

Analysis of Quadratic Surface Fitting for Subpixel Motion Extraction from Video Images

Bian Xiong

Univ Gustave Eiffel, Inria, France

bian.xiong@inria.fr

Qinghua Zhang

Univ Gustave Eiffel, Inria, France

qinghua.zhang@inria.fr

Vincent Baltazart

Univ Gustave Eiffel, IFSTTAR, France

vincent.baltazart@univ-eiffel.fr

Abstract—Digital image correlation is a popular method for estimating object displacement in successive images. At the pixel level, displacement is estimated by maximizing the cross-correlation between two images. To achieve subpixel accuracy, displacement estimation can be refined in the vicinity of the cross-correlation peak. Among existing refinement methods, quadratic surface fitting provides a good trade-off between accuracy and computational burden. The purpose of this paper is to analyze the quadratic surface fitting method. It is shown that the quadratic surface fitted to the cross-correlation values in the vicinity of the cross-correlation peak does not always have a maximum. Then the conditions ensuring the existence of a maximum are analyzed. The reported results consolidate the theoretic basis of the quadratic surface fitting method for subpixel motion extraction.

Index Terms—Digital image correlation, quadratic surface fitting, subpixel refinement.

I. INTRODUCTION

Image correlation is widely involved in various techniques for motion extraction from image sequences [1], [2], including digital image correlation techniques (DIC), phase-only correlation (POC) and virtual image correlation (VIC) techniques [3]. Correlation-based techniques are known to provide accurate measurement with high computational efficiency, along with good robustness against noise. It has been used for a huge variety of applications in remote sensing, computer vision, and more generally for image processing purposes.

In this paper, an image correlation technique is investigated for video-based Structural Health Monitoring (SHM), in order to survey civil engineering structures at long range [4], [5]. Within this scope of application, it is important to extract subpixel motion information (displacement and/or velocity) from video images.

The authors of [1], [2] provide performance comparisons of different correlation-based techniques to this aim. At the pixel level, displacement is estimated by maximizing the cross-correlation between two images. To achieve subpixel accuracy, displacement estimation is refined in the vicinity of the cross-correlation peak. Among such refinement methods, *quadratic surface fitting* provides a good trade-off between accuracy and computational burden, particularly suitable for video-based SHM [6]. The good performance of this method has also been confirmed by our own experiments.

The purpose of the present paper is to analyze the quadratic surface fitting method. It will be shown that, contrary to a widespread intuition, the quadratic surface fitted to the cross-correlation values within the 3×3 pixels vicinity of the correlation peak *does not* always have a maximum. Then the conditions ensuring the *existence* of a maximum of the fitted quadratic surface will be analyzed. The reported results consolidate the theoretic basis of the quadratic surface fitting method for subpixel motion extraction.

II. PROBLEM STATEMENT

In this paper, it is assumed that a moving object is observed through a video image time flow with small displacement amplitude between successive images. Correlation processing will focus on a rectangular template of $M \times N$ pixels, also known as a region of interest (ROI), which includes either the whole moving object or some identified edges of the object. The brightness of each pixel at instant t is denoted by $I(n, m, t)$ where the *integer* pair (n, m) indicates the position of the pixel in an image.

In contrast to the usual convention in image processing, in this paper the integer pair (n, m) denotes pixel coordinates in the Cartesian system on the plane of a rectangular image, with the origin at the lower-left corner of the image, and the horizontal and vertical axes coinciding respectively with the bottom and left borders of the image. This notation definition is more suitable for the surface fitting problem formulated in the Cartesian coordinate system, in agreement with usual mathematical notations.

Motion extraction will be carried out by accurately determining the horizontal and vertical shifts of the selected template between two images captured at instants t and $t + \Delta t$, assuming negligible rotational motion. The horizontal image shift is decomposed into an integer part \tilde{n} and a fractional (subpixel) part x , so that the total horizontal shift is equal to $\tilde{n} + x$ with $|x| < 1$. Similarly the vertical shift is decomposed into $\tilde{m} + y$ with $|y| < 1$.

Given two images (frames) captured at time instants t and $t + \Delta t$, as illustrated in Figure 1, template shifts are usually estimated through a two-step process [6]. The pixel level (integer) shifts (\tilde{n}, \tilde{m}) are first estimated by maximizing the cross-correlation between the two image templates. At the second

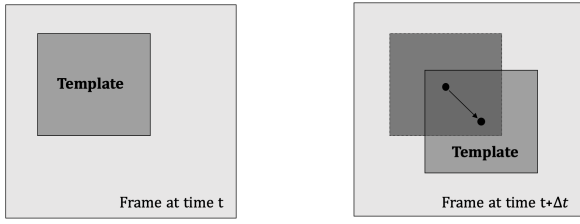


Fig. 1. Template shifts between two successive images

step, the cross-correlation is interpolated in the vicinity of the cross-correlation peak to estimate the subpixel shifts (x, y) .

The search for the cross-correlation peak at the pixel level is formulated as follows:

$$(\tilde{n}^*, \tilde{m}^*) = \arg \max_{\substack{-M \leq \tilde{n} \leq M \\ -N \leq \tilde{m} \leq N}} \rho(\tilde{n}, \tilde{m}), \quad (1)$$

with the cross-correlation defined as

$$\rho(\tilde{n}, \tilde{m}) \triangleq \sum_{(n,m) \in \mathbb{T}} I(n, m, t) I(n - \tilde{n}, m - \tilde{m}, t + \Delta t), \quad (2)$$

where \mathbb{T} denotes the integer pairs (n, m) corresponding to pixels belonging to the considered template. The dependences on $I(n, m, t)$ and on $I(n, m, t + \Delta t)$ are omitted in the notation $\rho(\tilde{n}, \tilde{m})$ for lighter presentations.

It is well known that maximizing the cross-correlation between two data patterns is equivalent to minimizing the sum of squared linear matching errors between them. This fact is stated in [6] as the equivalence of (1) to

$$(\tilde{n}^*, \tilde{m}^*) = \arg \min_{\substack{-M \leq \tilde{n} \leq M \\ -N \leq \tilde{m} \leq N}} \epsilon(\tilde{n}, \tilde{m}). \quad (3)$$

with

$$\epsilon(\tilde{n}, \tilde{m}) \triangleq \min_{\alpha \in \mathbb{R}} \sum_{(n,m) \in \mathbb{T}} [I(n, m, t) - \alpha I(n - \tilde{n}, m - \tilde{m}, t + \Delta t)]^2, \quad (4)$$

where again the dependences on $I(n, m, t)$ and on $I(n, m, t + \Delta t)$ are omitted in the notation $\epsilon(\tilde{n}, \tilde{m})$. This equivalence supports the practice of maximizing the cross-correlation, as formulated in (1).

At the second step, in order to gain subpixel accuracy, the cross-correlation $\rho(\tilde{n}, \tilde{m})$ is somehow interpolated for non integer shifts so that the correlation maximization (1) can be generalized to subpixel shifts.

As in [6], this interpolation is considered in the 3×3 pixels vicinity of the cross-correlation peak found at the integer shifts $(\tilde{n}^*, \tilde{m}^*)$, by fitting, in the least squares sense, a second degree polynomial (or, geometrically, a *quadratic surface*) to the value of $\rho(\tilde{n}^*, \tilde{m}^*)$ and to the 8 neighboring cross-correlation values $\rho(\tilde{n}, \tilde{m})$, namely $\rho(\tilde{n}^* + n, \tilde{m}^* + m)$ with $n, m \in \{-1, 0, +1\}$. Then the maximum of the fitted

polynomial yields the estimated subpixel shifts between the two templates.

More formally, the integer shifts $(\tilde{n}^*, \tilde{m}^*)$ being already estimated, let $p(x, y)$ denote the second degree polynomial fitted, in the least squares sense, to $\rho(\tilde{n}^* + x, \tilde{m}^* + y)$ for $x, y \in \{-1, 0, +1\}$. Then the subpixel shifts are estimated as

$$(x^*, y^*) = \arg \max_{(x,y) \in \mathbb{R}^2} p(x, y), \quad (5)$$

and the estimated total shifts amount to

$$(\tilde{n}^* + x^*, \tilde{m}^* + y^*). \quad (6)$$

Satisfactory experimental results of this method have been reported by different authors, e.g., [1], [2], [6]. Our own simulation studies confirm also its good performance compared to other existing methods for subpixel shift estimation. The purpose of this paper is to consolidate the theoretical basis of this method.

More precisely, the quadratic surface fitting method, as described in the aforementioned references, assumes *implicitly* that the second degree polynomial $p(x, y)$ fitted to the 9 correlation values $\rho(\tilde{n}^* + n, \tilde{m}^* + m)$, with $n, m \in \{-1, 0, +1\}$, *always* has a unique global maximum, corresponding to (x^*, y^*) located in the close vicinity of $(x, y) = (0, 0)$, so that the total shifts as expressed in (6) do not fall too far from the pixel level optimal shifts $(\tilde{n}^*, \tilde{m}^*)$.

This paper will investigate the following issues.

- 1) Does the second degree polynomial $p(x, y)$ fitted to the 9 cross-correlation values *always* have a unique global maximum?
- 2) If this fitted polynomial *does not always* have a unique global maximum, what are the conditions ensuring its existence?

III. ANALYSIS OF THE POLYNOMIAL FITTED TO CROSS-CORRELATION VALUES

Let $(\tilde{n}^*, \tilde{m}^*)$ be resulting from the pixel level maximization (1).

The 9 integer pairs (n, m) , with $n, m \in \{-1, 0, +1\}$, form a 3×3 grid:

$$\mathbb{G} = \begin{bmatrix} (-1, 1) & (0, 1) & (1, 1) \\ (-1, 0) & (0, 0) & (1, 0) \\ (-1, -1) & (0, -1) & (1, -1) \end{bmatrix}. \quad (7)$$

Accordingly, the 9 cross-correlation values $\rho(\tilde{n}^* + n, \tilde{m}^* + m)$, denoted also, *for shorter notations*, by

$$\gamma(n, m) \triangleq \rho(\tilde{n}^* + n, \tilde{m}^* + m), \quad (8)$$

form a matrix

$$\Gamma = \begin{bmatrix} \gamma(-1, 1) & \gamma(0, 1) & \gamma(1, 1) \\ \gamma(-1, 0) & \gamma(0, 0) & \gamma(1, 0) \\ \gamma(-1, -1) & \gamma(0, -1) & \gamma(1, -1) \end{bmatrix}. \quad (9)$$

The second degree polynomial

$$p_\theta(x, y) = \theta_1 + [\theta_2 \ \theta_3] \begin{bmatrix} x \\ y \end{bmatrix} + [x \ y] \begin{bmatrix} \theta_4 & \theta_5/2 \\ \theta_5/2 & \theta_6 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}, \quad (10)$$

with the vector $\theta \in \mathbb{R}^6$ collecting the scalar coefficients $\theta_1, \dots, \theta_6$, is then fitted to the entry values of Γ for $(x, y) \in \mathbb{G}$, by solving the least squares problem

$$\min_{\theta \in \mathbb{R}^6} \sum_{(n, m) \in \mathbb{G}} [p_\theta(n, m) - \gamma(n, m)]^2. \quad (11)$$

Does this fitted second degree polynomial $p_\theta(x, y)$ *always* have a unique global maximum?

Because the integer shifts $(\tilde{n}^*, \tilde{m}^*)$ are the results of the pixel level optimization (1), the *central entry* of the matrix Γ , namely $\gamma(0, 0) = \rho(\tilde{n}^*, \tilde{m}^*)$ as defined in (8), is the *maximum* value among all the 9 entries of Γ . It then *seems reasonable* to expect that the second degree polynomial $p_\theta(x, y)$ fitted to the 9 entries of Γ has a maximum somehow close to the (maximum) central entry of the matrix Γ corresponding to the origin $(x, y) = (0, 0)$.

Unfortunately, the fact that the central entry $\gamma(0, 0)$ is the maximum value among the 9 entries of Γ *does not* really ensure that the fitted second degree polynomial $p_\theta(x, y)$ *always* has a global maximum, as demonstrated by the following counterexample.

Consider the cross-correlations values filling up the Γ matrix (normalized by the central entry):

$$\Gamma = \begin{bmatrix} 0.7486 & 0.1558 & 0.1253 \\ 0.1558 & 1.0000 & 0.1558 \\ 0.1253 & 0.1558 & 0.7486 \end{bmatrix}.$$

Fitting the second degree polynomial $p_\theta(x, y)$ to Γ for $(x, y) \in \mathbb{G}$ yields

$$\theta = \begin{bmatrix} 0.4998 \\ 0.0000 \\ 0.0000 \\ -0.0940 \\ -0.3117 \\ -0.0940 \end{bmatrix}.$$

The corresponding quadratic surface exhibiting a *saddle point*, as illustrated in Figure 2, has no global maximum, despite the fact that the central entry of the matrix Γ is its largest entry.

This counterexample clearly invalidates the intuition that the fitted second degree polynomial $p_\theta(x, y)$ *always* has a global maximum, thus answering the first question raised at the end of section II.

The following section will answer the remaining question: given that the fitted second degree polynomial *does not always* have a unique global maximum, what are the conditions ensuring the existence of such a maximum?

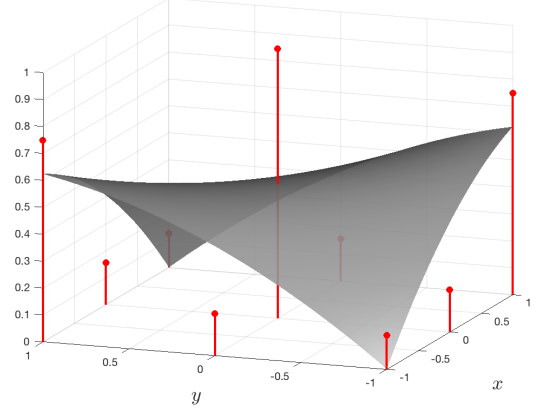


Fig. 2. Example of fitted quadratic surface exhibiting a saddle point (no global maximum). The vertical red line segments represent the entries of Γ .

IV. CONDITIONS FOR THE EXISTENCE A MAXIMUM

In elementary algebra [7], it is well known that the second degree polynomial $p_\theta(x, y)$ as expressed in (10) has a unique global maximum if its Hessian matrix

$$H \triangleq \begin{bmatrix} \theta_4 & \theta_5/2 \\ \theta_5/2 & \theta_6 \end{bmatrix} \quad (12)$$

is negative definite. However, this simply stated fact does not directly help to understand how the cross-correlation values $\gamma(n, m)$ (those filling up the Γ matrix in (9)) should be, so that $p_\theta(x, y)$ has a global maximum. Because the polynomial coefficients $\theta_1, \dots, \theta_6$ are determined from the values of $\gamma(n, m)$ by solving the least squares problem (11), it is straightforward to express the negative definiteness condition of H in terms of $\gamma(n, m)$. Then, *in principle*, the condition for the existence of a global maximum of $p_\theta(x, y)$ will be directly formulated in terms of the cross-correlations values $\gamma(n, m)$. However, this approach results in a sophisticated condition, notably an inequality involving the determinant of H expressed in terms of $\gamma(n, m)$. For a *better understanding*, the result presented below will be formulated with *simple and easily interpretable* inequalities about the cross-correlation values $\gamma(n, m)$ filling up Γ . For instance, one of these simple inequalities states that the central entry $\gamma(0, 0)$ of Γ is its largest entry. As shown by the previously presented counterexample, this condition alone is not sufficient. It is then completed by similar simple inequalities.

Theorem 1: If the cross-correlation values $\gamma(n, m)$ filling up the matrix Γ satisfy

$$\gamma(0, 0) \geq \gamma(n, m) \quad \text{for all } n, m \in \{-1, 0, +1\} \quad (13)$$

$$\gamma(0, m) > \gamma(n, m) \quad \text{for all } n, m \in \{-1, +1\} \quad (14)$$

$$\gamma(n, 0) > \gamma(n, m) \quad \text{for all } n, m \in \{-1, +1\}, \quad (15)$$

then the second degree polynomial $p_\theta(x, y)$ fitted to the entries of Γ on the grid \mathbb{G} defined in (7), by solving the least squares problem (11), has a unique global maximum.

Interpretation of the conditions of Theorem 1.

- Inequalities (13): the central entry $\gamma(0, 0)$ has the largest value among all the 9 entries of Γ .
- Inequalities (14): the middle entry $\gamma(0, \pm 1)$ is the largest entry of the top or the bottom row of Γ .
- Inequalities (15): the middle entry $\gamma(\pm 1, 0)$ is the largest entry of the right or the left column of Γ .

Proof of Theorem 1.

In order to shorten lengthy equations and inequalities, let us introduce more compact notations for the cross-correlation values $\gamma(m, m)$ filling up the matrix Γ defined in (9), so that Γ is rewritten as

$$\Gamma = \begin{bmatrix} a & e & b \\ f & i & g \\ c & h & d \end{bmatrix}. \quad (16)$$

Remark that the letters a, b, d, \dots, i fill Γ first at the 4 corners, then the middles of side rows and columns, before finishing at the central entry.

With these compact notations, the least squares solution (11) leads to

$$9\theta_1 = 2(e + f + g + h) + 5i - (a + b + c + d) \quad (17a)$$

$$6\theta_2 = (b - a) + (g - f) + (d - c) \quad (17b)$$

$$6\theta_3 = (a - c) + (e - h) + (b - d) \quad (17c)$$

$$-6\theta_4 = (e - a) + (e - b) + (h - c) + (h - d) + (i - f) + (i - g) \quad (17d)$$

$$4\theta_5 = (b - a + c - d) \quad (17e)$$

$$-6\theta_6 = (f - a) + (f - c) + (g - b) + (g - d) + (i - e) + (i - h). \quad (17f)$$

As already mentioned in this paper, the negative definiteness of the Hessian matrix H defined in (12) ensures that the polynomial $p_\theta(x, y)$ has a global maximum. Based on Sylvester's criterion¹, this negative definiteness will be checked through:

$$\theta_4 < 0 \quad (18)$$

$$\det(H) > 0. \quad (19)$$

According to the inequalities assumed in (14), e (or h , resp.) is the largest entry of the top (or bottom, resp.) row of Γ , then

$$e - a > 0, \quad e - b > 0 \quad (20)$$

$$h - c > 0, \quad h - d > 0 \quad (21)$$

and according to (13), i is the largest entry of Γ , then

$$i - f \geq 0, \quad i - g \geq 0. \quad (22)$$

These inequalities together with (17d) imply immediately (18).

¹Usually Sylvester's criterion [8] is about the *positive* definiteness of a real symmetric (or complex Hermitian) matrix. It is trivial to translate this criterion to the case of *negative* definiteness.

It is more involved to check (19). The inequalities in (20) imply

$$\left(1 + \frac{3}{4}\right)(e - a) > 0 > \left(-1 + \frac{3}{4}\right)(e - b), \quad (23)$$

then

$$(e - a) + (e - b) > -\frac{3}{4}(e - a) + \frac{3}{4}(e - b) \quad (24)$$

$$= -\frac{3}{4}(b - a). \quad (25)$$

Repeat the reasoning from (23) to (25) while interchanging the positions of $(e - a)$ and $(e - b)$:

$$\left(1 + \frac{3}{4}\right)(e - b) > 0 > \left(-1 + \frac{3}{4}\right)(e - a), \quad (26)$$

leading to

$$(e - a) + (e - b) > \frac{3}{4}(b - a). \quad (27)$$

Combining (25) and (27) yields

$$(e - a) + (e - b) > \frac{3}{4}|b - a|. \quad (28)$$

This result expresses a relationship between the entries in the top row of the matrix Γ . A similar reasoning then leads to the following relationship between the entries in the bottom row of Γ :

$$(h - c) + (h - d) > \frac{3}{4}|c - d|. \quad (29)$$

According to the inequalities assumed in (13), i is the largest entry of Γ , then

$$i - f \geq 0, \quad i - g \geq 0. \quad (30)$$

Take the sums of the respective sides of the 4 inequalities in (28), (29) and (30), then

$$(e - a) + (e - b) + (h - c) + (h - d) + (i - f) + (i - g) > \frac{3}{4}|b - a| + \frac{3}{4}|c - d| \quad (31)$$

$$\geq \frac{3}{4}|b - a + c - d|. \quad (32)$$

This result then implies that $-6\theta_4$ and $4\theta_5$, as expressed respectively in (17d) and (17e), satisfy

$$-6\theta_4 > \frac{3}{4}|4\theta_5|, \quad (33)$$

hence

$$-\theta_4 > \frac{1}{2}|\theta_5| \geq 0. \quad (34)$$

Following the same approach, it is then similarly shown that

$$-\theta_6 > \frac{1}{2}|\theta_5| \geq 0. \quad (35)$$

This last inequality can also be deduced from a certain "symmetry" between the formulae expressing θ_4 and θ_6 in (17d) and (17f).

It then follows from (12), (34) and (35) that

$$\det(H) = \theta_4\theta_6 - \frac{1}{4}\theta_5^2 > 0. \quad (36)$$

Therefore, the two inequalities (18) and (19) ensuring the negative definiteness of H are successfully checked. It is then established that the second degree polynomial $p_\theta(x, y)$ has a unique global maximum.

When the fitted polynomial $p_\theta(x, y)$ has a unique global maximum, it may happen that this maximum is far away from the origin $(x, y) = (0, 0)$ corresponding to the optimized integer shifts $(\tilde{n}^*, \tilde{m}^*)$, outside the square area corresponding to the 3×3 pixels vicinity of the cross-correlation peak. The following result ensures that the maximum of $p_\theta(x, y)$ stays inside this square area, under easily interpretable conditions.

Theorem 2: If, in addition to the conditions of Theorem 1, the cross-correlation values $\gamma(n, m)$ filling up the matrix Γ satisfy, for all $n, m \in \{-1, +1\}$,

$$\gamma(0, m) - \gamma(n, m) > \frac{1}{5} [\gamma(0, m) - \gamma(-n, m)] \quad (37)$$

$$\gamma(n, 0) - \gamma(n, m) > \frac{1}{5} [\gamma(n, 0) - \gamma(n, -m)], \quad (38)$$

then the maximum of the fitted polynomial $p_\theta(x, y)$ is located at (x^*, y^*) such that $|x^*| < 1$ and $|y^*| < 1$.

Interpretation of the conditions of Theorem 2.

The conditions inherited from Theorem 1 ensure that the middle entry in each row or column of Γ is the largest entry of the row or column, without imposing any “degree of symmetry”. For example, among the top row of Γ as expressed in (9), inequalities formulated in (14) ensure that $\gamma(0, 1)$ is the largest entry, but the ratio $[\gamma(0, 1) - \gamma(-1, 1)] / [\gamma(0, 1) - \gamma(1, 1)]$ can be any positive number. Two of the inequalities in the extra condition (37) of Theorem 2 constrain this ratio between $1/5$ and 5 , thus limiting the dissymmetry between $\gamma(-1, 1)$ and $\gamma(1, 1)$.

Due to the page limitation of this conference paper, the proof of Theorem 2, omitted here, will be presented elsewhere.

V. CONCLUSION

In this paper, within the scope of developing digital image correlation techniques for SHM, some theoretical aspects of the quadratic surface fitting method have been investigated. Compared to the literature, this paper has brought some theoretic insight for a better understanding of this method, by providing mathematical conditions ensuring expected results. As the existence of a maximum of the fitted quadratic surface can be numerically checked through the Hessian matrix, the results of this paper are mainly for theoretic purpose. Further works will include a complementary algorithm addressing the cases unexpected by the quadratic surface fitting method, and the performance assessment of the completed method through applications within the considered scope.

REFERENCES

- [1] X. Tong, Z. Ye, Y. Xu, S. Gao, H. Xie, Q. Du, S. Liu, X. Xu, S. Liu, K. Luan, and U. Stilla, “Image registration with fourier-based image correlation: A comprehensive review of developments and applications,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. PP, 08 2019.
- [2] A. Alba, J. F. Viguera-Gomez, E. R. Arce-Santana, and R. M. Aguilar-Ponce, “Phase correlation with sub-pixel accuracy: A comparative study in 1d and 2d,” *Computer Vision and Image Understanding*, vol. 137, pp. 76 – 87, 2015. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1077314215000685>
- [3] B. Semin, H. Auradou, and M. François, “Accurate measurement of curvilinear shapes by virtual image correlation,” *The European Physical Journal - Applied Physics*, vol. 56, no. 1, p. 10701, 2011.
- [4] D. Feng and M. Q. Feng, “Computer vision for shm of civil infrastructure: From dynamic response measurement to damage detection – a review,” *Engineering Structures*, vol. 156, pp. 105 – 117, 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0141029617315559>
- [5] B. F. Spencer, V. Hoskere, and Y. Narazaki, “Advances in computer vision-based civil infrastructure inspection and monitoring,” *Engineering*, 2019.
- [6] D. Zhang, J. Guo, X. Lei, and C. Zhu, “A high-speed vision-based sensor for dynamic vibration analysis using fast motion extraction algorithms,” *Sensors*, vol. 16, no. 4, p. 572, 2016. [Online]. Available: <https://doi.org/10.3390/s16040572>
- [7] O. T. O’Meara, *Introduction to Quadratic Forms*. Berlin Heidelberg: Springer-Verlag, 1973.
- [8] C. D. Meyer, *Matrix analysis and applied linear algebra*. Philadelphia: SIAM, 2010.