

# Comparaison de méthodes d'estimation de déplacement sous-pixel basées sur la phase

Cédric MARINEL<sup>1,2</sup>, Benjamin MATHON<sup>1</sup>, Olivier LOSSON<sup>1</sup>, Ludovic MACAIRE<sup>1</sup>

<sup>1</sup>Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRISTAL, F-59000 Lille, France

<sup>2</sup>EOMYS Engineering, Lille, France

cedric.marinel.etu@univ-lille.fr

**Résumé** – L'analyse des propriétés mécaniques de structures de génie civil contribue à la détection précoce de défaillances. L'analyse modale opérationnelle permet d'estimer ces propriétés mécaniques à partir de mesures de vibrations. Profitant des travaux sur l'estimation de déplacement, plusieurs méthodes d'analyse modale opérationnelle à partir de vidéos ont été développées ces dernières années. Ce travail compare deux stratégies d'estimation de déplacements inférieurs au pixel basées sur la phase de la décomposition multi-échelle. Des vidéos synthétiques de poutres verticales vibrantes sont générées pour étudier la robustesse de ces méthodes à l'amplitude de déplacement, à la quantification des niveaux de gris, ainsi qu'au bruit additif.

**Abstract** – Monitoring mechanical properties of structures such as bridges, buildings, or wind turbines is important to detect early stage failures. Operational modal analysis is a testing method to estimate these properties from vibration measurements. Taking advantage of works on motion estimation, several modal analysis methods based on videos have emerged in the last decade. This paper compares two strategies about sub-pixel phase-based motion estimations thanks to multi-scale decomposition. Synthetic videos of a vibrating cantilever beam are generated to assess the robustness of these methods against motion amplitude, gray scale quantization, and white noise.

## 1 Introduction

La surveillance des propriétés mécaniques des structures de génie civil permet la détection précoce de défaillances. L'analyse modale opérationnelle (AMO) a été développée pour estimer les propriétés mécaniques, telles que la forme du mode et les fréquences naturelles, à partir de mesures d'accélération, de vitesse et de déplacement. Ces mesures sont généralement acquises par des capteurs de contact tels que les accéléromètres ou les transducteurs différentiels variables linéaires. Toutefois, la mise en place de ces capteurs sur la structure peut être longue et coûteuse. C'est pourquoi des méthodes peu coûteuses ont récemment été développées pour effectuer ces mesures sans contact grâce à l'analyse de vidéos [1]. Comme les vibrations peuvent être considérées comme de petits déplacements périodiques d'un élément de surface observé par la caméra, les déplacements doivent être estimés avec une précision inférieure au pixel. Chen *et al.* [2] exploitent la phase de la réponse complexe de filtres en quadrature appliqués aux images d'une vidéo pour estimer les déplacements et analyser les vibrations de structures simples. Ils supposent que le déplacement est représenté selon une seule échelle spatiale. Cependant, les fréquences spatiales d'intérêt peuvent varier selon les zones de l'image.

Deux approches existent pour estimer le déplacement à l'aide de la décomposition pyramidale multi-échelle. Wadhwa *et al.* [3] prennent en compte chaque échelle de la pyramide pour obtenir une seule estimation de déplacement par pixel qui est ex-

ploitée par une AMO. En revanche, Yang *et al.* [4] estiment le déplacement selon chaque échelle et effectuent une AMO par échelle. Comme, à notre connaissance, aucune étude ne compare ces méthodes d'estimation de déplacement basées sur la phase, nous proposons d'évaluer leurs performances à l'aide de vidéos synthétiques d'une poutre verticale. Ces vidéos sont générées avec différentes amplitudes de déplacement afin d'étudier la précision des méthodes. Nous nous focalisons sur leur robustesse à la quantification des niveaux de gris, ainsi qu'au bruit additif qui dégrade la qualité des vidéos.

## 2 Estimation du déplacement basée sur la phase

### 2.1 Décomposition pyramidale

Soit  $I(x, y; t)$  l'intensité au pixel  $(x, y)$  de l'image  $t$ . On souhaite estimer le déplacement en chaque pixel de l'image  $t$  :

$$\delta(x, y; t) = \begin{pmatrix} \delta^h(x, y; t) \\ \delta^v(x, y; t) \end{pmatrix} \in \mathbb{R}^2. \quad (1)$$

En supposant que l'éclairage est spatialement et spectralement constant au cours du temps, l'intensité en un pixel associé à un élément de surface donné peut être considérée constante :

$$I(x, y; 0) = I(x + \delta^h(x, y; t), y + \delta^v(x, y; t); t). \quad (2)$$

Les méthodes suivantes se basent sur la décomposition en pyramide complexe orientée (PCO) pour séparer chaque image

en sous-bandes spatio-fréquentielles à l'aide de filtres complexes. Les fréquences spatiales  $(\omega^h, \omega^v)$  de chaque image sont décomposées en coordonnées polaires  $(\omega_r \cos(\theta), \omega_r \sin(\theta))$  correspondant aux différentes échelles  $r = 1, \dots, \mathcal{N}_r$  et orientations  $\theta = 0, \dots, \frac{\mathcal{N}_\theta - 1}{\mathcal{N}_\theta} \pi$ , où  $\mathcal{N}_r$  et  $\mathcal{N}_\theta$  sont respectivement les nombres d'échelles et d'orientations.

L'amplitude  $\rho_{r,\theta}$  et la phase  $\varphi_{r,\theta}$  associées à la réponse complexe  $S_{r,\theta}(x, y; t) = G_{r,\theta} * I(x, y; t)$  du filtre  $G_{r,\theta}$  appliquée à l'image  $t$  s'expriment par :

$$\rho_{r,\theta}(x, y; t) = |S_{r,\theta}(x, y; t)|, \quad (3)$$

$$\varphi_{r,\theta}(x, y; t) = \arctan \left( \frac{\Im(S_{r,\theta}(x, y; t))}{\Re(S_{r,\theta}(x, y; t))} \right). \quad (4)$$

## 2.2 Estimation du déplacement

La réponse du filtre appliqué à l'image 0 peut être exprimée en fonction du déplacement dans l'image  $t$  :

$$S_{r,\theta}(x, y; 0) = G_{r,\theta} * I(x, y; 0) \quad (5)$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I(\tilde{x}, \tilde{y}; 0) G_{r,\theta}(x - \tilde{x}, y - \tilde{y}) d\tilde{x} d\tilde{y} \quad (6)$$

$$\stackrel{(2)}{=} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I(\tilde{x} + \delta^h(\tilde{x}, \tilde{y}; t), y + \delta^v(\tilde{x}, \tilde{y}; t); t) G_{r,\theta}(x - \tilde{x}, y - \tilde{y}) d\tilde{x} d\tilde{y}. \quad (7)$$

En supposant que  $\delta^h$  et  $\delta^v$  sont constants sur le support (de taille  $b_r^h \times b_r^v$ ) du filtre  $G_{r,\theta}$  à chaque image  $t$ , nous obtenons :

$$S_{r,\theta}(x, y; 0) = \int_{x-b_r^h}^{x+b_r^h} \int_{y-b_r^v}^{y+b_r^v} I(\tilde{x} + \delta^h(t), \tilde{y} + \delta^v(t); t) G_{r,\theta}(x - \tilde{x}, y - \tilde{y}) d\tilde{x} d\tilde{y}. \quad (8)$$

Par les changements de variables  $\hat{x} = \tilde{x} + \delta^h(t)$  et  $\hat{y} = \tilde{y} + \delta^v(t)$ , la réponse du filtre à l'image 0 peut être déduite de celle de l'image  $t$  :

$$S_{r,\theta}(x, y; 0) = S_{r,\theta}(x + \delta^h(t), y + \delta^v(t); t). \quad (9)$$

Pour chaque échelle, l'amplitude et la phase de la réponse du filtre s'expriment ainsi par :

$$\rho_{r,\theta}(x, y; 0) = \rho_{r,\theta}(x + \delta^h(t), y + \delta^v(t); t), \quad (10)$$

$$\varphi_{r,\theta}(x, y; 0) = \varphi_{r,\theta}(x + \delta^h(t), y + \delta^v(t); t). \quad (11)$$

En considérant que  $\delta^h(t)$  et  $\delta^v(t)$  dépendent de la localisation du pixel, on peut déduire que l'amplitude et la phase de la réponse du filtre sont approximativement constantes :

$$\rho_{r,\theta}(x, y; 0) \approx \rho_{r,\theta}(x + \delta^h(x, y; t), y + \delta^v(x, y; t); t), \quad (12)$$

$$\varphi_{r,\theta}(x, y; 0) \approx \varphi_{r,\theta}(x + \delta^h(x, y; t), y + \delta^v(x, y; t); t). \quad (13)$$

On suppose que  $\varphi_{r,\theta} \in \mathcal{C}^1$  pour tout  $r, \theta$ , et  $t$  pour effectuer le développement de Taylor à l'ordre 1 de (13) :

$$\varphi_{r,\theta}(x, y; 0) - \varphi_{r,\theta}(x, y; t) \approx \nabla \varphi_{r,\theta}(x, y; t) \cdot \delta(x, y; t). \quad (14)$$

Comme le gradient de la phase  $\nabla \varphi_{r,\theta}$  est approximativement égal aux fréquences spatiales centrales du filtre [5], le déplacement peut être estimé en remplaçant  $\nabla \varphi_{r,\theta}$  par  $(\omega^h, \omega^v)$  dans (14).

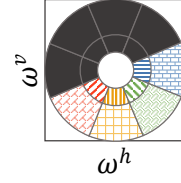


FIGURE 1 – Supports des filtres fréquentiels idéaux de Simoncelli et Freeman  $G_{r,\theta}$  avec  $\mathcal{N}_r = 2$  échelles et  $\mathcal{N}_\theta = 4$  orientations.

Nous utilisons le peigne de Dirac pour échantillonner les données spatio-temporelles continues  $I$ ,  $\rho_{r,\theta}$  et  $\varphi_{r,\theta}$ . On note ainsi leurs versions discrètes :  $I[x, y; t]$ ,  $\rho_{r,\theta}[x, y; t]$ ,  $\varphi_{r,\theta}[x, y; t]$  et  $\delta[x, y; t]$ , avec  $[x, y; t] \in \mathbb{J}1, \mathcal{N}_x \mathbb{K} \times \mathbb{J}1, \mathcal{N}_y \mathbb{K} \times \mathbb{J}0, \mathcal{N}_t - 1 \mathbb{K}$ , où  $\mathcal{N}_x$ ,  $\mathcal{N}_y$  et  $\mathcal{N}_t$  sont respectivement le nombre de colonnes de pixels, de lignes de pixels, et d'images.

## 2.3 Estimation multiéchelle de déplacement

Pour décomposer chaque image en une PCO, Wadhwa *et al.* [3] appliquent les filtres fréquentiels de Simoncelli et Freeman [6] dont les supports sont représentés sur la Fig. 1. Ils résolvent un problème des moindres carrés pondérés pour estimer le déplacement  $\hat{\delta}$  en fusionnant les phases de chaque sous-bande :

$$\hat{\delta}[x, y; t] = \underset{\delta[x, y; t]}{\operatorname{argmin}} \sum_{r=1}^2 \sum_{\theta=0}^{\frac{3\pi}{4}} \sum_{k=-9}^9 \sum_{\ell=-9}^9 \mathcal{G}[k, \ell] \rho_{r,\theta}^2[x + k, y + \ell; t] \left[ (\omega^h, \omega^v) \cdot \delta[x, y; t] - (\varphi_{r,\theta}[x + k, y + \ell; 0] - \varphi_{r,\theta}[x + k, y + \ell; t]) \right]^2. \quad (15)$$

Les poids sont basés sur le carré  $\rho_{r,\theta}^2$  de l'amplitude de la réponse du filtre utilisé pour la décomposition en sous-bandes. En effet, la phase dans chaque sous-bande ne doit être prise en compte que si l'amplitude associée est élevée. Les auteurs supposent aussi que le déplacement est localement constant et ajoutent une contrainte spatiale. Par conséquent, ils pondèrent cette réponse avec un noyau Gaussien  $\mathcal{G}$  d'écart type 3 px et de support  $19 \text{ px} \times 19 \text{ px}$ . De plus, avant de résoudre (15), la phase  $\varphi_{r,\theta}[x, y; t]$ , calculée dans l'intervalle  $(-\pi, \pi]$ , est temporellement déroulée pour comparer le déphasage entre les images  $t$  et 0. En pratique,  $\varphi_{2,\theta}$  et  $\rho_{2,\theta}$  sont interpolées de manière biquadratique pour atteindre les mêmes définitions spatiales que  $\varphi_{1,\theta}$  et  $\rho_{1,\theta}$ .

## 2.4 Estimation monoéchelle de déplacement

Yang *et al.* [4] utilisent aussi les filtres fréquentiels de [6] (cf. Fig. 1) pour la décomposition de chaque image en PCO. De plus, ils supposent que le déplacement vertical dans leur vidéo de poutre verticale en porte-à-faux peut être négligé, *i.e.*  $\delta^v(x, y; t) \approx 0$ , ce qui donne pour (12) :

$$\varphi_{r,\theta}(x, y; 0) \approx \varphi_{r,\theta}(x + \delta^h(x, y; t), y; t). \quad (16)$$

Ainsi, en utilisant le développement de Taylor et l'approximation des dérivées partielles de la phase [5], (14) devient :

$$\varphi_{r,\theta}(x, y; 0) = \varphi_{r,\theta}(x, y; t) + \omega^h \delta^h(x, y; t). \quad (17)$$

Les auteurs utilisent uniquement les réponses des filtres horizontaux ( $\theta = 0$ ) et estiment les déplacements horizontaux aux échelles  $r \in \{1, 2\}$  par :

$$\hat{\delta}_r^h[x, y; t] = \frac{\varphi_{r,0}[x, y; 0] - \varphi_{r,0}[x, y; t]}{\omega_r}. \quad (18)$$

La phase est aussi déroulée temporellement avant l'estimation.  $\hat{\delta}_1^h$  est estimé sur l'image en pleine définition alors que  $\hat{\delta}_2^h$  est d'abord calculé avec la phase  $\varphi_{2,0}$  sous-échantillonnée, puis une interpolation bicubique permet d'obtenir une estimation en pleine définition.

## 3 Résultats expérimentaux

### 3.1 Dispositif expérimental

Pour comparer ces méthodes, nous générons des vidéos synthétiques d'une poutre verticale encastree à sa base à l'aide de l'équation d'Euler-Bernoulli. Ce modèle nécessite de préciser les paramètres physiques de la poutre : la longueur  $L$  (m), l'aire de la section ( $\text{m}^2$ ), le module de Young  $E$  (Pa), le moment d'inertie  $J$  ( $\text{m}^4$ ) et la masse par unité de longueur  $\mu$  ( $\text{kg}\cdot\text{m}^{-1}$ ). La ligne centrale de la poutre verticale est définie dans le système de coordonnées de la scène par l'ensemble  $\{(g(Y, t), Y; t) \in \mathbb{R} \times [0, L] \times ]0, \mathcal{N}_t - 1\}$ , où  $g$  est solution de l'équation d'Euler-Bernoulli :

$$EJ \frac{\partial^4 g(Y, t)}{\partial Y^4} + \mu \frac{\partial^2 g(Y, t)}{\partial t^2} = f(Y, t). \quad (19)$$

Dans cette expérience, la force d'entrée  $f$  (N) est représentée par un Dirac en temps et en espace pour simuler un coup de marteau à l'extrémité libre de la poutre. Pour simuler le comportement de notre poutre expérimentale, nous fixons son volume à  $900 \times 30 \times 6 \text{ mm}^3$ , sa masse à 1,413 kg et son module de Young  $E$  à  $210 \cdot 10^9$  Pa. La taille de l'image est de  $720 \times 40$  px et un pixel observe un élément de surface de 1,289 mm de côté. La cadence d'acquisition des images est de 436 i/s pour correspondre aux caractéristiques de notre caméra expérimentale. Ainsi, les vibrations de fréquences inférieures à 218 Hz peuvent être étudiées. Les vidéos générées durent 2,3 s et contiennent  $\mathcal{N}_t = 1000$  images. Les niveaux de gris des pixels dans les images sont normalisés entre 30 et 225 et quantifiés sur  $\mathcal{N}_b = 8$  bits.

Nos analyses se concentrent sur deux pixels représentés sur la Fig. 2 et situés sur le bord gauche de la poutre, l'un au milieu et l'autre au sommet. Le tableau 1 montre les amplitudes des déplacements horizontaux  $\delta^h$  du modèle calculés au milieu et à l'extrémité libre de la poutre pour une force  $f$  donnée. L'amplitude de  $\delta^h$  à l'extrémité est approximativement trois fois plus importante que celle au milieu de la poutre.

Pour estimer le déplacement  $\delta^h$ , chaque vidéo est analysée par l'approche multiéchelle à l'aide de (15) pour obtenir  $\hat{\delta}^h$  et par l'approche monoéchelle en utilisant (18) pour calculer  $\hat{\delta}_1^h$  et  $\hat{\delta}_2^h$ . Comme la largeur de la poutre couvre 5 px, le noyau Gaussien  $\mathcal{G}$  utilisé par l'approche multiéchelle est adapté avec

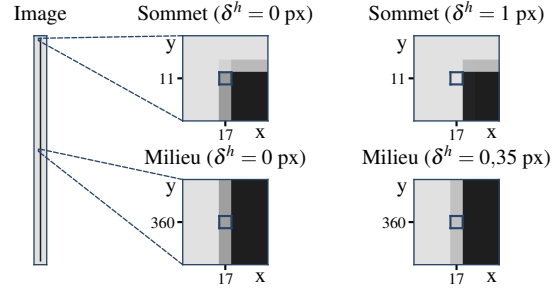


FIGURE 2 – Pixels d'intérêt pour l'estimation du mouvement dans deux images de poutre (à l'équilibre et en déplacement).

Force $f$ (N)	0,04	0,08	0,16	0,33	0,65	1,31	2,62	5,24	10,47
$\delta^h$ au sommet	0,016	0,03	0,06	0,12	0,25	0,5	1	2	4
$\delta^h$ au milieu	0,005	0,01	0,02	0,04	0,09	0,18	0,35	0,7	1,4

TABLE 1 – Amplitude du déplacement horizontal théorique  $\delta^h$  (px) aux deux pixels de la Fig. 2 en fonction de la force  $f$ .

un écart-type de 1 px et un support de  $7 \text{ px} \times 7 \text{ px}$ . Les déplacements horizontaux théoriques et estimés pour les 400 premières images sont représentés sur la Fig. 3 pour une poutre excitée par une force  $f = 10,47$  N. Comme les déplacements sont estimés par rapport à la première image, ils ne sont pas centrés autour de 0 px. En pratique, la moyenne temporelle des déplacements estimés devrait être soustraite pour obtenir des vibrations centrées autour de la position de repos. Cette figure montre que pour les deux pixels,  $\hat{\delta}^h$  et  $\hat{\delta}_1^h$  sous-estiment les déplacements alors que  $\hat{\delta}_2^h$  les surestime.

### 3.2 Robustesse à l'amplitude de déplacement

Une analyse de sensibilité à l'amplitude de déplacement est nécessaire pour mettre en évidence la précision de l'estimation au niveau sous-pixel. Neuf vidéos sont donc générées avec une force  $f$  allant de 0,04 N à 10,47 N selon une échelle logarithmique (cf. Table 1). Nous comparons ensuite les coefficients de corrélation de Pearson entre le déplacement théorique  $\delta^h$  et ceux estimés ( $\hat{\delta}^h$ ,  $\hat{\delta}_1^h$ , et  $\hat{\delta}_2^h$ ) au sommet et au milieu du bord de la poutre (cf. Fig. 4(a)).

Pour  $f$  allant de 0,08 N à 10,47 N, les estimateurs  $\hat{\delta}^h$  et  $\hat{\delta}_2^h$  atteignent les mêmes corrélations élevées. Les coefficients de  $\hat{\delta}_1^h$  sont légèrement moins élevés. Pour les déplacements de petites amplitudes ( $f = 0,04$  N), aucune méthode n'atteint un résultat satisfaisant pour le pixel du milieu.

### 3.3 Robustesse à la quantification des niveaux de gris

Comme l'éclairage en conditions extérieures ne peut pas être contrôlé, le contraste au sein de l'image peut varier. Pour tester la robustesse des méthodes à ce paramètre, nous avons quantifié les vidéos avec un nombre de bits  $\mathcal{N}_b$  variant de 2 à 8 avant d'estimer les déplacements. Cette étude a été faite sur deux vidéos, l'une avec de grandes amplitudes de déplacement ( $f = 2,62$  N) et l'autre avec de faibles amplitudes ( $f = 0,08$  N). Les trois méthodes estiment avec succès les déplacements de grandes amplitudes pour  $\mathcal{N}_b \geq 4$ . Pour les déplacements de

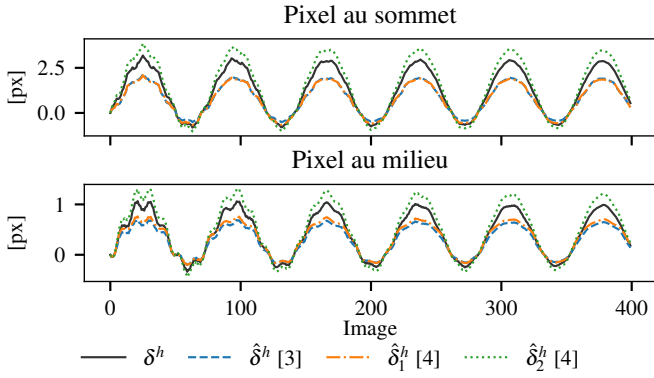


FIGURE 3 – Déplacements théoriques et estimés pour  $f = 10,47$  N.

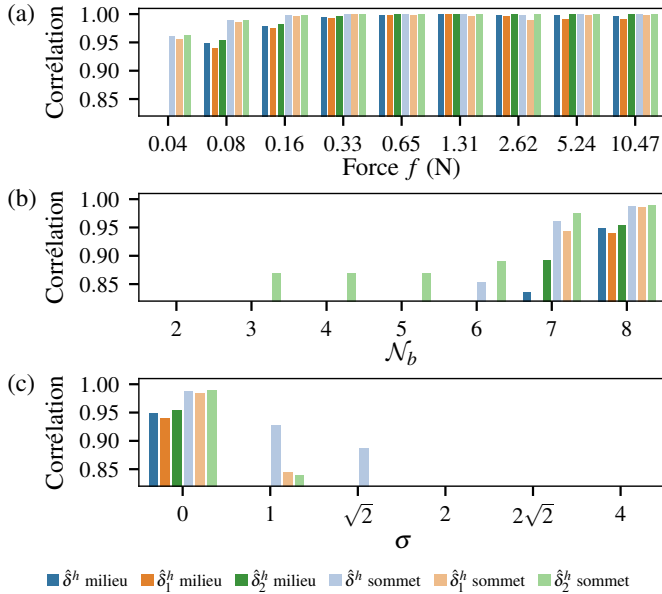


FIGURE 4 – Corrélation entre les déplacements théoriques et estimés en fonction de la force (a), du nombre de bits de quantification pour  $f = 0,08$  N (b) et de l'écart-type du bruit pour  $f = 0,08$  N (c).

faibles amplitudes ( $f = 0,08$  N),  $\mathcal{N}_b = 8$  bits de quantification sont nécessaires pour avoir une estimation correcte en chaque pixel. Pour le pixel au sommet, la robustesse à la quantification de  $\delta_2^h$  est bonne (cf. Fig. 4(b)).

### 3.4 Robustesse au bruit

Pour cette étude, un bruit Gaussien a été ajouté dans chaque image des deux vidéos précédemment utilisées (avec  $\mathcal{N}_b = 8$ ) avant d'estimer les déplacements. Différents écarts-types de bruit ont été appliqués pour tester la robustesse des méthodes avec différentes amplitudes de déplacements. Pour les grandes amplitudes ( $f = 2,62$  N), la corrélation des trois estimateurs est proche de 1,0 en chaque pixel. Pour les petites amplitudes ( $f = 0,08$  N),  $\delta^h$  est plus robuste au bruit que  $\delta_1^h$  et  $\delta_2^h$  (cf. Fig. 4(c)). Cela montre que la pondération par un filtre Gaussien pour la résolution de (15) réduit les perturbations liées au bruit.

## 4 Conclusion

Dans cette étude, nous comparons les performances de deux méthodes d'estimation de déplacement basée sur la phase. Pour ce faire, des vidéos synthétiques d'une poutre sont générées à l'aide d'un modèle physique qui simule le comportement d'une poutre vibrante. Deux pixels d'intérêt sont retenus au milieu et au sommet de la poutre pour estimer les déplacements. Les déplacements horizontaux sont estimés soit par une approche multiéchelle en résolvant un problème des moindres carrés pondérés pour fusionner les phases à différentes échelles, soit par une approche rapide monoéchelle. La seconde méthode fournit les moins bons résultats sur la première échelle. Sur la seconde échelle, elle donne globalement les meilleures estimations et semble plus robuste à la quantification des niveaux de gris. L'approche multiéchelle donne de bonnes estimations et se révèle être la plus robuste au bruit. En revanche, elle est la plus coûteuse en temps de calculs. Cette étude montre que la sélection de l'échelle adéquate reste un problème ouvert pour l'estimation du déplacement. De plus, cette étude se concentre sur l'estimation d'un déplacement horizontal. Par conséquent, les performances devraient être comparées a priori sur la direction de déplacement.

## Références

- [1] J.-Y. CHOU et C.-M. CHANG : Image motion extraction of structures using computer vision techniques: A comparative study. *Sensors*, 21(18), septembre 2021.
- [2] J. G. CHEN, N. WADHWA, Y.-J. CHA, F. DURAND, W. T. FREEMAN et O. BUYUKOZTURK : Modal identification of simple structures with high-speed video using motion magnification. *Journal of Sound and Vibration*, 345:58–71, juin 2015.
- [3] N. WADHWA, J. G. CHEN, J. B. SELLON, D. WEI, M. RUBINSTEIN, R. GHAFFARI, D. M. FREEMAN, O. BÜYÜKÖZTÜRK, P. WANG, S. SUN, S. H. KANG, K. BERTOLDI, F. DURAND et W. T. FREEMAN : Motion microscopy for visualizing and quantifying small motions. *Proceedings of the National Academy of Sciences*, 114(44):11639–11644, octobre 2017.
- [4] Y. YANG, C. DORN, T. MANCINI, Z. TALKEN, G. KENYON, C. FARRAR et D. MASCAREÑAS : Blind identification of full-field vibration modes from video measurements with phase-based video motion magnification. *Mechanical Systems and Signal Processing*, 85:567–590, février 2017.
- [5] D. J. FLEET et A. D. JEPSON : Computation of component image velocity from local phase information. *International Journal of Computer Vision*, 5(1):77–104, août 1990.
- [6] E. P. SIMONCELLI et W. T. FREEMAN : The steerable pyramid: A flexible architecture for multi-scale derivative computation. In *Procs. International Conference on Image Processing*, Washington, DC, USA, octobre 1995.