does not try to assign nonzero flow estimates where there is no difference in the image data, but care is required to determine the direction of the flow vector. No coarse to fine resolution proceedure was necessary for the test set used.

V. Conclusion

Optical flow from warping by the nonequispaced Fourier transform was proposed and tested. The target problem for the proposed optical flow estimator was low textured objects moving moderate distances in front of a flat background, an objective that is difficult for some optical flow approaches. Whilst poor results were obtained for some standard data commonly used in the image processing literature, reasonable success was achieved for the test set specifically generated for this experiment.

The optimisation routine devised was simple and deliberately so, that we might separate features of the algorithm that a dependent on the implementation from features of the mathematical formulation. The most noteworthy aspect is that the data constraint yeilds a procedure that simply scans to find the nearest matching point and then stops. The current algorithm is computationally inefficient. In future work mathematical analysis will transform the Fourier data constraint into a more practical solution, with focus on partially sparating the application of the warping from the Fourier transform itself by way of a series expansion.

VI. Acknowledgements

This work was supported by The University of Waikato Strategic Research Fund. Adrian Dorrington is acknowledged for providing the genesis and environment for the research.

References

[1] B. K. P. Horn and B. G. Schunck, “Determining optical flow,” tech. rep., Massachesetts Institute of Technology, 1980. AI Memo 572.

[2] B. D. Lucas and T. Kanade, “An iterative image registration technique with an application to stereo vision,” in *Proceedings of Imaging Understanding Workshop*, pp. 121–130, 1981.

[3] S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, and R. Szeliski, “A database and evaluation methodology for optical flow,” *International Journal of Computer Vision*, vol. 92, no. 1, pp. 1–31, 2011.

[4] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, “High accuracy optical flow estimation based on a theory for warping,” in *8th European Conference on Computer Vision*, (Prague, Czech Republic), pp. 25–36, May 2004.

[5] T. Brox and J. Malik, “Large displacement optical flow: Descriptor matching in variational motion estimation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 3, pp. 500–513, 2011.

[6] C. Zach, T. Pock, and H. Bischof, “A duality based approach for realtime tv-l1 optical flow,” in *Proceedings of the 29th DAGM conference on Pattern recognition*, (Berlin, Heidelberg), pp. 214–223, Springer-Verlag, 2007.

[7] A. Chambolle, “An algorithm for total variation minimization and applications,” *Journal of Mathematical Imaging and Vision*, vol. 20, no. 1-2, pp. 89–97, 2004.

[8] S. M. Seitz and S. Baker, “Filter flow,” in *2009 IEEE 12th International Conference on Computer Vision*, pp. 143–150, 2009.

[9] R. N. Bracewell, *The Fourier Transform and Its Applications*. Boston, Massachusetts, USA: McGraw-Hill, 1986.

[10] F. Pla and M. Bober, “Estimating translation/deformation motion through phase correlation,” in *Proceedings of the 9th International Conference on Image Analysis and Processing-Volume I*, pp. 653–660, 1997.

[11] D. Potts, G. Steidl, and M. Tasche, *Modern Sampling Theory: Mathematics and Applications*, ch. 12 Fast Fourier transforms for nonequispaced data: A tutorial, pp. 249–274. Applied and Numerical Harmonic Analysis Series, Birkhauser Boston, 2001.

[12] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Upper Saddle River, New Jersey, USA: Prentice-Hall, second edition ed., 2002.

[13] D. Sun, S. Roth, and M. Black, “Secrets of optical flow estimation and their principles,” in *2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (San Francisco, CA, USA), pp. 2432 – 2439, 2010.