Article

Error Analysis and Condition Estimation of the Pyramidal Form of the Lucas-Kanade Method in Optical Flow

Joab R. Winkler

Department Computer Science, The University of Sheffield, Regent Court, 211 Portobello, Sheffield S1 4DP, UK; j.r.winkler@sheffield.ac.uk; Tel.: +44-(0)114-2221834

Abstract: Optical flow is the apparent motion of the brightness patterns in an image. The pyramidal form of the Lucas-Kanade (LK) method is frequently used for its computation but experiments have shown that the method has deficiencies. Problems arise because of numerical issues in the least squares (LS) problem $\min\|Ax - b\|_2^2$, $A \in \mathbb{R}^{m \times 2}$ and $m \gg 2$, which must be solved many times. Numerical properties of the solution $x_0 = A^\dagger b = (A^T A)^{-1} A^T b$ of the LS problem are considered and it is shown that the property $m \gg 2$ has implications for the error and stability of x_0. In particular, it can be assumed that b has components that lie in the column space (range) $\mathcal{R}(A)$ of A, and the space that is orthogonal to $\mathcal{R}(A)$, from which it follows that the upper bound of the condition number of x_0 is inversely proportional to $\cos\theta$, where θ is the angle between b and its component that lies in $\mathcal{R}(A)$. It is shown that the maximum values of this condition number, other condition numbers and the errors in the solutions of the LS problems increase as the pyramid is descended from the top level (coarsest image) to the base (finest image), such that the optical flow computed at the base of the pyramid may be computationally unreliable. The extension of these results to the problem of total least squares is addressed by considering the stability of the optical flow vectors when there are errors in A and b. Examples of the computation of the optical flow demonstrate the theoretical results, and the implications of these results for extended forms of the LK method are discussed.

Keywords: optical flow; Lucas-Kanade; condition estimation; Gaussian pyramid

1. Introduction

Optical flow is the apparent motion of the brightness patterns in an image. This is different from the motion field, which is the projection onto two dimensions of motion in a three dimensional environment. Both these motions are usually termed optical flow even though they are not necessarily the same ([§2] [1]). There are several methods for the computation of optical flow and it is assumed in all the methods that it is locally smooth. The methods can be classified as global, for example, the Horn-Schunck method [2], or local, for example, the Lucas-Kanade (LK) method [3], and a review of these methods, and other methods, is in [4–8]. Local methods are more robust to noise but they do not yield a dense optical flow field, and global methods are more sensitive to noise but they yield a dense optical flow field. The computation of the optical flow has been the focus of much research, particularly in images that are corrupted by significant noise, or images in which two or more objects move, or images in which there are temporal changes in illumination [9].

Comparisons of the different methods for the computation of optical flow are in [1,10]. The conclusions from these comparisons are empirical and based on quantitative measures of the errors in the computed results, but numerical and linear algebraic considerations of the methods are not included in these comparisons. These issues are addressed in this paper by error analysis and condition estimation of the pyramidal form of the LK method. This method requires that many least squares (LS) problems be solved at each level of the pyramid and it is shown that, even in the absence of occlusions, motion discontinuities and other problematic phenomena, the maximum values of the errors and condition numbers

of the solutions of the LS problems (the optical flow) increase as the pyramid is descended from the top level to the bottom level.

Kearney, Thompson and Boley [11] consider condition estimation and error analysis for the computation of the optical flow from two points with pixel coordinates $p_i = (x_i, y_i)$ and $p_j = (x_j, y_j)$ at time t, where p_i is near p_j such that a first order Taylor approximation is appropriate. This approximation yields

$$G\omega = -b, \tag{1}$$

where

$$G = \begin{bmatrix} I_x^{(i)} & I_y^{(i)} \\ I_x^{(j)} & I_y^{(j)} \end{bmatrix}, \qquad \omega = \begin{bmatrix} u \\ v \end{bmatrix}, \qquad b = \begin{bmatrix} I_t^{(i)} \\ I_t^{(j)} \end{bmatrix},$$

$$I_x^{(i)} = \frac{\partial I^{(i)}}{\partial x}, \qquad I_y^{(i)} = \frac{\partial I^{(i)}}{\partial y}, \qquad I_t^{(i)} = \frac{\partial I^{(i)}}{\partial t},$$

ω is the optical flow and $I^{(i)} = I^{(i)}(x, y, t)$ is the intensity of the ith pixel. The work in this paper follows the analysis in [11] because issues of computational linear algebra, specifically condition estimation and error analysis, are addressed. There are, however, differences because an arbitrary number of points, rather than two points, is considered in this paper since the LK method yields an overdetermined set of equations, rather than the problem (1), for which the coefficient matrix is square and of order two. It is therefore necessary to compute x_0, the optical flow,

$$x_0 = \arg\min_x \|Ax - b\|^2, \qquad A \in \mathbb{R}^{m \times n}, \tag{2}$$

where $\|\cdot\| = \|\cdot\|_2$, $m > n$ and rank $A = n$, for each window at each level of the pyramid. The optical flow problem requires that $n = 2$ but it is assumed in the theoretical analysis that $n \leq m$ is arbitrary.

The first and second columns of A contain, respectively, the intensity gradients with respect to x and y, and b contains the negative of the temporal intensity gradients, and errors are therefore present in A and b, such that it is necessary to solve the total least squares (TLS) problem ([§6.3] [12]),

$$\{x_{tls}, \delta A_{tls}, \delta b_{tls}\} = \arg\min_{x, \delta A, \delta b} \left\| \begin{bmatrix} \delta A & \delta b \end{bmatrix} \right\|_F$$

$$subject\,to \quad (A + \delta A)x = b + \delta b, \tag{3}$$

where $\|\cdot\|_F$ denotes the Frobenius norm. The method of TLS has been used for the computation of optical flow [13], but an estimate of its numerical stability, which is a measure of its sensitivity to noise, has not been established. This measure is called the condition number and this paper addresses this issue by considering the condition number of the computed optical flow. This condition number must be derived for the solution x_{tls} of the TLS problem (3), but it is shown that it is first necessary to consider the condition number of the solution x_0 of the LS problem (2).

The results of this paper are summarised:

1. It is shown that the maximum value of the error in the solution x_0 of (2) increases as the pyramid is descended from the top level to the lowest level.

2. The condition number $\kappa(A) = \|A\|\|A^\dagger\|$ of A, where $A^\dagger = (A^T A)^{-1} A^T$, is a function of A only but the solution x_0 of the LS problem is a function of A and b. A refined condition number, called the effective condition number $\eta(A, b)$ that is a function of A and b, is therefore developed.

3. An upper bound of $\eta(A, b)$ is $\kappa(A)/\cos\theta$, where θ is the angle between b and its component that lies in the column space (range) $\mathcal{R}(A)$ of A. It follows that even if A is well conditioned, the solution x_0 may be ill conditioned if θ is large.

4. The dimensions of the coefficient matrix A satisfy $m \gg n$ and it can therefore be assumed that a significant component of b lies in the space that is orthogonal to $\mathcal{R}(A)$. It is shown that the error in x_0 is therefore large and that there is a simple relationship between this error and $\cos\theta$. It is concluded that the angle θ arises in the error and condition numbers of x_0, and it is therefore an important measure to consider when assessing the computational reliability of the optical flow.

5. It is shown that the maximum values of the condition numbers and errors in the solutions of the LS problems (one LS problem for each window at each level of the pyramid) increase as the pyramid is descended from the coarsest image to the finest image. This shows that the advantage of the pyramidal implementation of the LK method - the ability to cope with large displacements between successive images - must be balanced against the numerical measures that quantify the robustness and reliability of the solutions of the many LS problems that are solved in the LK method.

6. The condition number $\kappa(A)/\cos\theta$ of x_0 is limited because it quantifies the effect on x_0 of errors in b, but errors in A are not considered. A condition number $\rho(A, b)$ of x_0 that includes errors in A and b is developed because it is more appropriate for the LK method. This is a non-linear condition number, in contrast to the linear condition number $\kappa(A)/\cos\theta$, and it is shown that the dominant term in the expression for $\rho(A, b)$ is $\kappa(A)\eta(A, b)$.

The method of least median of squares (LMedS) is used in [14] to solve simultaneously the problems of temporal variation of illumination and motion discontinuities induced by the relative motion of two or more objects. The LMedS method requires the solution of the LS problem (2), and the pyramidal form of the LK method is also used in feature tracking [15]. More generally, the theory and results in this paper are applicable to all problems in which the LK method is used. There are several forms of the LK method and they differ in the expressions used for the calculation of the derivatives of the intensity. For example, one form of the LK method requires that A and b in (2) are premultiplied by a diagonal matrix, such that a larger weight is assigned to pixels in the centre of each window, and a smaller weight is assigned to pixels at its borders ([§4.3.2] [16]).

The LK method is described in Section 2 and the error in the solution x_0 of the LS problem is considered in Section 3. It is shown that there is a simple relationship between $\cos\theta$ and the error, and an expression for $\cos\theta$ is developed in Section 4. Linear and non-linear condition numbers that consider errors in b, and errors in A and b, respectively, of the optical flow are considered in Section 5. Examples that show the errors and condition numbers of the optical flow at each level of the pyramid are in Section 6. The paper is summarised in Section 7 and it is shown that the numerical problems with the solution of the LS problem that are addressed in this paper provide the mathematical motivation for the generalised form of the LK method that includes warping functions [17].

2. The Lucas-Kanade Method

The LK method is derived from the assumption of constant intensity $I(x, y, t)$ as a pixel at position (x, y) at time t moves to position $(x + \delta x, y + \delta y)$ at time $t + \delta t$, such that

$$I(x, y, t) = I(x + \delta x, y + \delta y, t + \delta t),$$

to first order. It therefore follows that

$$\frac{\partial I}{\partial x}\delta x + \frac{\partial I}{\partial y}\delta y + \frac{\partial I}{\partial t}\delta t = 0, \tag{4}$$

and thus if

$$u = \frac{\partial x}{\partial t} \quad and \quad v = \frac{\partial y}{\partial t},$$

then

$$I_x u + I_y v + I_t = 0, \tag{5}$$

where

$$I_x = \frac{\partial I}{\partial x}, \qquad I_y = \frac{\partial I}{\partial y} \quad and \quad I_t = \frac{\partial I}{\partial t}.$$

Equation (5) is one equation in two unknowns (the components u and v of the optical flow), and it therefore possesses an infinite number of solutions. This problem of a non-unique solution is addressed by considering a square window in each image at times t and $t + \delta t$, and if each window contains p points, then the application of (5) to each of these points yields,

$$I_x^{(i)} u + I_y^{(i)} v + I_t^{(i)} = 0, \qquad i = 1, \ldots, p,$$

which can be combined into one equation,

$$Ax = b, \quad A \in \mathbb{R}^{p \times 2}, \quad x \in \mathbb{R}^2 \quad and \quad b \in \mathbb{R}^p. \tag{6}$$

Equation (6) is solved in the LS sense (2), which yields the solution,

$$x_0 = A^\dagger b, \qquad A^\dagger = \left(A^T A \right)^{-1} A^T, \tag{7}$$

where $x_0 = [u\, v]^T$ is the optical flow. It follows from (6) that the motion in each window is a translation only, where u and v represent its horizontal and vertical components, respectively.

The solution x_0 assumes there are errors in b only, and that A does not contain errors. The qualifications that follow from this assumption must be addressed because A and b contain, respectively, the spatial and temporal derivatives of the intensity $I(x, y, t)$. The inclusion of errors in A and b requires that the TLS problem be considered ([§6.3] [12]), and the effect of perturbations in A and b on the optical flow x_0 is considered by the development of a non-linear condition number of x_0 in Section 5.2.

The computation of the translation x_0 between two images of a dynamic scene occurs in several applications, including remote sensing and medical imaging, and it has been shown that noise and the first order approximation (4) introduce bias in the estimate of x_0 [11,18–20]. Robinson and Milanfar [20] use prior knowledge of the image and the translation to design gradient-estimation filters but the bias due to noise is not considered, and thus the LK method yields poor results for images whose signal-to-noise ratio is less than about 25 dB. Pham et al. [19] reduce the bias by iteratively computing the translation and upsampling the matrices that store the components u and v of the optical flow, which is the pyramid procedure proposed by Lucas and Kanade [3], and refined by Baker and Matthews [17]. Pham et al. [19] report that a pyramid with three levels yields an optical flow field with a very small bias.

The first order approximation (4) requires that

$$|\delta x|, |\delta y|, |\delta t| \ll 1, \tag{8}$$

and if this condition is not satisfied, a pyramid is formed in which the images at the *i*th level are formed by passing the images at the $(i-1)$th level through a low pass filter and then downsampling the filtered images. The coefficients of the low pass filter are [21,22],

$$\left(\frac{1}{16}\right)\begin{bmatrix} 1 \\ 4 \\ 6 \\ 4 \\ 1 \end{bmatrix}\left(\frac{1}{16}\right)\begin{bmatrix} 1 & 4 & 6 & 4 & 1 \end{bmatrix}, \tag{9}$$

which is approximately equal to a Gaussian filter with standard deviation $\sigma = 1$.

The pyramidal implementation is shown in Figure 1 for two images *F* and *G* that are taken at times *t* and $t + \delta t$ respectively. The optical flow is computed at the top of the pyramid, level 4 (the coarsest image) in Figure 1, which is the initial condition for the computation of the optical flow at level 3 of the pyramid. It follows from (8) that (5) yields good estimates of *u* and *v* if the translation between the images at times *t* and $t + \delta t$ is small. Only the increments between successive levels of the pyramid are therefore calculated, after warping the two images based on the flow field estimate from the coarse level. This operation yields two matrices, one for each component of the optical flow, and each matrix is upsampled to yield matrices of twice the number of rows and columns, and then passed through a modified form of the filter (9),

$$\left(\frac{1}{8}\right)\begin{bmatrix} 1 \\ 4 \\ 6 \\ 4 \\ 1 \end{bmatrix}\left(\frac{1}{8}\right)\begin{bmatrix} 1 & 4 & 6 & 4 & 1 \end{bmatrix}. \tag{10}$$

This process is continued to the base of the pyramid, level 1, and the optical flow at this level is the desired optical flow.
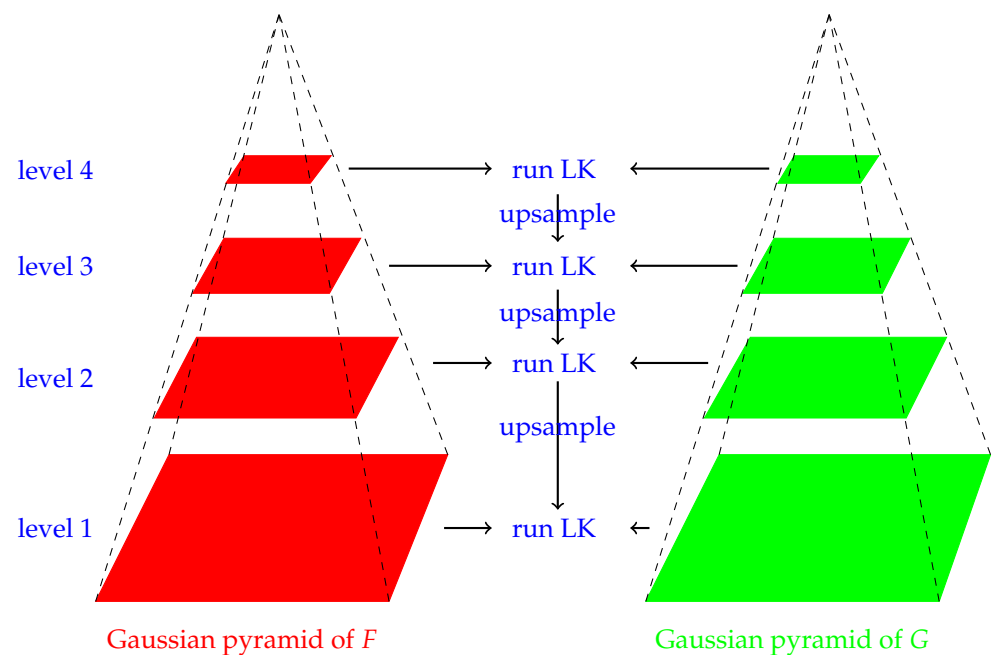
Figure 1. The pyramid implementation of the Lucas-Kanade method.

The difference in the scale factors for the filters for downsampling the images and upsampling the optical flow field arises because all the pixels in the images that are

downsampled are included in the computation, but rows and columns of zeros are added to the matrices that store the components u and v before they are upsampled, such that a matrix of order $M \times N$ is increased to a matrix of order $2M \times 2N$. The computation of all the pixel values in these larger matrices cannot include the added rows and columns of zeros and thus only one quarter of the pixels in the matrices of order $2M \times 2N$ are included in the computation.

The matrices of the components of the optical flow must be upsampled by the filter (10) and then multiplied by two in order to preserve the values of its components between successive levels of the pyramid. In particular, two adjacent pixels have unit separation before upsampling, but their separation increases to two after upsampling. The preservation of the values of the components of the optical flow in the upsampled images therefore requires that they be multiplied by two.

The discussion above shows that the pyramidal form of the LK method requires that the LS problem be solved many times at each level of the pyramid, and thus the error in its solution must be considered. This issue is addressed in Section 3 and it is shown in Section 4 that this error is closely related to the value of the angle θ between b and its component that lies in the column space $\mathcal{R}(A)$ of A. The computed optical flow must be numerically stable and it is therefore necessary to consider its numerical condition. This issue is addressed in Section 5, and linear and non-linear condition numbers are considered.

3. The Error in the Solution of the LS Problem

This section considers the error e in the solution of the LS problem (7),

$$e = \frac{\|b - Ax_0\|}{\|b\|} = \frac{\|(I - AA^\dagger)b\|}{\|b\|} = \frac{\|(I - \Sigma\Sigma^\dagger)c\|}{\|c\|}, \tag{11}$$

where the singular value decomposition of A is $U\Sigma V^T$, U and V are orthogonal matrices of orders m and n respectively,

$$\Sigma = \left[\begin{array}{c} \Sigma_1 \\ 0 \end{array} \right] \in \mathbb{R}^{m \times n}, \qquad \Sigma_1 = diag \left[\begin{array}{cccc} \sigma_1 & \sigma_2 & \cdots & \sigma_n \end{array} \right],$$

the singular values σ_i satisfy $\sigma_i \geq \sigma_{i+1}, i = 1, \ldots, n-1$, and

$$c = U^T b. \tag{12}$$

The matrix U is partitioned into two matrices U_1 and U_2,

$$U = \left[\begin{array}{cc} U_1 & U_2 \end{array} \right], \quad U_1 \in \mathbb{R}^{m \times n}, \quad U_2 \in \mathbb{R}^{m \times (m-n)},$$

and it follows from the orthogonal property of U that

$$\begin{array}{ll} U_1^T U_1 = I_n, & U_2^T U_2 = I_{m-n}, \\ U_1^T U_2 = 0_{n,m-n}, & U_2^T U_1 = 0_{m-n,n}, \end{array} \tag{13}$$

and

$$U_1 U_1^T + U_2 U_2^T = I_m. \tag{14}$$

It follows from (12) that c can be written as

$$c = \left[\begin{array}{c} \bar{c}_1 \\ \bar{c}_2 \end{array} \right] = \left[\begin{array}{c} U_1^T \\ U_2^T \end{array} \right] b, \quad \bar{c}_1 \in \mathbb{R}^n, \quad \bar{c}_2 \in \mathbb{R}^{m-n},$$

where $\bar{c}_1$ and $\bar{c}_2$ are the components of b that lie in, respectively, the spaces spanned by the columns of U_1 and U_2. It follows from (11) that

$$e^2 = \frac{\sum_{i=n+1}^m c_i^2}{\sum_{i=1}^m c_i^2} = \frac{\|\bar{c}_2\|^2}{\|c\|^2} = \frac{\|\bar{c}_2\|^2}{\|\bar{c}_1\|^2 + \|\bar{c}_2\|^2} = 1 - \cos^2\theta, \qquad (15)$$

where

$$\cos\theta = \frac{\|\bar{c}_1\|}{\|c\|}, \qquad (16)$$

and thus the error in x_0 is a measure of the proportion of b that lies in the space that is orthogonal to $\mathcal{R}(A)$, and a large value of θ leads to a large error in x_0. The significance of θ is considered in Section 4 and it is shown that it allows a geometric interpretation of the conditions that define the minimum and maximum values of e.

4. The Angle between b and $\mathcal{R}(A)$

It is shown in this section that θ, where $\cos\theta$ is defined in (16), is the angle between b and its component that lies in the column space $\mathcal{R}(A)$ of A. This interpretation follows from the decomposition of b into a component $r_1 \in \mathcal{R}(A)$ and a component $r_2 \in \mathcal{R}(A)^\perp$,

$$b = r_1 + r_2.$$

Since $r_1 \in \mathcal{R}(A)$ and $r_2 \in \mathcal{R}(A)^\perp$, there exist vectors t_1 and t_2 such that

$$r_1 = U_1 t_1 \quad \text{and} \quad r_2 = U_2 t_2,$$

because the columns of U_1 and U_2 form orthonormal bases for $\mathcal{R}(A)$ and $\mathcal{R}(A)^\perp$, respectively, and thus

$$b = U_1 t_1 + U_2 t_2.$$

It follows from (13) that premultiplication of this equation by U_1^T and then U_2^T yields

$$b = U_1 U_1^T b + U_2 U_2^T b,$$

where the first and second terms on the right are, respectively, the orthogonal projections of b onto $\mathcal{R}(A)$ and $\mathcal{R}(A)^\perp$. It follows from (13) that the second term is orthogonal to the first term, which confirms that r_1 is orthogonal to r_2. Also, it follows from (14) that

$$\left\| U_1 U_1^T b \right\|^2 + \left\| U_2 U_2^T b \right\|^2 = \|b\|^2,$$

where

$$\left\| U_1 U_1^T b \right\| = \left\| U_1^T b \right\| = \|\bar{c}_1\|,$$

and thus $\cos\theta$ is equal to the ratio of the magnitude of the orthogonal projection of b onto $\mathcal{R}(A)$ to the magnitude of b. Equation (15) shows that there is a simple relationship between the error in x_0, and the angle between b and its component that lies in $\mathcal{R}(A)$. The limiting cases are:

1. $b \in \mathcal{R}(A)$: It follows that $\cos\theta = 1$ and $U_2 = 0$, and thus $e = 0$.
2. $b \in \mathcal{R}(A)^\perp$: It follows that $\cos\theta = 0$ and $U_1 = 0$, and thus $e = 1$.

The ratio $\cos\theta$ provides a geometric interpretation of the error in the solution of the LS problem, and it is shown in Section 5 that $\cos\theta$ also arises in the expressions for the condition numbers of x_0.

5. The Numerical Condition of the Optical Flow

This section considers the numerical condition of the optical flow x_0, the solution of the LS problem. The condition number of A, $\kappa(A) = \sigma_1/\sigma_2$, where $\sigma_1 > \sigma_2$, is frequently used as a measure of the stability of x_0, but it requires qualification:

1. The condition number $\kappa(A)$ is a function of A and it is independent of b, but $x_0 = A^\dagger b$ is a function of A and b. This independence of b follows because $\kappa(A)$ is the upper bound of the ratio of the relative error in $x_0 = A^\dagger b$ to the relative error in b with respect to all vectors $b \in \mathcal{R}(A)$ and perturbations $\delta b \in \mathbb{R}^m$. The vector b is specified in a given problem and thus a better measure of the stability of x_0 is an upper bound of the ratio Δ of the relative error in x_0 to the relative error in b with respect to all perturbations δb, for the given vector b. The condition number $\kappa(A)$ may therefore overestimate the true value of Δ by many orders of magnitude if $b \in \mathcal{R}(A)$.

2. The property $b \in \mathcal{R}(A)$ is not, in general, satisfied in the LK method and thus a measure of the stability of x_0 with respect to perturbations in b that is appropriate when $b \notin \mathcal{R}(A)$ must be developed.

Although these points show that $\kappa(A)$ as a measure of the stability of x_0 has disadvantages, it has been used to assess the computational reliability of x_0 ([§II.A] [14]), ([§4.3.2] [16]), ([§2] [10]). These disadvantages of $\kappa(A)$ are overcome by the effective condition number of x_0, which is a function of A and b, and therefore a more accurate measure of the stability of x_0. This condition number is considered in Section 5.1.

The effective condition number is a linear condition number because it is assumed that the entries of A (the spatial derivatives of the intensity) are not subject to error, and only errors in b (the temporal derivatives of the intensity) are subject to error. A more realistic condition number of the optical flow requires that errors in the spatial and temporal derivatives of the intensity be considered. This inclusion leads to a non-linear extension of the effective condition number, and it is considered in Section 5.2.

5.1. The Effective Condition Number

This section considers the effective condition number of the solution x_0 of the LS problem, which must be solved many times in the LK method. It is assumed for generality that the coefficient matrix A is of order $m \times n$ where $m \geq n$, rather than $m \times 2$.

The relative errors of x_0 and b are, respectively,

$$\Delta x_0 = \frac{\|\delta x_0\|}{\|x_0\|} \qquad and \qquad \Delta b = \frac{\|\delta b\|}{\|b\|}, \tag{17}$$

and the effective condition number of x_0 is defined as the maximum value of the ratio of Δx_0 to Δb for the given vector b, with respect to all perturbations δb in b,

$$\eta(A, b) = \max_{\delta b \in \mathbb{R}^m} \frac{\Delta x_0}{\Delta b}.$$

Theorem 1. *The effective condition number $\eta(A, b)$ of x_0 is*

$$\eta(A, b) = \frac{\|A^\dagger\| \|b\|}{\|x_0\|} = \frac{\|A^\dagger\| \|b\|}{\|A^\dagger b\|}, \tag{18}$$

and it follows from (12) that

$$\eta(A, b) = \frac{\|c\|}{\sigma_n \|\Sigma^\dagger c\|}. \tag{19}$$

Proof. It follows from (7) and (17) that

$$\|\delta x_0\| = \left\|A^\dagger \delta b\right\| \le \left\|A^\dagger\right\| \|\delta b\| = \left\|A^\dagger\right\| \|b\| \Delta b. \tag{20}$$

The division of this inequality by $\|x_0\| = \left\|A^\dagger b\right\|$ yields (18), from which (19) follows. □

The condition that must be satisfied for $\kappa(A)$ to be a measure of the stability of x_0 follows from (20) because $\|b - Ax_0\| = 0$ if and only if $b \in \mathcal{R}(A)$. This condition is necessarily satisfied by all vectors b if A is square, but by some vectors b, not all vectors b, if $m > n$. If $b \in \mathcal{R}(A)$, then $b = Ax_0$ and thus $\|b\| \le \|A\| \|x_0\|$, and (20) yields

$$\|\delta x_0\| \le \left\|A^\dagger\right\| \|b\| \Delta b \le \left\|A^\dagger\right\| \|A\| \|x_0\| \Delta b.$$

This leads to the definition of the condition number $\kappa(A)$ of $A \in \mathbb{R}^{m \times n}$, $m \ge n$, as the upper bound of the ratio of the relative error of x_0 to the relative error of b if and only if $b \in \mathcal{R}(A)$,

$$\kappa(A) = \max_{\delta b \in \mathbb{R}^m, b \in \mathcal{R}(A)} \frac{\Delta x_0}{\Delta b} = \left\|A^\dagger\right\| \|A\| = \frac{\sigma_1}{\sigma_n}.$$

Furthermore, it follows from (12) and (19) that

$$\max_{b \in \mathcal{R}(A)} \eta(A, b) = \max_{c \in \mathcal{R}(A)} \frac{\|c\|}{\sigma_n \|\Sigma^\dagger c\|} = \frac{\sigma_1}{\sigma_n},$$

and thus $\kappa(A)$ is the maximum value of $\eta(A, b)$ with respect to all vectors $b \in \mathcal{R}(A)$.

The superiority of the effective condition number $\eta(A, b)$ with respect to the condition number $\kappa(A)$ as a measure of the stability of x_0 follows because $\eta(A, b)$ is a function of A and b, but $\kappa(A)$ is a function of A only. The example of regression in ([§3] [23]) shows the difference between these condition numbers. The dependence of $\eta(A, b)$ on A and b is clearly advantageous, but $\eta(A, b)$ must be used with care because it follows from (18) that the denominator contains the term $\|x_0\|$. It therefore follows that $\eta(A, b)$ is ill conditioned if x_0 is ill conditioned, and it is well conditioned if x_0 is well conditioned. It is shown in ([§4] [24]) that $\eta(A, b)$, and therefore x_0, are ill conditioned if the discrete Picard condition is satisfied ([§4.5] [25]). This condition states that if the constants $|c_i|$ decay to zero faster than the singular values σ_i decay to zero, that is,

$$\frac{|c_i|}{\sigma_i} \to 0 \quad \text{as} \quad i \to n, \tag{21}$$

then x_0 is ill conditioned. This condition can be derived from x_0,

$$x_0 = A^\dagger b = \sum_{i=1}^n \left(\frac{c_i}{\sigma_i}\right) v_i, \tag{22}$$

where v_i is the ith column of V. If b is perturbed to $b + \delta b$, then x_0 is perturbed to

$$x_0 + \delta x_0 = \sum_{i=1}^n \left(\frac{c_i + \delta c_i}{\sigma_i}\right) v_i, \tag{23}$$

and there exists a constant s such that the perturbations $|\delta c_i| \approx \epsilon, i = 1, \ldots, n$, satisfy

$$\begin{aligned} |c_i| &> \epsilon, & i &= 1, \ldots, s, \\ |c_i| &< \epsilon, & i &= s+1, \ldots, n. \end{aligned}$$

These inequalities follow because the perturbations $|\delta c_i|$ are approximately constant but the coefficients $|c_i|$ decay to zero because the discrete Picard condition (21) is satisfied, and thus from (23),

$$x_0 + \delta x_0 = \sum_{i=1}^{n} \left(\frac{c_i + \delta c_i}{\sigma_i} \right) v_i \approx \sum_{i=1}^{s} \left(\frac{c_i}{\sigma_i} \right) v_i + \sum_{i=s+1}^{n} \left(\frac{\delta c_i}{\sigma_i} \right) v_i.$$

Since the discrete Picard condition is satisfied, x_0 can be approximated with a small error by only considering the first s singular values in (22),

$$x_0 = A^\dagger b \approx \sum_{i=1}^{s} \left(\frac{c_i}{\sigma_i} \right) v_i,$$

and thus

$$\delta x_0 \approx \sum_{i=s+1}^{n} \left(\frac{\delta c_i}{\sigma_i} \right) v_i.$$

It follows that

$$\|\delta x_0\| \approx \epsilon \left(\sum_{i=s+1}^{n} \frac{1}{\sigma_i^2} \right)^{\frac{1}{2}} \approx \frac{\epsilon}{\sigma_n},$$

and thus the norm of the perturbation δx_0 is approximately proportional to the magnitude of the noise and inversely proportional to the smallest singular value of A, from which it follows that x_0 is ill conditioned if the discrete Picard condition (21) is satisfied. It is seen that x_0 is ill conditioned if it is dominated by the large singular values of A, in which case δx_0 is dominated by the small singular values of A.

The satisfaction of the discrete Picard condition is necessary for the application of Tikhonov regularisation because it guarantees that the regularisation error is small and the regularised form of x_0 is numerically stable ([§5] [23]). Tikhonov regularisation is used for the removal of blur from images because experiments have shown that many images satisfy the discrete Picard condition [26]. This condition forms the prior information that guarantees that Tikhonov regularisation is effective for image deblurring, and more generally, the ill conditioned property of $\eta(A, b)$ limits its practical use if prior information is not available. The condition number $\eta(A, b)$ is, however, an important condition measure because, as shown above, it defines the conditions between A and b for which x_0 is well conditioned, and the conditions for which x_0 is ill conditioned.

It has been shown that $\eta(A, b)$ is ill conditioned if x_0 is ill conditioned, but its upper bound is numerically stable, and an expression for this bound is developed in Theorem 2. This bound includes the term $\cos \theta$ that is defined in (16), and thus the error in the solution x_0 of the LS problem is related to the numerical condition of x_0.

Theorem 2. *If $A \in \mathbb{R}^{m \times n}$, $m \geq n$, then the effective condition number (19) of the solution x_0 of the LS problem is*

$$\eta(A, b) = \frac{1}{\sigma_n} \left(\frac{\sum_{i=1}^{m} c_i^2}{\sum_{i=1}^{n} \left(\frac{c_i}{\sigma_i} \right)^2} \right)^{\frac{1}{2}} = \left(\frac{\sigma_1}{\sigma_n} \right) \left(\frac{\sum_{i=1}^{m} c_i^2}{\sum_{i=1}^{n} \left(\frac{\sigma_1}{\sigma_i} \right)^2 c_i^2} \right)^{\frac{1}{2}},$$

where $c = \{c_i\}_{i=1}^{m}$ is defined in (12), and thus $\eta(A, b)$ satisfies

$$\eta(A, b) \leq \begin{cases} \kappa(A) & if & m = n, \\ \frac{\kappa(A)}{\cos \theta} & if & m > n. \end{cases} \tag{24}$$

Proof. If $m = n$, (19) yields

$$\eta(A, b) = \left(\frac{\sigma_1}{\sigma_n} \right) \left(\frac{\sum_{i=1}^n c_i^2}{\sum_{i=1}^n \left(\frac{\sigma_1}{\sigma_i} \right)^2 c_i^2} \right)^{\frac{1}{2}} \leq \frac{\sigma_1}{\sigma_n} = \kappa(A),$$

and the result (24) for $m = n$ follows.

Consider now the situation $m > n$,

$$\eta(A, b) = \left(\frac{\sigma_1}{\sigma_n} \right) \left(\frac{\sum_{i=1}^m c_i^2}{\sum_{i=1}^n \left(\frac{\sigma_1}{\sigma_i} \right)^2 c_i^2} \right)^{\frac{1}{2}} \leq \left(\frac{\sigma_1}{\sigma_n} \right) \left(\frac{\sum_{i=1}^m c_i^2}{\sum_{i=1}^n c_i^2} \right)^{\frac{1}{2}} = \kappa(A) \frac{\|c\|}{\|\bar{c}_1\|},$$

and the result (24) for $m > n$ follows from (16). $\quad\square$

Equation (24) shows that the condition number of x_0 increases rapidly as $\cos \theta \to 0$. For example, if a window of order 7×7 is used, then $A \in \mathbb{R}^{49 \times 2}$ and the error in x_0 is equal to zero if and only if b lies in the two dimensional subspace $\mathbb{R}^2_{colA}$ spanned by the columns of A, in the space $\mathbb{R}^{49}$, and the error in x_0 is due to the component of b that lies in the space that is orthogonal to $\mathbb{R}^2_{colA}$. The condition number of x_0 is large if $1/\cos \theta \gg 1$, even if $\sigma_1/\sigma_2 = O(1)$.

Example 1. *Consider the matrix A and vector b,*

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & -1 \end{bmatrix} \quad and \quad b = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}.$$

The singular values of A are $\sigma_1 = 1$ and $\sigma_2 = \sqrt{3}$, and thus $\kappa(A) = \sqrt{3}$. It may therefore be thought that the solution x_0 of the LS problem is well conditioned but this is incorrect because b is orthogonal to $\mathcal{R}(A)$,

$$b^T A = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}.$$

It follows that $\cos \theta = 0$ and $x_0 = \begin{bmatrix} 0 & 0 \end{bmatrix}^T$.

The left singular matrix U of $A = U\Sigma V^T$, and the matrices U_1 and U_2 are

$$U = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{3}} \\ \frac{-1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \end{bmatrix}, \quad U_1 = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\ \frac{2}{\sqrt{6}} & 0 \end{bmatrix} \quad and \quad U_2 = \begin{bmatrix} \frac{-1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix},$$

and thus

$$U_1^T b = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad and \quad U_1 U_1^T b = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

which confirm that b is orthogonal to $\mathcal{R}(A)$.

5.2. A Non-Linear Condition Number

This section extends the linear effective condition number considered in Section 5.1 by including perturbations in the coefficient matrix A, such that the perturbed solution of the LS problem is

$$x_0 + \delta x_0 = (A + \delta A)^\dagger (b + \delta b). \tag{25}$$

An expression for the condition number of x_0 when A and b are perturbed is developed in Theorem 3.

Theorem 3. *The condition number of x_0 when A and b are perturbed by, respectively, δA and δb such that*

$$\|\delta b\| \leq \epsilon \|b\| \qquad and \qquad \|\delta A\| \leq \epsilon \|A\|, \tag{26}$$

is

$$
\begin{aligned}
\rho(A, b) &= \max_{\|\delta b\| \leq \epsilon \|b\|, \|\delta A\| \leq \epsilon \|A\|} \frac{1}{\epsilon} \frac{\|\delta x_0\|}{\|x_0\|} \\
&= \frac{\|A^\dagger\| \|b\| + \|A\| \left\|\left(A^T A\right)^{-1}\right\| \left\|(I - AA^\dagger) b\right\|}{\|x_0\|} + \|A\| \left\|A^\dagger\right\| \\
&= \eta(A, b) + \kappa(A) + \frac{\kappa(A) \left\|(I - AA^\dagger) b\right\|}{\sigma_n \|x_0\|}.
\end{aligned} \tag{27}
$$

Proof. The pseudo-inverse of $A + \delta A$ is

$$(A + \delta A)^\dagger = \left((A + \delta A)^T (A + \delta A)\right)^{-1} (A + \delta A)^T,$$

where, to first order,

$$
\begin{aligned}
\left((A + \delta A)^T (A + \delta A)\right)^{-1} &= \left(I + A^\dagger \delta A + \left(A^T A\right)^{-1} \delta A^T A\right)^{-1} \left(A^T A\right)^{-1} \\
&= \left(I - A^\dagger \delta A - \left(A^T A\right)^{-1} \delta A^T A\right) \left(A^T A\right)^{-1},
\end{aligned}
$$

and thus

$$
\begin{aligned}
(A + \delta A)^\dagger &= \left((A + \delta A)^T (A + \delta A)\right)^{-1} (A + \delta A)^T \\
&= \left(I - A^\dagger \delta A - \left(A^T A\right)^{-1} \delta A^T A\right) \left(A^T A\right)^{-1} (A + \delta A)^T \\
&= A^\dagger + \left(A^T A\right)^{-1} \delta A^T - A^\dagger \delta A A^\dagger - \left(A^T A\right)^{-1} \delta A^T A A^\dagger,
\end{aligned}
$$

to first order. It therefore follows from (25) that

$$x_0 + \delta x_0 = \left(A^\dagger + \left(A^T A\right)^{-1} \delta A^T - A^\dagger \delta A A^\dagger - \left(A^T A\right)^{-1} \delta A^T A A^\dagger\right)(b + \delta b),$$

and thus

$$\delta x_0 = A^\dagger \delta b - A^\dagger \delta A x_0 + \left(A^T A\right)^{-1} \delta A^T \left(I - AA^\dagger\right) b.$$

Hence,

$$\|\delta x_0\| \;\leq\; \left\|A^\dagger\right\|\|\delta b\| + \left\|A^\dagger\right\|\|\delta A\|\|x_0\| + \left\|\left(A^T A\right)^{-1}\right\|\|\delta A\|\left\|\left(I - AA^\dagger\right)b\right\|,$$

and the substitution of (18) and (26) into this expression yields (27). □

An upper bound for the last term in (27) allows $\rho(A,b)$ to be expressed in terms of $\eta(A,b)$ and $\kappa(A)$. In particular,

$$\left\|\left(I - AA^\dagger\right)b\right\| = \left\|\begin{bmatrix} 0 & 0 \\ 0 & I_{m-n} \end{bmatrix} U^T b\right\| \leq \|b\|,$$

and thus (27) becomes

$$\begin{aligned}
\rho(A,b) &= \eta(A,b) + \kappa(A) + \frac{\kappa(A)\left\|\left(I - AA^\dagger\right)b\right\|}{\sigma_n \|x_0\|} \\
&\leq \eta(A,b) + \kappa(A) + \frac{\kappa(A)\|b\|}{\sigma_n \|A^\dagger b\|} \\
&= \eta(A,b) + \kappa(A) + \kappa(A)\eta(A,b),
\end{aligned} \tag{28}$$

and thus the inclusion of errors in the spatial derivatives of the intensity is significant because of the term $\kappa(A)\eta(A,b)$.

It was stated in Section 5.1 that the effective condition number $\eta(A,b)$ is ill conditioned if the discrete Picard condition is satisfied, in which case (27) and its upper bound (28) cannot be computed reliably. This problem is addressed by using the upper bound (24) for the value of $\eta(A,b)$ in (28). The condition number (27) is, however, useful because it shows that the inclusion of perturbations in the spatial derivatives of the intensity causes a significant increase in the condition number of the optical flow.

6. Examples

This section contains two examples that show the errors and condition numbers of the optical flow at each level of the pyramid. The Middlebury data set [1] is used in Examples 2 and 3, and the following numerical measures are considered:

1. The condition number $\kappa(A)$, the upper bound of the effective condition number, $\kappa(A)/\cos\theta$, and the non-linear condition number $\rho(A,b)$.
2. The angle θ between b and its component that lies in $\mathcal{R}(A)$.
3. The error e in the solution x_0 of the LS problem.

The examples use the images shown in Figure 2 [1].

Figure 2. The images used in Examples 2 and 3.

Example 2. *The five measures listed above (the three condition numbers, the angle θ and the error e in x_0) were computed in each window for windows of side lengths $2s + 1$, $s = 2, 3, \ldots, 8, 9$, at each level of a pyramid that has four levels, for the images in Figure 2. The mean and standard deviation of $\log_{10} \kappa(A)$, the upper bound $\log_{10} \kappa(A)/\cos\theta$ of the effective condition number, and the non-linear condition number $\log_{10} \rho(A, b)$, at each level of the pyramid and for each window size, are shown in Figure 3. The figures in the left column show that, as expected, the mean of $\rho(A, b)$ at each of the 32 points in the grid is larger than the corresponding values of $\kappa(A)$ and $\kappa(A)/\cos\theta$. Also, the means of $\rho(A, b)$ and $\kappa(A)/\cos\theta$ increase, but not monotonically, as the pyramid is descended from the coarsest image (level 4) to the finest image (level 1), and they decrease as the window size increases, at each level of the pyramid. The variation of the mean of each condition number, Figure 3 (left), is similar to the variation of its standard deviation, Figure 3 (right).*

The mean and standard deviation of the angle θ, where $\cos\theta$ is defined in (16), and the error $\log_{10} e$, where e is defined in (11), are shown in Figure 4. It is seen that the mean of θ is approximately constant for windows of side length 7 pixels or more at levels 2, 3 and 4 of the pyramid. Figure 4 (left) shows that the variation of the means of θ and $\log_{10} e$ are very similar, but Figure 4 (right) show that, for each measure, the standard deviation is significant, which implies that there is considerable variation of each measure.
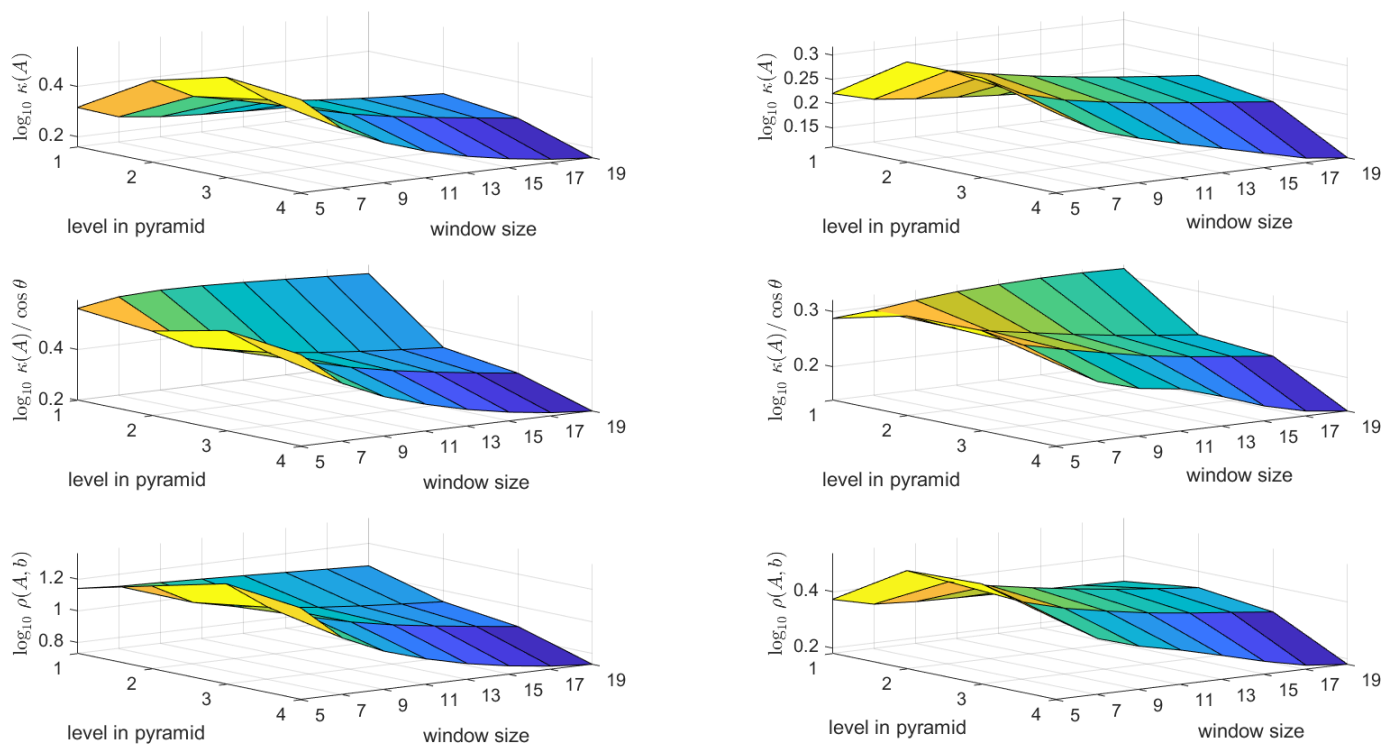

Figure 3. The variation of the mean (**left**) and standard deviation (**right**) of the condition number $\log_{10} \kappa(A)$, upper bound $\log_{10} \kappa(A)/\cos\theta$ of the effective condition number, and the non-linear condition number $\log_{10} \rho(A, b)$ with the size of the window and number of levels in the pyramid, for Example 2.
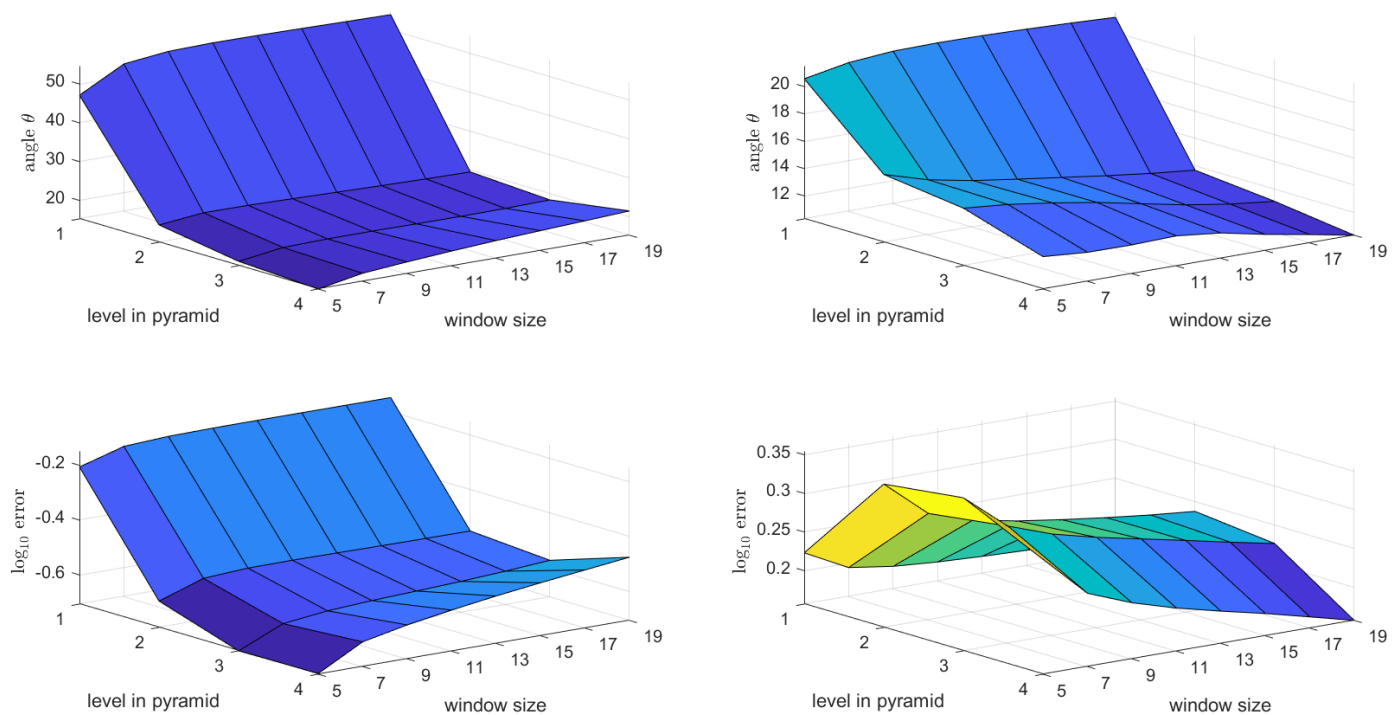
Figure 4. The variation of the mean (**left**) and standard deviation (**right**) of the angle θ and the error $\log_{10} e$, with the size of the window and number of levels in the pyramid, for Example 2.

Example 3. *The LK method with a pyramid of three levels was applied to the images in Figure 2, with a window of size 15×15. The condition number $\log_{10} \kappa(A)$, upper bound $\log_{10} \kappa(A)/\cos\theta$ of the effective condition number, the non-linear condition number $\log_{10} \rho(A,b)$, angle θ and error $\log_{10} e$ were computed for every LS problem in the windows for each level of the pyramid. Figure 5 shows the axes and size of the image at each level of the pyramid for Figures 6–10. These figures show that the quantities above increase, and the standard deviation of each of these quantities increases, as the pyramid is descended. The figures are therefore consistent with Figures 3 and 4, and it follows that the computational reliability of the optical flow decreases as the pyramid is descended.*

Figure 9 shows that the values of the local maxima of the angle θ increase significantly as the pyramid is descended. In particular there are many LS problems for which θ is about 80 degrees, from which it follows that the errors in the solutions of these problems are large. This result is consistent with the plots of the condition numbers in Figures 7 and 8. Furthermore, it follows from (15) that large values of θ are associated with large errors e, which is confirmed in Figure 10.
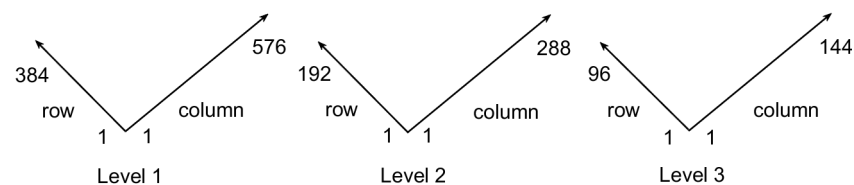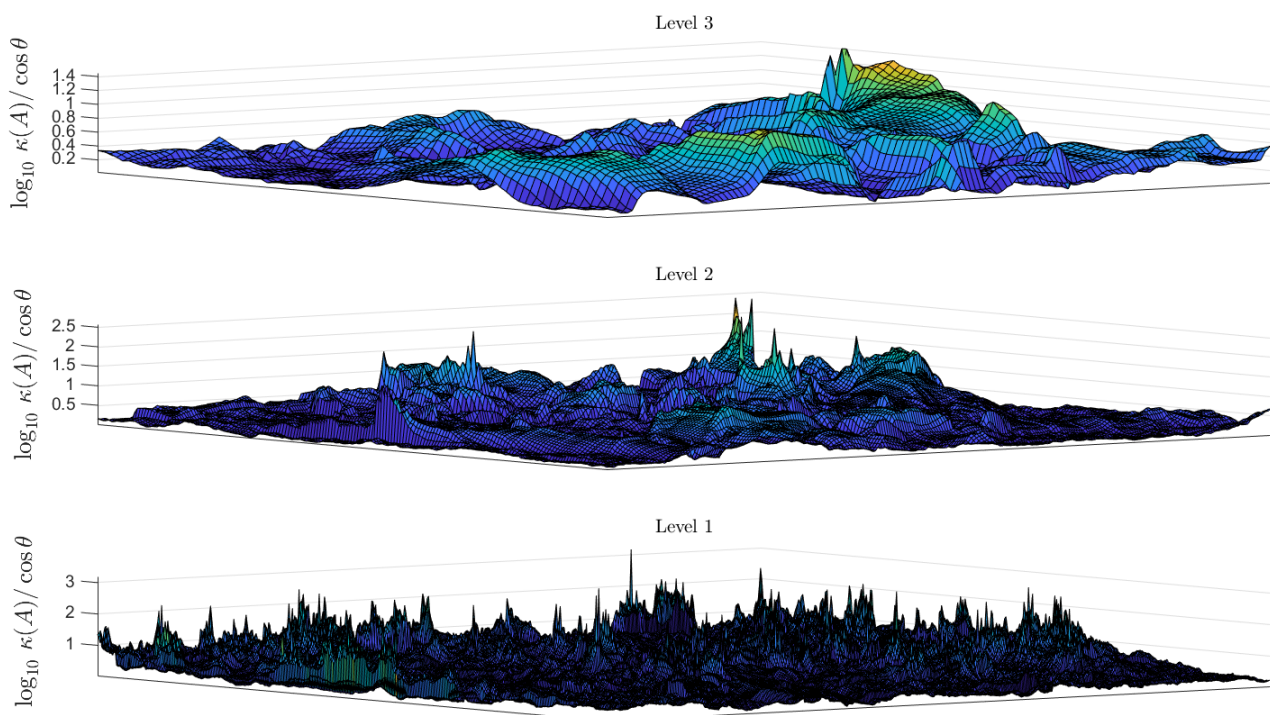

Figure 5. The sizes of the images at level 1 (384×576), level 2 (192×288) and level 3 (96×144) in Figures 6–10, for Example 3.
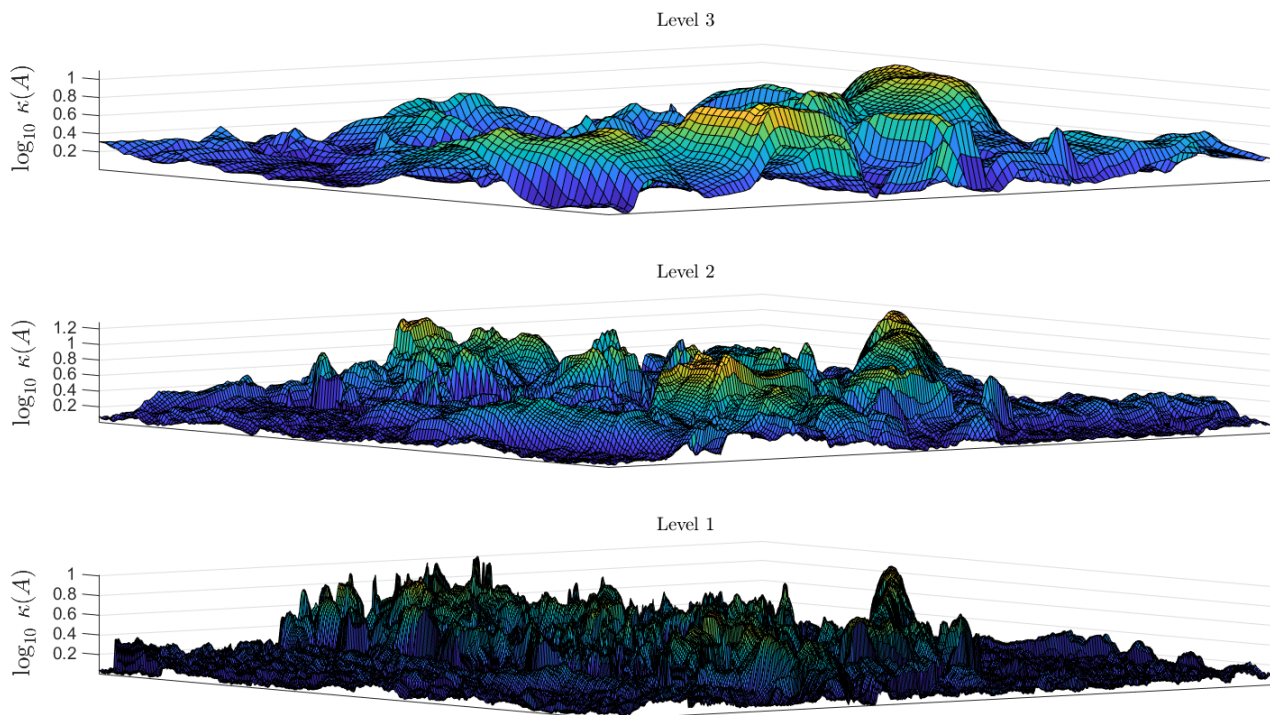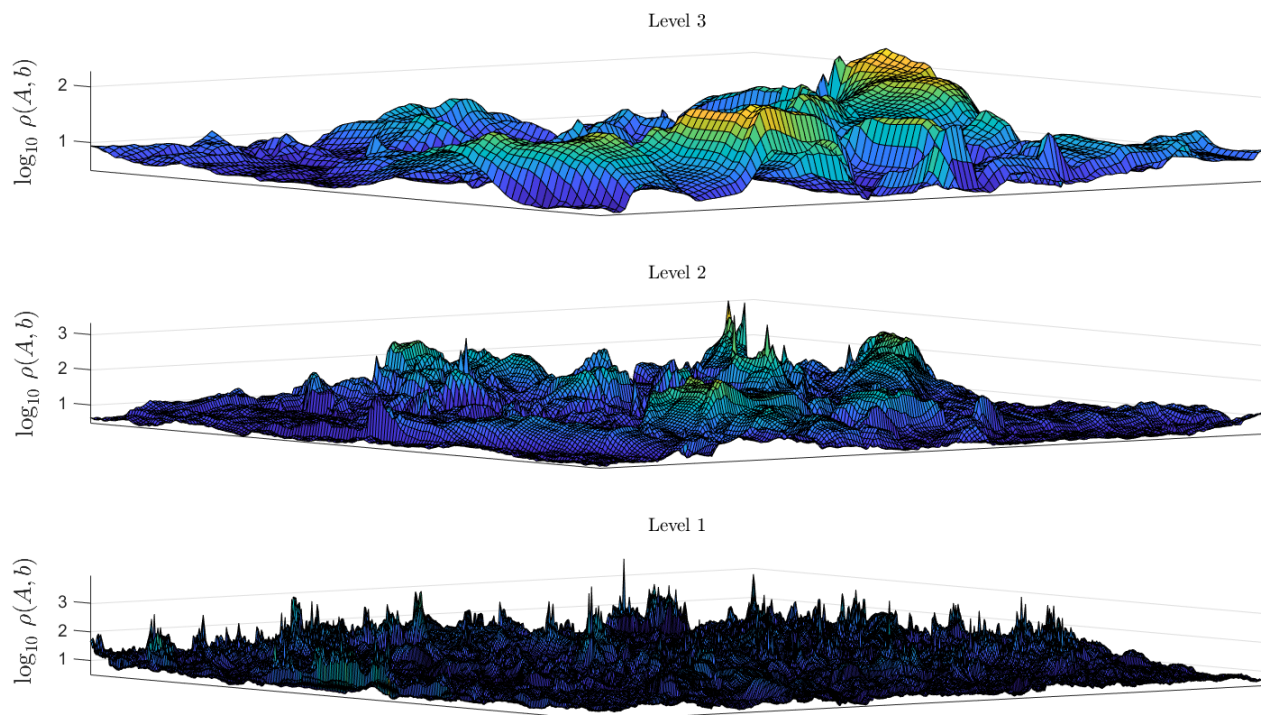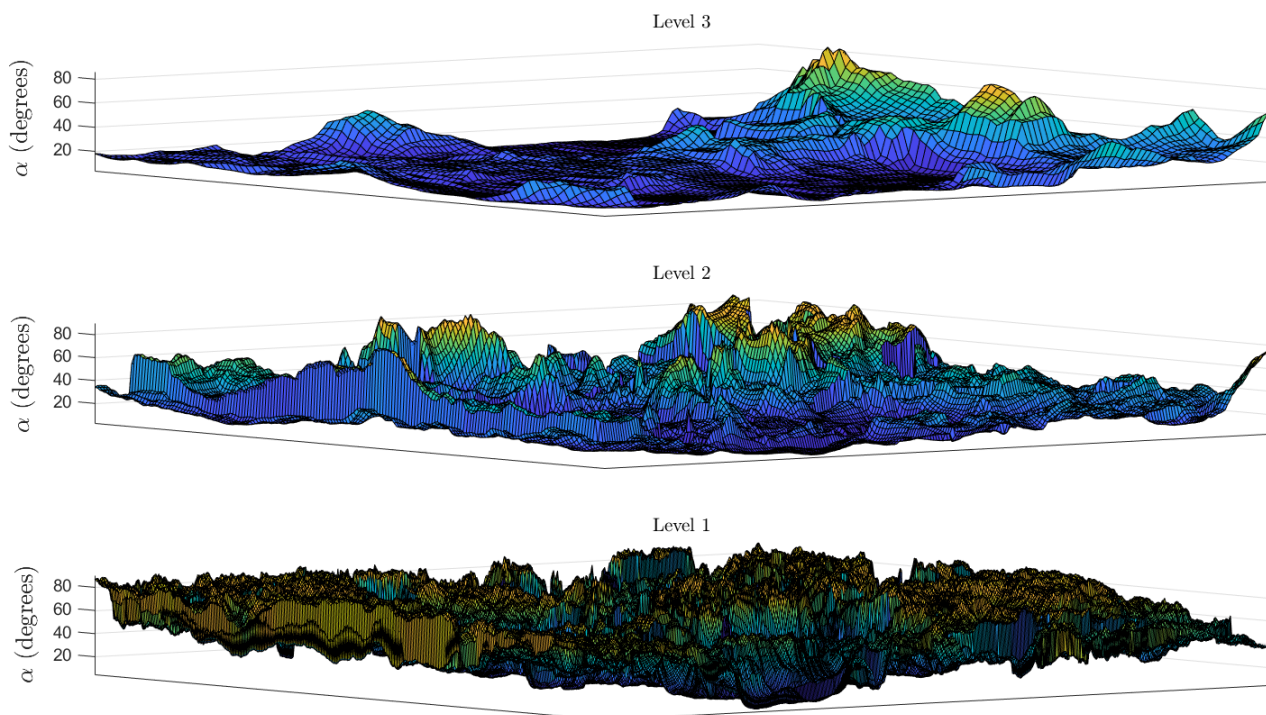
Figure 6. The condition number $\log_{10} \kappa(A)$ at level three (**top**), level two (**middle**) and level one (**base**) of the pyramid, for Example 3.
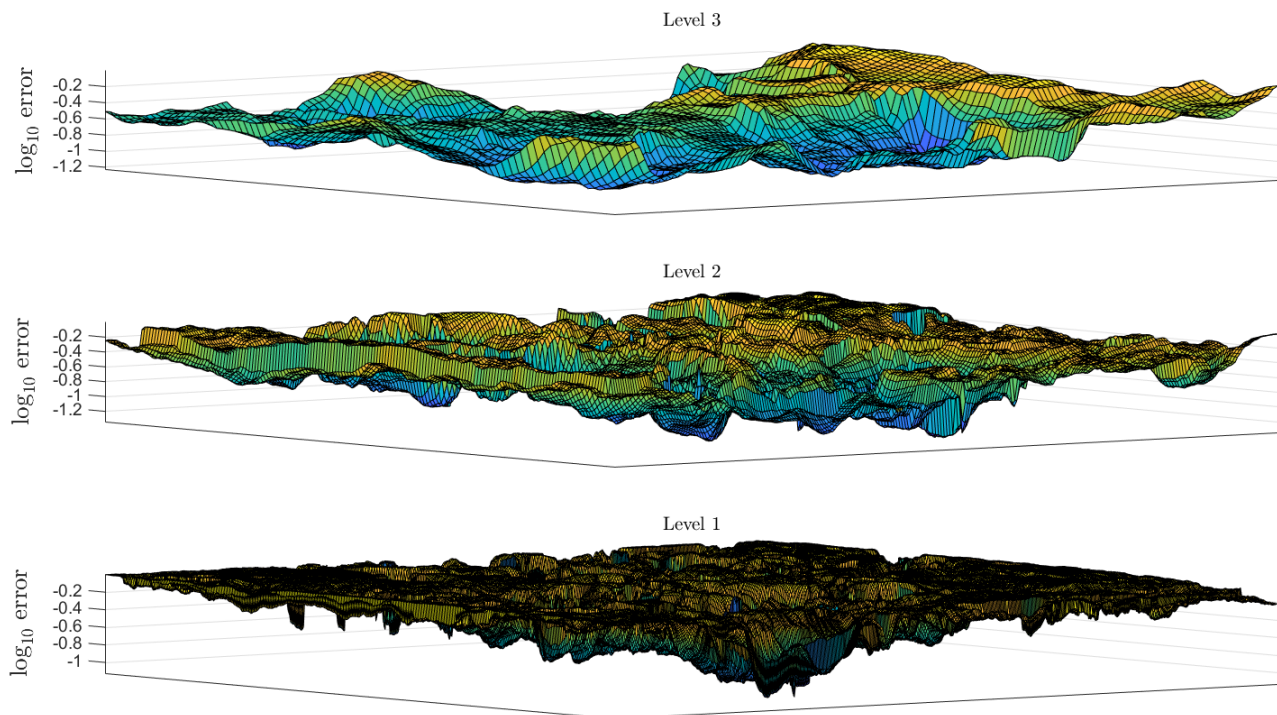
Figure 7. The upper bound $\log_{10} \kappa(A)/\cos\theta$ of the effective condition number at level three (**top**), level two (**middle**) and level one (**base**) of the pyramid, for Example 3.

Level 3

$\log_{10} \rho(A, b)$

Level 2

$\log_{10} \rho(A, b)$

Level 1

$\log_{10} \rho(A, b)$

Figure 8. The non-linear condition number $\log_{10} \rho(A, b)$ at level three (**top**), level two (**middle**) and level one (**base**) of the pyramid, for Example 3.

Level 3

α (degrees)

Level 2

α (degrees)

Level 1

α (degrees)

Figure 9. The angle θ at level three (**top**), level two (**middle**) and level one (**base**) of the pyramid, for Example 3.

Figure 10. The error $\log_{10} e$ at level three (**top**), level two (**middle**) and level one (**base**) of the pyramid, for Example 3.

7. Summary and Conclusions

The work in this paper considers the numerical properties of the solution of the LS problem because of their consequences on the computed optical flow. These consequences are compounded by the large number of LS problems that must be solved in the LK method, where the solutions of the LS problems at level l of the pyramid form the data, after upsampling, for the LS problems at level $l-1$ of the pyramid. It has been shown that the large values of the condition numbers $\eta(A, b)$ and $\rho(A, b)$, and error e, of the solution x_0 of the LS problem arise because a significant component of b lies in the space that is orthogonal to $\mathcal{R}(A)$, and thus the terms $1/\cos\theta$ and $\cos\theta$ in, respectively, the expressions for the condition numbers and error are significant. It has been shown that the error and condition numbers of the LS problem increase as the pyramid is descended from the top level (coarsest image) to the bottom level (finest image) and thus the fidelity of the computed optical flow decreases as the pyramid is descended. There have been several studies that consider the errors associated with the LK method, but the new aspect of the work described in this paper is the extensive use of computational linear algebra, specifically, condition estimation and error analysis of the solution of the LS problem.

The large values of the condition numbers and errors follow from the assumption that the optical flow in each window at each level of the pyramid is constant, that is, the motion in each window can be represented by a translation. It follows that an improved form of the LK method requires that a richer class of motions be allowed in each window, and this is achieved by the introduction of warping functions in the LK method [17]. These warping functions are defined by parametric models, for example, affine and projective models, and the values of the parameters of these models are computed in the optimisation procedure in an extended form of the LK method.

Data Availability Statement: The data used in this paper are taken from the Middlebury data set [1], which is in the public domain and frequently used in computer vision.

Conflicts of Interest: The author declares that he does have not relationships or interests that have direct or potential influence, or impart bias, on the work.

References

1. Baker, S.; Scharstein, D.; Lewis, J.; Roth, S.; Black, M.; Szeliski, R. A database and evaluation methodology for optical flow. *Int. J. Computer Vision* **2011**, *92*, 1–31. [CrossRef]
2. Horn, B.; Schunck, B. Determining optical flow. *Artif. Intell.* **1981**, *17*, 185–203. [CrossRef]
3. Lucas, B.D.; Kanade, T. An iterative image registration technique with an application to stereo vision. In Proceedings of the 7th International Joint Conference on Artificial Intelligence, Vancouver, BC, Canada, 24–28 August 1981; pp. 674–679.
4. Fortun, D.; Bouthemy, P.; Kervrann, C. Optical flow modeling and computation: A survey. *Comput. Vis. Image Underst.* **2015**, *134*, 1–21. [CrossRef]
5. Hofinger, M.; Bulò, S.; Porzi, L.; Knapitsch, A.; Pock, T.; Kontschieder, P. Improving optical flow on a pyramidal level. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; pp. 770–786.
6. Shah, S.T.H.; Xuezhi, X. Traditional and modern strategies for optical flow: An investigation. *SN Appl. Sci.* **2021**, *3*. [CrossRef]
7. Tu, Z.; Xie, W.; Zhang, D.; Poppe, R.; Veltkamp, R.; Li, B.; Yuan, J. A survey of variational and CNN-based optical flow techniques. *Signal Process. Image Commun.* **2019**, *72*, 9–24. [CrossRef]
8. Zhai, M.; Xiang, X.; Lv, N.; Kong, X. Optical flow and scene flow estimation: A survey. *Pattern Recognit.* **2021**, *114*, 107861. [CrossRef]
9. Savian, S.; Elahi, M.; Tillo, T. Optical flow estimation with deep learning: A survey on recent advances. In *Deep Biometrics*; Jiang, R., Li, C.T., Crookes, D., Meng, W., Rosenberger, C., Eds.; Springer: Berlin/Heidelberg, Germany, 2020; pp. 257–287.
10. Lin, T.; Barron, J. Image reconstruction error for optical flow. *Res. Comput. Robot. Vis.* **1995**, 269–290.
11. Kearney, J.; Thompson, W.; Boley, D. Optical flow estimation: An error analysis of gradient-based methods with local optimization. *IEEE Trans. Pattern Anal. Mach. Intell.* **1987**, *9*, 229–244. [CrossRef]
12. Golub, G.H.; Loan, C.F.V. *Matrix Computations*; John Hopkins University Press: Baltimore, MD, USA, 2013.
13. Tsai, C.; Galatsanos, N.; Katsaggelos, A. Total least squares estimation of stereo optical flow. In Proceedings of the IEEE International Conference on Image Processing, Chicago, IL, USA, 4–7 October 1998; pp. 622–626.
14. Kim, Y.; Kak, A. Error analysis of robust optical flow estimation by least median of squares methods for the varying illumination model. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 1418–1435. [CrossRef]
15. Tarasenko, V.; Park, D. Detection and tracking over image pyramids using Lucas and Kanade algorithm. *Int. J. Appl. Eng. Res.* **2016**, *11*, 6117–6120.
16. Klette, R. *Concise Computer Vision: An Introduction into Theory and Applications*; Springer: London, UK, 2014.
17. Baker, S.; Matthews, I. Lucas-Kanade 20 years on: A unifying framework. *Int. J. Computer Vision* **2004**, *56*, 221–255. [CrossRef]
18. Brandt, J.W. Analysis of bias in gradient-based optical flow estimation. In Proceedings of the 28th Asilomar Conference of Signals, Systems and Computers, Pacific Grove, CA, USA, 31 October–2 November 1994; pp. 721–725.
19. Pham, T.Q.; Bezuijen, M.; van Vliet, L.J.; Schutte, K.; Hendriks, C.L.L. Performance of optimal registration estimators. *Proc. SPIE* **2005**, *5817*, 133–144.
20. Robinson, D.; Milanfar, P. Fundamental performance limits in image registration. *IEEE Trans. Image Process.* **2004**, *13*, 1185–1199. [CrossRef] [PubMed]
21. Burt, P. Fast filter transforms for image processing. *Comput. Graph. Image Process.* **1981**, *16*, 20–51. [CrossRef]
22. Burt, P.; Adelson, E. The Laplacian pyramid as a compact image code. *IEEE Trans. Comm.* **1983**, *31*, 532–540. [CrossRef]
23. Winkler, J.R.; Mitrouli, M. Condition estimation for regression and feature estimation. *J. Comput. Appl. Math.* **2020**, *373*, 112212. [CrossRef]
24. Winkler, J.R.; Mitrouli, M.; Koukouvinos, C. The application of regularisation to variable selection in statistical modelling. *J. Comput. Appl. Math.* **2022**, *404*, 113884. [CrossRef]
25. Hansen, P.C. *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*; SIAM: Philadelphia, PA, USA, 1998.
26. Hansen, P.C.; Nagy, J.G.; O'Leary, D.P. *Deblurring Images: Matrices, Spectra, and Filtering*; SIAM: Philadelphia, PA, USA, 2006.