



HAL
open science

Contribution à des architectures de stabilisation d'images basées sur la perception visuelle et la physiologie du tremblement humain

Fabien Gavant

► **To cite this version:**

Fabien Gavant. Contribution à des architectures de stabilisation d'images basées sur la perception visuelle et la physiologie du tremblement humain. Autre. Université de Grenoble, 2012. Français. NNT : 2012GRENT061 . tel-01124089

HAL Id: tel-01124089

<https://theses.hal.science/tel-01124089>

Submitted on 6 Mar 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Nanoélectronique et Nanotechnologies**

Arrêté ministériel : 7 août 2006

Présentée par

Fabien GAVANT

Thèse dirigée par **Dominique DAVID** et
codirigée par **Laurent ALACOQUE**

préparée au sein du **Laboratoire CEA-LETI/DACLE/L3I**
dans l'**École Doctorale EEATS**

Contribution à des architectures de stabilisation d'images basées sur la perception visuelle et la physiologie du tremblement humain

Thèse soutenue publiquement le **11 décembre 2012**,
devant le jury composé de :

M. Christian JUTTEN

Professeur EEATS, Université Joseph Fourier, Grenoble, Président

M. Michel PAINDAVOINE

Professeur, Université de Bourgogne, Dijon, Rapporteur

M. Jacques-Olivier KLEIN

Professeur, Université Paris-Sud, Orsay, Rapporteur

M. Virginio CANTONI

Professeur, Università di Pavia, Italia, Membre

M. Dominique DAVID

HDR, CEA-LETI/DSIS/SPICE, Grenoble, Membre

M. Laurent ALACOQUE

Docteur, CEA-LETI/DACLE/L3I, Grenoble, Membre



À mon épouse

Remerciements

Cette thèse n'aurait pas été possible sans le bienveillant soutien de certaines personnes je ne suis pas non plus capable de dire avec les mots qui conviennent, le rôle qu'elles ont pu jouer à mes côtés pour en arriver là. Cependant, je voudrais les remercier du fond du cœur.

Mes premiers remerciements vont d'abord à mon directeur de thèse, **Dominique David**, qui m'a accompagné tout au long de ma formation. Sa disponibilité et ses généreux secours au cours de certains moments difficiles ont été d'une très grande qualité, et d'un immense réconfort; merci infiniment ! Mes seconds remerciements vont à **Laurent Alacoque**, mon encadrant direct. Merci de m'avoir fait grandir dans la rigueur et l'argumentation scientifique. Puis mes remerciements vont à **Antoine Dupret**, notre encadrant des doctorants du laboratoire. Merci pour ta précieuse relecture de ce manuscrit et les suggestions d'amélioration. Merci à notre chef de laboratoire, **Michael Tchagaspagnian** pour ta confiance renouvelée depuis le début de cette thèse.

Je remercie chaleureusement mes rapporteurs **Jacques-Olivier Klein** et **Michel Paindavoine** pour leurs relectures de ce manuscrit. Je remercie aussi les autres membres de mon jury **Christian Jutten**, **Virginio Cantoni** qui m'ont permis de passer la soutenance de thèse.

Mes remerciements vont aussi à tous mes collègues du laboratoire, **Tien, Fadoua, Anaïs, Bertrand, Laurent, Jean-Pierre, Jean-Luc, Josep, Rémi, Fabrice, Arnaud, Manuel, Assim, Gilles, Timothé, Leonardo, Sébastien, Franz, Romain, Hawraa, William** qui, d'une manière ou d'une autre ont joué un rôle dans l'accomplissement de cette thèse.

Je remercie également toutes les personnes que j'ai eu l'occasion de côtoyer pendant ces années de thèse, particulièrement celles qui ont contribué aux différents tests visuels dont **Margaux, Alejandro, Santhosh, Nathaniel, Olivier, Hung, Antoine, Kawtar, Laure, Romain, Thomas**.

Toute cette belle organisation ne serait rien sans le dévouement de l'administration. J'ai été heureux de mener ces travaux au sein du département DACLE sous la direction de **Jean-René Lequepeys** puis de **Thierry Collette**, ainsi que la supervision scientifique de **Marc Belleville** avec la fameuse journée annuelle des doctorants. Je remercie tout particulièrement **Armelle** et **Caroline**, nos adjointes de direction qui sont de vraies fées.

Merci aussi à mes relecteurs, **Françoise, Pierre, Louis-Pierre, Alain** d'avoir chassé un bon nombre de fautes pour ce présent manuscrit.

Je voudrais aussi remercier mes **parents** qui je pense s'attendaient à beaucoup de chose sauf que je rédige une thèse ! Mais aussi mes frères et sœurs **Nath, Flo, AnneC, Fréd** et **MC** aussi que leurs « **pièces rapportées** ». Merci **Papi** et **Mami** de m'avoir transmis le virus de la photo et des inventions. Pendant que je suis dans les remerciements de la famille j'en profite pour remercier ma **famille élargie** ainsi que ma **belle-famille**.

Il y a une personne de très grande importance dans ma vie, qui m'accompagne et qui me soutient tout le temps : **Marie**. Merci ma douce épouse chérie.

Mes remerciements vont aussi à tous mes amis, qu'ils soient d'Alpha, d'Isèreanybody? ou d'ailleurs, pour nos différents échanges, vos soutiens pendant ces trois années. Je cite dans l'ordre alphabétique **Adrien, Agnès, Alain, Alban, Albane, Alix, Anaïs, Anne, Anne-Claire, Anne-Gab, Anthony, Antoine, Arnaud, Augustin, Aurélie, Baptiste, Benoît, Bernadette, Bertrand, Bruno, Caroline, Cécile, Charlotte, Chloé, Chris, Christophe, Christopher, Claire, Clémence, Colin, Corentine, Cyril, David, Delphine, Elise, Edward, Elizabeth, Elodie, Emma, Emmanuel, Eric, Etienne, Fanny, Florence, Florent, Florian, Francesca, Franck, François, Jean-Matt, Gabrielle, Gaëlle, Geoffroy, Grégoire, Guénaël, Guilhem, Guillaume, Guillemette, Gwendo, Hannah, Hélène, Hugues, Jean, J-Jo, J-B, Jean-Luc, Jean-Max, Jérémy, Jérôme, Justine, Koumi, Laura, Laurène, Livia-Marie, Loïc, Lucie, Ludo, Marc, Marcelle, Margaux, Marie, Marie-Hélène, Marie-Laure, Marie-Liesse, Marion, Marthe, Martin, Mathieu, Mathilde, Matthieu, Maud, Menehould, Michel, Myriam, Nathanaël, Nicolas, Nina, Olivier, Paola, Pascale, Patrick, Paul, Pauline, Pema, Perrine, Pierre, Quentin, Raphaël, Rebecca, Rémi, Richard, Rodolphe, Roshan, Sabine, Sébastien, Simon, Sophie, Stéphanie, Sylvain, Thérèse, Thibault, Thomas, Vincent, Yuxi, Yves et ceux que j'oublie !**

Enfin, en dernier ressort je souhaite remercier infiniment celui qui est mon rocher, qui donne un sens à ma vie et me redit « Je suis le Chemin, la Vérité et la Vie » : **Jésus-Christ**.

Table des Matières

| | |
|---|----------|
| Remerciements | a |
| Table des Matières | c |
| Liste des Figures | f |
| Liste des Tables..... | k |
| Introduction Générale..... | 1 |
| Chapitre 1 Modélisation de l'impact du tremblement humain sur le flou de bougé... 10 | |
| 1.1 État de l'art sur le tremblement humain et sa modélisation | 11 |
| 1.1.1 Les origines physiologiques | 11 |
| 1.1.2 Les caractéristiques du bougé humain..... | 13 |
| 1.1.2.1 Méthodes de mesure et caractéristiques du tremblement physiologique | 13 |
| 1.1.2.2 Caractéristiques du bougé de l'appareil photo | 19 |
| 1.1.2.3 Conclusion..... | 23 |
| 1.1.3 Modèle de bougé dans la littérature | 23 |
| 1.1.3.1 Modélisation du tremblement avec une approche mécanique | 23 |
| 1.1.3.2 Modèle de déplacement linéaire..... | 24 |
| 1.1.3.3 Modèle de marche aléatoire | 24 |
| 1.1.3.4 Conclusion sur l'insuffisance des modèles | 25 |
| 1.2 Modèle de tremblement proposé | 26 |
| 1.2.1 Hypothèses pour la construction d'un nouveau modèle de tremblement adapté aux besoins | 26 |
| 1.2.2 Notre modèle de tremblement physiologique de la main..... | 28 |
| 1.2.2.1 Vue d'ensemble..... | 28 |
| 1.2.2.2 Détails..... | 28 |
| 1.2.2.3 Conclusion..... | 30 |
| 1.2.3 Notre modèle photographique du bougé | 30 |
| 1.2.3.1 Du tremblement au bougé | 30 |

| | | |
|-------------------|---|-----------|
| 1.2.3.2 | Modèle du système optique..... | 31 |
| 1.2.3.3 | Modèle d'intégration du bougé sur le capteur..... | 32 |
| 1.3 | Comparaison et validation..... | 35 |
| 1.3.1 | Comparaison subjective aux autres modèles..... | 35 |
| 1.3.1.1 | Formes de bougés..... | 35 |
| 1.3.1.2 | Comparaison aux sorties gyroscopes..... | 37 |
| 1.3.2 | Validation statistique..... | 38 |
| 1.4 | Conclusion sur notre modèle de génération de flous de bougé..... | 40 |
| 1.4.1 | Avantages et points nouveaux..... | 40 |
| 1.4.2 | Conclusion sur le modèle..... | 41 |
| Chapitre 2 | Métrique de qualité d'image pour le flou de bougé..... | 42 |
| 2.1 | État de l'art sur la qualité d'image..... | 43 |
| 2.1.1 | Mesure liée à la perception visuelle..... | 43 |
| 2.1.1.1 | L'homme comme seul juge..... | 43 |
| 2.1.1.2 | Les mesures psycho-visuelles développées..... | 46 |
| 2.1.1.3 | Conclusion et limitations des mesures psycho-visuelles..... | 50 |
| 2.1.2 | Mesures objectives..... | 51 |
| 2.1.2.1 | Positionnement du problème..... | 51 |
| 2.1.2.2 | Métriques orientés autofocus..... | 52 |
| 2.1.2.3 | Métriques orientées compression d'image..... | 53 |
| 2.1.2.4 | Métriques généralistes..... | 54 |
| 2.1.2.5 | Métriques spécialisées dans le flou..... | 56 |
| 2.1.2.6 | Métriques orientées optiques de l'œil..... | 57 |
| 2.1.2.7 | Métriques orientées chaîne de l'image de l'appareil photo..... | 58 |
| 2.1.2.8 | Conclusion et limites sur les métriques objectives existantes..... | 58 |
| 2.2 | Métrique de qualité proposée pour le flou de bougé..... | 61 |
| 2.2.1 | Limite des méthodes de l'état de l'art..... | 61 |
| 2.2.2 | Hypothèses pour la construction d'une métrique de bougé..... | 61 |
| 2.2.3 | Fonctionnement de notre métrique..... | 63 |
| 2.3 | Expériences visuelles de validation de la métrique..... | 64 |
| 2.3.1 | 1 ^{ère} expérience de « <i>just noticeable difference</i> »..... | 64 |
| 2.3.1.2 | Résultats de la 1 ^{ère} expérience JND..... | 74 |
| 2.3.2 | 2 ^{ème} expérience de règle de qualité ajustable..... | 76 |
| 2.3.2.2 | Résultats de l'expérience et de la performance de prédiction de la métrique.. | 79 |
| 2.3.2.3 | L'influence des conditions de visualisation..... | 81 |

| | |
|---|-------------|
| 2.3.3 Application de la métrique au flou de bougé lors d'une prise de vue | 84 |
| 2.3.3.1 Ce que souhaite l'utilisateur en terme de qualité d'image | 84 |
| 2.3.3.2 Du temps de pose au bougé sur l'imageur | 86 |
| 2.4 Conclusion sur la métrique de Qualité | 89 |
| Chapitre 3 Systèmes de stabilisation d'image | 91 |
| 3.1 Etat de l'art sur la stabilisation des images | 91 |
| 3.1.1 Position du problème..... | 92 |
| 3.1.2 Solution classique de stabilisation d'images | 93 |
| 3.1.3 Autres solutions de stabilisation existantes | 98 |
| 3.1.4 Imageur rapide..... | 101 |
| 3.2 Algorithmes et architectures économiques pour la stabilisation | 102 |
| 3.2.1 Nouvelle approche pour l'estimation de mouvement sans gyroscope | 103 |
| 3.2.1.1 Méthodes d'estimation de mouvement existantes basées sur l'analyse d'image | 103 |
| 3.2.1.2 Apport du modèle de tremblement pour l'estimation de mouvement..... | 104 |
| 3.2.1.3 Description du système d'estimation de mouvement sans gyroscope | 107 |
| 3.2.2 Nouvelle approche pour la réduction du flou | 111 |
| 3.2.3 Impact des nouvelles solutions sur l'architecture de stabilisation | 112 |
| 3.2.3.1 Architecture sans gyroscope avec système de stabilisation optique | 113 |
| 3.2.3.2 Architecture sans gyroscope ni système de stabilisation optique..... | 118 |
| 3.2.3.3 Architecture sans système de stabilisation optique avec gyroscope | 120 |
| 3.2.4 Comparaison des performances de stabilisation des architectures..... | 122 |
| 3.2.5 Conclusion et limitations..... | 128 |
| 3.3 Nouveau paradigme : segmentation du temps de pose pour garantir la qualité de l'image | 132 |
| 3.3.1 Présentation de l'architecture de stabilisation avec consigne de qualité..... | 133 |
| 3.3.2 Résultats et conclusion | 135 |
| 3.4 Conclusion sur les algorithmes de stabilisation priorisant la netteté | 143 |
| Conclusions et Contributions | 149 |
| Bibliographie | i |
| Annexe 1 Instructions d'expérience de différence juste visible | xii |
| Annexe 2 Instructions pour l'expérience de règle de qualité d'image ajustable | xiii |
| Liste des Publications | xiv |

Liste des Figures

| | | |
|------------|--|----|
| Figure 1. | Evolution du nombre de photos prises chaque année. | 1 |
| Figure 2. | Illustrations des éléments de l'appareil photo ainsi que d'un module d'appareil photo pour téléphone portable. (1) Ecran de visualisation des images. (2) Imageur. (3) Carte mémoire. (4) Batterie. (5) Flash. (6) Obturateur. (7) Optique. | 2 |
| Figure 3. | Graphique synthétique des paramètres mis en jeux lors d'une prise de vue. | 5 |
| Figure 4. | Evolution de l'exigence de la qualité d'image perçue par l'observateur. | 6 |
| Figure 5. | Chaine de stabilisation d'image dans les appareils photos. | 7 |
| Figure 6. | Schéma de la démarche de la thèse avec les numéros de chapitre associés. | 8 |
| Figure 7. | Exemple de relevé d'expérience de mesure de bougé par laser [29]. Cas pour un temps de pose de 5 secondes. | 14 |
| Figure 8. | Exemple de tremblement physiologique de la main d'un participant de 31 ans [22] pendant une durée de 20 secondes. | 15 |
| Figure 9. | Exemple de relevé des histogrammes du spectre fréquentiel des déplacements du tremblement de la main et des doigts lorsque le bras du sujet est tendu. | 16 |
| Figure 10. | Densité spectrale de puissance de la rotation du poignet pendant une tache d'aller-retour du doigt entre le genou et le nez avec un patient ayant un tremblement essentiel [32]. | 17 |
| Figure 11. | Exemple de densité spectrale de puissance du tremblement de la main avec ses différentes sources. | 18 |
| Figure 12. | Mesure de bougé avec un appareil photo reflex Pentax K-7 avec <i>shake reduction</i> (SR) ou sans système de réduction du tremblement. Le bougé est approximé par un bougé linéaire. | 20 |
| Figure 13. | À gauche : bougé de l'appareil photo avec son axe long (rouge) et son axe court (vert). À droite : 20 bougés différents suivants des temps de pose allant de 0,01 à 0,8 seconde. | 21 |
| Figure 14. | Exemple de résultat des travaux de [10]. Évolution de l'écart type suivant le temps d'exposition sur l'axe long pour trois appareils photo avec des masses différentes. | 22 |
| Figure 15. | Cinq bougés linéaires aléatoires 2D avec des angles et vitesses différentes pour le même temps de simulation. | 24 |
| Figure 16. | Cinq marches aléatoires 2D différentes avec 100 pas pour le même temps de pose. | 25 |
| Figure 17. | Exemple de PSF convoluée avec une image idéale. La PSF est la « carte d'identité » de l'image floutée. | 27 |
| Figure 18. | Schéma fonctionnel du modèle de tremblement humain physiologique. | 28 |
| Figure 19. | Exemple de densités spectrales de puissance d'accélération, vitesse et déplacement du tremblement de la main, avec les différentes contributions des membres du corps humain. | 29 |
| Figure 20. | Les 6 degrés de liberté de l'appareil photo. | 30 |

| | | |
|------------|---|----|
| Figure 21. | Système optique simplifié de l'appareil photo avec ses 6 degrés de liberté. | 31 |
| Figure 22. | Modèle de l'appareil photo utilisé dans notre simulation. | 33 |
| Figure 23. | Exemple de notre modèle d'accumulateur spatial pour un déplacement sous pixel. Ligne pleine : déplacement sous pixel, ligne pointillée : position actuelle de la matrice de pixel. | 33 |
| Figure 24. | Modèle de simulation de flou de bougé pendant une prise de vue. | 34 |
| Figure 25. | Exemple de sortie du modèle de génération de flou de bougé. Bougé de rotation angulaire (haut), Matrices de PSF (milieu), Trois images résultantes à 100% de l'éclairage (bas). | 34 |
| Figure 26. | Comparaison des différents modèles pendant un temps d'exposition court (0,1 s). | 36 |
| Figure 27. | Comparaison des différents modèles pour un temps d'exposition long (0,8 s). | 36 |
| Figure 28. | Comparaison de sorties gyroscopes [52] (gauche), avec notre modèle (droite). | 37 |
| Figure 29. | Sorties gyroscopiques 2D de [40] comparées à notre modèle de tremblement. | 38 |
| Figure 30. | Comparaison de l'évolution des écarts-types suivant le temps de pose entre la courbe d'équation d'ajustement de [10], le bougé linéaire, la marche aléatoire et notre modèle. | 39 |
| Figure 31. | Réponses normalisées de la perception visuelle vis-à-vis des contrastes spatiaux. (a) Réponse de la luminance. (b) Réponse des composantes couleurs (notées C_{r1} et C_{r2}). | 45 |
| Figure 32. | La perception du flou est similaire pour les images en couleurs (a) ou en niveaux de gris (b). | 46 |
| Figure 33. | Courbe de l'équation 9 présentant les résultats en JND suivant le pourcentage de bonnes réponses. | 49 |
| Figure 34. | Sélection de scènes typiques pour les différentes expériences visuelles. | 65 |
| Figure 35. | L'interface graphique de l'expérience de JND. | 65 |
| Figure 36. | Schéma bloc représentant l'algorithme implémenté dans Matlab pour réaliser l'expérience de JND. | 66 |
| Figure 37. | Deux manières différentes de générer un flou de 1 pixel dans le domaine numérique. | 67 |
| Figure 38. | Exemple de résultats de deux observateurs différents ayant utilisé deux calibres différents. (a) Résultats p_1 de couple d'images suivant leur Δ flou linéaire, d'un observateur avec une vision normale. (b) Résultats p_2 de couple d'images suivant leur Δ flou linéaire, d'un observateur avec une bonne vision. | 70 |
| Figure 39. | Exemple d'exploitation de résultats pour un observateur. | 73 |
| Figure 40. | Résultats de l'exemple en JND de l'observateur. (a) Echelle linéaire. (b) Echelle logarithmique. | 74 |
| Figure 41. | Évolution de la perception visuelle pour chaque observateur suivant une dégradation de type flou linéaire. En orange la valeur moyenne de la perception visuelle des 20 observateurs. | 74 |
| Figure 42. | Prédiction de la métrique de qualité d'image (points) par rapport à la vérité de terrain pour l'évolution des flous linéaires (courbe). | 76 |

| | | |
|------------|--|-----|
| Figure 43. | Dix-huit différentes PSF obtenues avec notre modèle de tremblement suivant différents temps d'exposition. | 77 |
| Figure 44. | Interface graphique de l'expérience de règle de qualité d'image ajustable. | 78 |
| Figure 45. | Comparaison de la prédiction de la métrique (abscisses) avec la vérité de terrain issu de l'expérience d'ajustement de la qualité des images (ordonnées). L'écart-type entre observateurs pour chaque flou est représenté par la barre d'erreur. | 80 |
| Figure 46. | Exemple d'image floutée, cas de l'image dédoublée. | 81 |
| Figure 47. | Résultats de la 1 ^{ère} expérience visuelle de la section 2.3.1.2 montrant l'évolution de la qualité d'image perçue suivant la taille du flou linéaire. (a) Pour une distance de visualisation à 2,97 m. (b) Pour une distance de visualisation à 0,97 m (1 pixel vu à l'acuité visuelle). | 82 |
| Figure 48. | Ajustement de la qualité d'image après réduction de celle-ci. | 83 |
| Figure 49. | Évolution de l'exigence de la qualité des images au cours du temps exprimé en JND par [9]. 1979 : format 110 ; 1990 : format 35-mm ; 2002 : Qualité améliorée. | 84 |
| Figure 50. | Qualité subjective observée suivant la valeur de qualité objective en JND. | 85 |
| Figure 51. | Évolution de la qualité de netteté en JND suivant le temps de pose pour trois types d'appareils photo. | 88 |
| Figure 52. | Evolution du pourcentage de bonne image (Qualité > -3 JND) suivant le temps d'exposition pour 3 types d'appareils photo. | 88 |
| Figure 53. | Chaine de stabilisation d'image classiquement implémentée dans les appareils photos. | 92 |
| Figure 54. | Illustration de la stabilisation d'image dans le cas de lentille mobile. | 93 |
| Figure 55. | Illustration présentant la force de Coriolis. | 94 |
| Figure 56. | Système de stabilisation optique. (a) Stabilisation par déplacement de l'optique. (b) Stabilisation par déplacement de l'imageur. | 95 |
| Figure 57. | Gain des systèmes de stabilisation optique. | 97 |
| Figure 58. | Densité de probabilité de la taille des déplacements normalisés, vis-à-vis de la distance focale et de la taille des pixels, suivant la fréquence d'estimations de mouvement. | 105 |
| Figure 59. | Entropie de la taille de déplacement sur le plan imageur suivant la fréquence d'estimation de celui-ci. | 106 |
| Figure 60. | Schéma fonctionnel de l'estimation de mouvement sans gyroscope. | 107 |
| Figure 61. | Implémentation de l'estimateur de mouvement par mise en correspondance. | 109 |
| Figure 62. | Principe de prédiction pour estimer de manière efficace le déplacement de bougé. | 110 |
| Figure 63. | Schéma de principe présentant la sommation cohérente des trames. | 112 |
| Figure 64. | Détail de l'architecture de stabilisation d'image sans gyroscope avec système de stabilisation optique. | 114 |
| Figure 65. | Résultat de simulation de l'architecture sans gyroscope suivant la fréquence de fonctionnement du système. Deux types d'estimation de mouvement sont présentés celui de mise en correspondance (BM) et <i>Motion 2D</i> (M2D). . | 115 |
| Figure 66. | Qualité d'image lors de la stabilisation suivant la fréquence de cadence des imageurs avec leurs écarts types respectifs. | 117 |
| Figure 67. | Détail de l'architecture de stabilisation d'image sans gyroscope ni système de stabilisation optique. | 118 |

| | | |
|------------|--|-----|
| Figure 68. | Résultat de simulation de l'architecture de stabilisation sans gyroscope ni système de stabilisation optique en qualité d'image suivant la fréquence de fonctionnement du système. | 119 |
| Figure 69. | Détaille de l'architecture de stabilisation sans système optique avec gyroscope. | 120 |
| Figure 70. | Résultat de simulation de l'architecture de stabilisation sans système de stabilisation optique mais avec gyroscope en qualité d'image suivant la fréquence de fonctionnement du système. | 122 |
| Figure 71. | Comparaison des différentes architectures suivant leurs fréquences de fonctionnement pour le cas du smartphone avec un temps de pose de 1 seconde. A.1 et A.2 : Architecture de stabilisation avec système de stabilisation optique sans gyroscope utilisant l'estimateur de mouvement M2D (A.1) et BM (A.2). | 123 |
| Figure 72. | Résultat de l'entropie de la taille des déplacements du signal résiduelle de bougé suivant la fréquence trame pour l'architecture avec système de stabilisation optique sans gyroscope. | 124 |
| Figure 73. | Evolution de la fréquence de fonctionnement du système pour garantir une stabilisation optimale (-3 JND) suivant la distance focale de l'appareil photo pour un temps de pose 1/10 de seconde. | 125 |
| Figure 74. | Evolution du nombre de trame requis suivant la distance focale pour garantir une bonne qualité de service. | 126 |
| Figure 75. | Evolution de la fréquence de fonctionnement du système suivant le temps de pose afin d'obtenir une qualité de service définit à -3 JND. | 127 |
| Figure 76. | Evolution du nombre de trame des architectures de stabilisation suivant le temps de pose pour garantir une qualité de -3 JND. | 128 |
| Figure 77. | Chaîne simplifiée de l'acquisition d'un pixel de l'imageur avec les différentes sources de bruit. | 129 |
| Figure 78. | Schéma de la chaîne du bruit utilisée pour les simulations Matlab permettant de mesurer le PSNR de l'image. | 129 |
| Figure 79. | Evolution du PSNR d'une trame en fonction du nombre de trames requis pour les architectures de stabilisation des images. Leur seuil de limite de fonctionnement est présenté pour les estimateurs de mouvement <i>Motion 2D</i> (M2D) et <i>block-matching</i> (BM). | 131 |
| Figure 80. | Schéma de principe de l'évaluation du bougé par la métrique de qualité au cours du temps de pose. | 133 |
| Figure 81. | Détaille de l'architecture basée sur la qualité de service. | 134 |
| Figure 82. | 10 prises de vue de 1/2 seconde avec leurs trames ayant des temps d'intégration différents. | 135 |
| Figure 83. | Qualité des images perçues en JND avec ou sans stabilisation pour 10 prises de vue. | 136 |
| Figure 84. | Rendu visuel de l'architecture de stabilisation d'une prise de vue. (a) Image non stabilisée. (b) Image stabilisée (sans les sources de bruit parasites). (c) Image stabilisée (avec simulation du bruit). (d) Image de référence (scène). Pour visualiser les résultats correctement, attention à la résolution d'impression couleur (>180 dpi) ainsi qu'à la distance de visualisation (48,5 cm). | 137 |
| Figure 85. | (a) Histogramme normalisé du nombre de trames suivant la cadence de l'imageur et leur temps d'intégration. (b) Pourcentage cumulé du nombre | |

| | | |
|------------|---|-----|
| | de trames et du temps de pose en fonction de la fréquence maximum de fonctionnement de l'imageur et du temps d'intégration des trames. | 138 |
| Figure 86. | Principe de suppression des trames afin de relâcher la contrainte fréquentielle de l'architecture de stabilisation. Dans cet exemple, le temps d'intégration (Ti) final de l'image est de 88 % du temps de pose (Tp) requis de ½ seconde. | 139 |
| Figure 87. | Solutions afin d'obtenir l'exposition requise pour avoir une image suffisamment lumineuse : par augmentation du temps de pose ou par augmentation de l'ISO. | 140 |
| Figure 88. | Rendu de l'architecture de stabilisation d'image avec consigne de qualité fixée à -3 JND utilisant la technique de l'augmentation du temps de pose. (a) Image initiale comprenant toutes les trames. (b) Image avec seulement les trames générées à une fréquence inférieure à 60 Hz. (c) Image avec seulement les trames générées à une fréquence inférieure à 23 Hz. Pour visualiser les résultats correctement, attention à la résolution d'impression couleur (>180 dpi) ainsi qu'à la distance de visualisation (48,5 cm)..... | 141 |
| Figure 89. | Rendu de l'architecture de stabilisation d'image avec consigne de qualité fixée à -3 JND utilisant la technique de l'augmentation de l'ISO. (a) Image initiale comprenant toutes les trames. (b) Image avec seulement les trames générées à une fréquence inférieure à 60 Hz. (c) Image avec seulement les trames générées à une fréquence inférieure à 23 Hz. | 142 |
| Figure 90. | Synthèse des différentes architectures de stabilisation. | 144 |

Liste des Tables

| | | |
|------------|--|-----|
| Tableau 1. | Récapitulatif du type d'éclairage suivant la luminosité de la scène en lux avec l'équivalent en indice de luminance (EV)..... | 4 |
| Tableau 2. | Tableau comparatif sur la pertinence des différents modèles de bougé : Bougé linéaire, marche aléatoire et notre modèle proposé. | 40 |
| Tableau 3. | Avantages et inconvénients des deux différentes méthodes des mesures psycho-visuelles. | 50 |
| Tableau 4. | Etat de l'art sur les métriques de qualité d'image pouvant servir à mesurer le flou sur une image. Légende : D = Déplacement de l'appareil ; Opt = Optique de l'œil ; SR = Sans référence ; ? = Non disponible. | 60 |
| Tableau 5. | Résultats partiels de l'étude de l'influence de la distance pour deux PSF différentes simulant le même type de flou de 1 pixel..... | 68 |
| Tableau 6. | Exemple d'images ayant subi un flou de bougé linéaire avec leurs qualités objectives et subjectives respectives. Ces images ayant une définition de 180 dpi, une fois imprimée en 180 dpi ou plus, devront être visualisées à 48,5 cm afin de retrouver les résultats inscrits dans le tableau. | 86 |
| Tableau 7. | Etat de l'art des plus petits gyroscopes MEMS disponibles sur le marché adapté à la stabilisation d'image. | 95 |
| Tableau 8. | Etat de l'art des imageurs et camera les plus rapides ayant une résolution supérieur à 1 méga pixel..... | 101 |
| Tableau 9. | Comparatif des architectures de stabilisation d'image proposées ainsi que la stabilisation classique. Plus il y a d'étoiles, meilleure est la note. Les étoiles entourées s'ajoutent lorsque l'architecture avec consigne de qualité est ajoutée en surcouche. | 146 |

Introduction Générale

Avec l'intégration de capteurs d'images dans tous nos appareils mobiles tels que les téléphones, les tablettes, les appareils photo... nous prenons de plus en plus de photos. Aujourd'hui prendre une photo est quasiment à portée de tous. Par exemple en 2009 il a été estimé que 2,5 milliards de personnes sur la planète ont un appareil photo numérique dans leur téléphone portable [1]. Ceci explique l'explosion du nombre de photos prises chaque année ce que présente la figure 1, obtenue d'après les estimations de [2]. En 2011, 380 milliards de photos ont été prises. Ce qui fait un total d'environ 3,5 Téra (3500 milliards) d'images depuis l'invention de la photo.

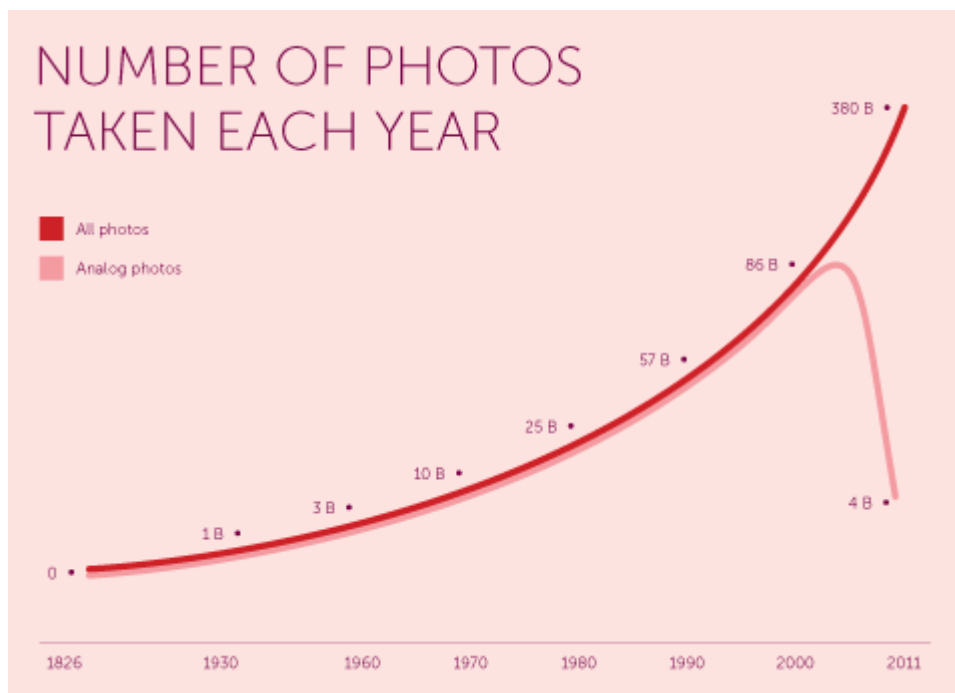


Figure 1. Evolution du nombre de photos prises chaque année¹.

Dans les années 1960 il est estimé que 55% des photos prises étaient celles de bébés. Aujourd'hui, les raisons de prendre une photo se sont démultipliées [3], que ce soit pour des

¹ Figure extraite de [2].

photos souvenir, pour des événements, des actions insolites... Toutes les raisons sont bonnes pour figer l'instant présent. De plus, quelle que soit la destination [4], les utilisateurs sont à la recherche de qualité d'image accrue. Un des principaux critères déterminants de qualité des images est la netteté de celle-ci. Lorsque la mise au point est réalisée, opération qui consiste à régler la netteté de l'image, seul le flou de bougé va réduire la netteté de celle-ci. Ce flou de bougé provient des tremblements de la main du photographe. Il va dépendre de la dextérité du photographe, du temps d'exposition et des performances de l'appareil photo.

Base des éléments de l'appareil photo

L'appareil photo est un système composé de parties optiques, mécaniques et électroniques. Afin de bien définir les termes pour la suite de ce manuscrit, une description des éléments de l'appareil photo est présentée figure 2.

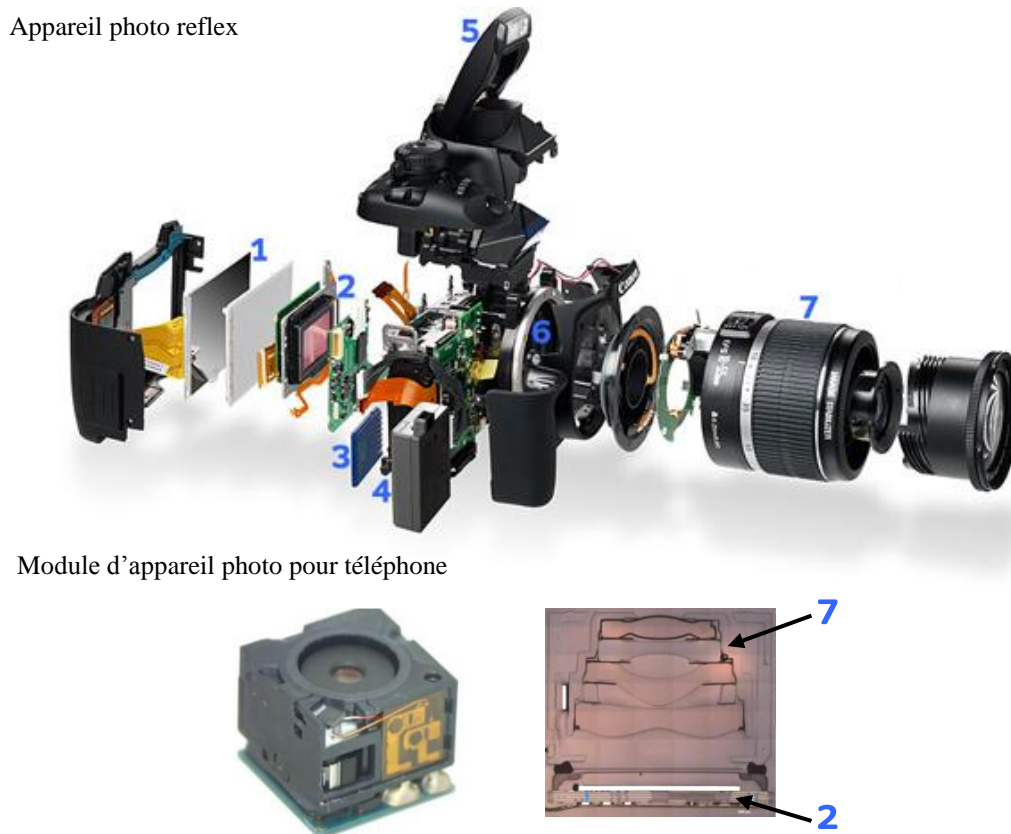


Figure 2. Illustrations² des éléments de l'appareil photo ainsi que d'un module d'appareil photo pour téléphone portable. (1) Ecran de visualisation des images. (2) Imageur. (3) Carte mémoire. (4) Batterie. (5) Flash. (6) Obturateur. (7) Optique.

² L'illustration présentant l'appareil photo reflex de manière éclatée est tirée de [173]. Les images du module Sony Ericsson Cyber-shot proviennent de [174].

Que ce soit dans l'appareil photo reflex, dans l'appareil photo compact ou dans l'appareil photo pour téléphone nous allons retrouver quasiment les mêmes éléments de manière plus ou moins miniaturisée, tels que présentés figure 2 :

- (1) Un écran de contrôle pour afficher ou prendre les photos.
- (2) Le capteur de l'appareil photo appelé l'imageur.
- (3) Une mémoire pour sauvegarder les images numérique.
- (4) Une batterie pour fournir l'énergie électrique à l'ensemble de l'appareil.
- (5) Un flash pour augmenter la luminosité de la scène en cas de besoin.
- (6) L'obturateur mécanique permettant de laisser passer la lumière ou non, ainsi que de régler la quantité de lumière entrant dans l'appareil grâce à la taille d'ouverture du diaphragme. Dans le cas des appareils photo pour téléphone cette fonction est directement réalisée électroniquement par l'imageur afin de gagner de la place.
- (7) L'optique, composée de différentes lentilles.

Dans le cas des appareils photo pour téléphone, le module photographique regroupe les éléments principaux de la fonction appareil photo du téléphone, c'est-à-dire les lentilles ainsi que l'imageur.

De la luminosité de la scène au flou de bougé

Pour prendre une photo le capteur de l'appareil photo doit recevoir assez de lumière afin de reproduire la scène. L'appareil capte les rayons lumineux réfléchis par la scène. Excepté lors de l'utilisation du flash, les rayons lumineux proviennent de sources externes telles que le soleil, l'éclairage et les étoiles suivant le type de scène. Ceci revient à dire que la scène contient sa propre luminosité ambiante. Avant de prendre une photo, l'appareil va ajuster trois paramètres pour capter suffisamment de luminosité afin d'avoir une image bien exposée. Malgré l'importance de ces paramètres et surtout depuis la démocratisation des appareils, l'utilisateur se soucie peu de leurs réglages. Ces paramètres sont la vitesse d'obturation, l'ouverture du diaphragme et la sensibilité ISO du capteur :

- La vitesse d'obturation, détermine combien de temps le capteur de l'appareil photo va recevoir les rayons lumineux de la scène. Cette durée s'appelle précisément le temps de pose. Plus le temps de pose sera long, plus la quantité de lumière captée sera importante.

- Le diaphragme est un système mécanique opaque circulaire qui permet le passage ou non des rayons lumineux vers le capteur. La largeur d'ouverture du diaphragme va déterminer l'intensité de lumière qui arrivera sur le capteur. Plus l'ouverture du diaphragme sera grande, plus l'intensité de lumière sera forte.
- La sensibilité ISO est la mesure de sensibilité des surfaces sensibles du capteur, elle est définie par la norme ISO 12232:1998 [5]. Par contre plus la sensibilité ISO sera importante, plus l'image sera bruitée.

Tableau 1. Récapitulatif du type d'éclairage suivant la luminosité de la scène en lux avec l'équivalent en indice de luminance (EV).

| <i>Luminosité [Lux]</i> | <i>EV (ISO 100)</i> | <i>Type de situation</i> |
|-----------------------------|-------------------------|---|
| 0,04 | -6 | Dans la nuit, loin des lueurs de la ville, sous les étoiles seulement |
| 0,08 | -5 | Dans la nuit, loin des lueurs de la ville, avec un premier croissant de lune |
| 0,16 | -4 | Dans la nuit, loin des lueurs de la ville, avec une demi-lune |
| 0,31 | -3 | Dans la nuit, loin des lueurs de la ville, avec la pleine lune |
| 0,63 | -2 | Dans la nuit, loin des lueurs de la ville, dans la neige sous la pleine lune |
| 1,25 | -1 | Sujet éclairé par une très faible lumière artificielle |
| 2,5 | 0 | Sujet éclairé par une faible lumière artificielle |
| 5 | 1 | Paysage lointain avec éclairage artificiel |
| 10 | 2 | Immeubles lointains avec éclairage artificiel, Lune en situation d'éclipse totale |
| 20 | 3 | Bâtiments, monuments et fontaines éclairés ou un feu d'artifice |
| 40 | 4 | Illuminations de Noël, sujet photographié à la lueur des bougies |
| 80 | 5 | Intérieurs faiblement éclairés, églises, sujet éclairé par un feu de camp |
| 160 | 6 | Intérieurs moyennement éclairés, foires, parcs d'attractions |
| 320 | 7 | Intérieurs éclairés, vitrines de magasins, spectacles sur scène, lieux de travail |
| 640 | 8 | Eclairage de stade de foot, éclairage sur scène |
| 1280 | 9 | Crépuscule 10 min après le coucher du Soleil, objet éclairés par des projecteurs |
| 2600 | 10 | Crépuscule juste après le coucher du Soleil |
| 5120 | 11 | Coucher du Soleil, galeries d'art |
| 10240 | 12 | Paysages par temps très nuageux, paysages juste avant le coucher du Soleil |
| 20480 | 13 | Sujet par temps nuageux clair sans ombres |
| 40960 | 14 | Temps brumeux avec ombres légères, arc-en-ciel sur un ciel nuageux |
| 81900 | 15 | Paysage en plein soleil avec ombres très marquées |
| 164000 | 16 | Soleil brillant sur sable ou neige |
| 328000 | 17 | Reflets du Soleil sur l'eau |

Avec une sensibilité standard de 100 ISO, une vitesse d'obturation ainsi que l'ouverture du diaphragme fixé, l'indice de luminance de la scène peut être calculé grâce à l'équation 1. Cet indice correspond à une quantité de lumière dans la scène. Cet indice de luminance, plus connu par *exposure value* (EV), est une mesure de l'exposition lumineuse sur une échelle logarithmique.

$$EV = \log_2 \left(\frac{N^2}{T_p} \right) \quad (1)$$

Avec N , l'ouverture et T_p , le temps de pose de l'appareil. Ainsi une correspondance peut être réalisée entre la luminosité de la scène exprimée en lux, les réglages de l'appareil photo exprimés en EV et le type d'éclairage de la scène présenté dans le tableau ci-dessus.

Tous ces paramètres photographiques tels que l'EV, ouverture, l'ISO et le temps d'exposition ont été synthétisés par [6] et présentés figure 3.

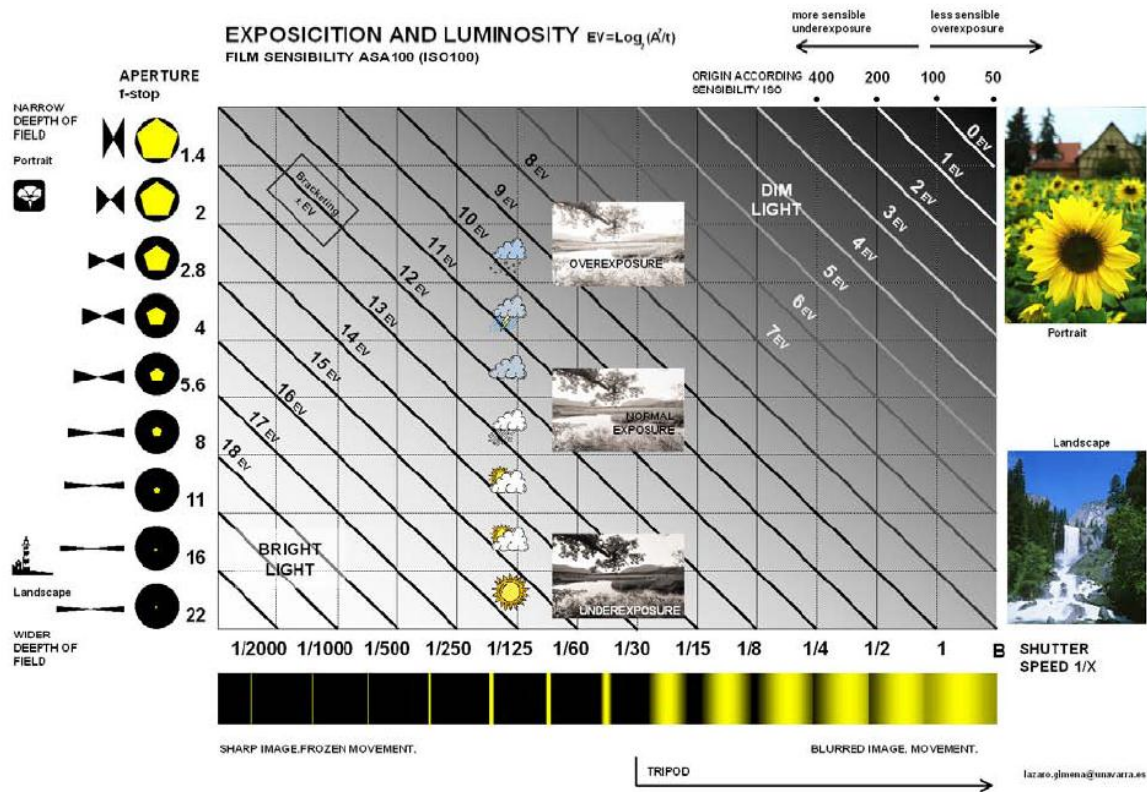


Figure 3. Graphique synthétique des paramètres mis en jeu lors d'une prise de vue.

Suivant les conditions d'éclairage et des réglages de l'appareil, un temps de pose optimum sera requis pour avoir une bonne luminosité sur le capteur. Ce graphique met en valeur que les trois raisons de l'augmentation du temps de pose seront :

- Une faible luminosité (prise de vue d'intérieur, de nuit...).
- Une sensibilité faible (faible ISO).
- Une ouverture petite (f-stop grand).

Par conséquent, plus le temps de pose est long, plus il est probable que le flou soit « important ». La notion d'importance du flou sera explicitée chapitre 2.

Tendances à l'accentuation du flou

Aujourd'hui, nous arrivons à deux situations conflictuelles. D'un côté, nous avons une tendance à l'accentuation du flou de bougé causé par la miniaturisation des capteurs [7, 8]. De l'autre côté l'utilisateur final est de plus en plus exigeant en termes de qualité d'image incluant une exigence plus élevée en termes de netteté comme le démontre les études de [9] présentées figure 4. Par la suite nous reviendrons sur ce graphique pour expliciter l'échelle et ces résultats.

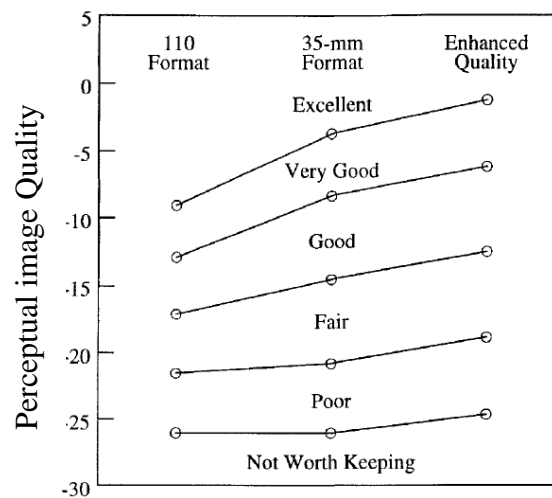


Figure 4. Evolution de l'exigence de la qualité d'image perçue par l'observateur.

Du fait de la miniaturisation des différents éléments tels que : l'optique, la taille du capteur et de ses pixels qui composent le module photographique, ces modules captent de moins en moins de lumière du fait de leurs petites tailles. Cela entraîne une augmentation du temps d'exposition pour compenser ce manque de nombre de rayon lumineux. De plus, au niveau du pixel, le même bougé va impacter plus de pixels d'autant que ceux-ci seront petits. La réduction des modules permettant leur intégration dans des appareils toujours plus petits entraîne un autre problème. Du fait de la réduction de taille des appareils, ils ont par conséquent des masses plus faibles. Ceci les rends plus sensibles à une perturbation externe tel que le bougé. Cette tendance a été observée par [10]. Par la suite nous détaillerons ces différents points.

A ces différentes raisons techniques de l'augmentation du bougé s'ajoutent d'autres raisons plus humaines. Du fait de la démocratisation des appareils photos, les conditions d'utilisation en basse lumière sont de plus en plus fréquentes. Conjugué à cela, des photographes de moins en moins expérimentés prennent de moins en moins de précaution

pour limiter le bougé (appareil tenu d'une main, à bout de bras...) [11]. Nous arrivons à une forte augmentation du flou de bougé sur l'image finale...

Solutions pour réduire ce flou

Heureusement, plusieurs solutions existent pour réduire le flou de bougé. La première solution consiste simplement à réduire le temps d'exposition et à compenser le manque de lumière par un gain électronique en augmentant l'ISO. Le problème de cette solution est que celle-ci augmente le bruit dans l'image. Ce qui entraîne une diminution de la qualité de l'image finale [9].

La deuxième solution consiste à utiliser un stabilisateur d'image. Son principe a été inventé en 1974 [12] mais le premier appareil photo avec cette fonctionnalité est arrivé sur le marché en 1994 avec le Nikon Zoom 700 VR [13]. Ceci a été permis grâce à la miniaturisation des gyroscopes permettant de capter les mouvements de l'appareil. Depuis, la chaîne de stabilisation des images dans l'appareil photo peut être synthétisée par la figure 5.

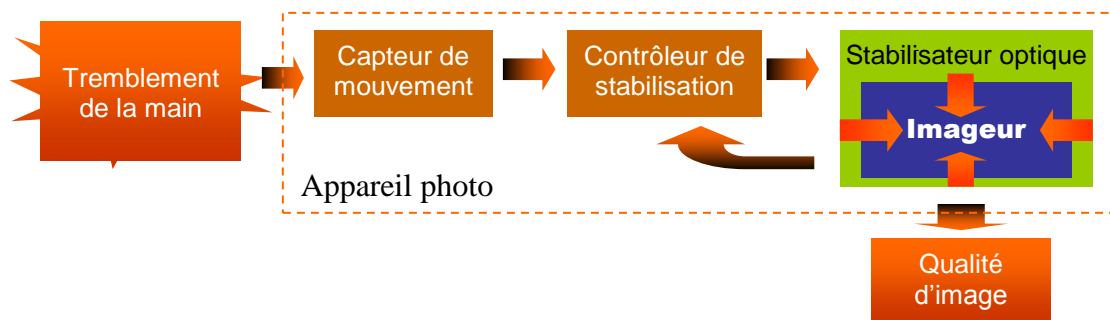


Figure 5. Chaîne de stabilisation d'image dans les appareils photos.

Les mouvements de l'appareil sont mesurés par un capteur de mouvement, le gyroscope. Il envoie ses résultats au contrôleur de stabilisation afin d'asservir le système de stabilisation optique. Prenant comme référence spatiale le capteur de l'appareil photo, appelé imageur, le système optique va déplacer l'axe optique afin de maintenir la projection de la scène stable sur le plan focal du capteur. Ainsi au cours de l'intégration de l'image sur l'imageur, les tremblements de la main seront compensés afin d'augmenter la netteté de l'image finale, entraînant une amélioration de la qualité des images.

Néanmoins les stabilisateurs d'image actuels présentent deux problèmes majeurs :

- Ils ont des modules de dimensions importantes qui requièrent des éléments externes. Cela limite leur intégration dans les téléphones portables et entraîne un coût supérieur.

De plus ces solutions, ne sont pas clé en main pour les fabricants [14]. C'est-à-dire que le fabricant doit ajuster les différents composants entre eux, que sont le capteur de mouvement, le contrôleur de stabilisation et le stabilisateur optique, au lieu d'avoir un seul et unique module réalisant la stabilisation.

- Ils ne tiennent pas en compte de la netteté finale souhaitée par l'utilisateur. Car ils n'arrivent pas à garantir la qualité de sortie de l'image.

Cette thèse vise à l'amélioration de la qualité des images en proposant des algorithmes de stabilisation d'image. L'objectif est, à terme, de répondre à ces problématiques en proposant des architectures de stabilisation qui permettent de supprimer les composants externes de la chaîne de l'image tout en garantissant une qualité d'image optimale.

Démarche de la thèse

Présentée de manière synthétique figure 6, la démarche de la thèse est la suivante. Pour atteindre une image stabilisée de qualité, une architecture de stabilisation d'image doit concilier de la meilleure manière : sa perturbation d'entrée qui est le bougé dû aux tremblements de la main, et son image de sortie qui est perçue par l'homme.

Nous proposons dans un premier temps de s'intéresser à l'entrée du système en allant jusqu'à la modélisation de celui-ci. Ainsi, une bonne connaissance et modélisation du flou de bougé nous permet d'avoir des entrées réalistes dans nos systèmes de stabilisation d'image.

Puis dans un second temps de s'intéresser à la sortie du système en allant jusqu'à la modélisation d'une métrique de qualité. Ainsi, par une connaissance fine de comment sont perçues les images, nous développons des architectures de stabilisation prenant en compte la qualité d'image de sortie.

De cette manière, par la prise en compte de la chaîne globale de stabilisation d'image, dans un troisième temps, nous développons des architectures de stabilisation permettant de garantir une image de bonne qualité.

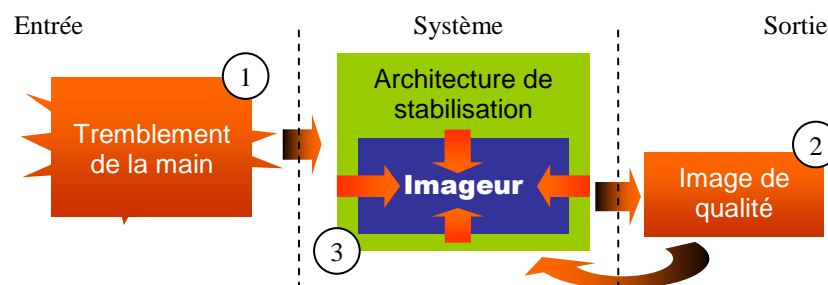


Figure 6. Schéma de la démarche de la thèse avec les numéros de chapitre associés

Organisation de la thèse

Le manuscrit est présenté de la manière suivante :

- Le chapitre 1 répond à la nécessité de bien connaître les mouvements que l'on souhaite stabiliser, en explorant l'origine des tremblements de la main dans le domaine physiologique. Nous les avons confrontés aux observations de bougé faites dans le domaine photographique. De ces études, nous avons réalisé le premier modèle de tremblement humain physiologique. Ce modèle nous permet de reproduire de manière fidèle les tremblements de la main, ainsi que leurs impacts sur l'image.
- L'absence potentielle de compensation de mouvement et la validation de techniques algorithmiques de compensation de flou nécessitent de définir une mesure du flou. Le chapitre 2 s'intéresse donc à la question : « Comment l'homme perçoit la qualité de netteté des images ? ». Pour répondre à cette question, j'ai développé une métrique de qualité objective. Cette métrique permet de quantifier la netteté perçue par l'homme afin d'optimiser les systèmes de stabilisation d'image sur ce critère de netteté.
- Le chapitre 3 présente les techniques de stabilisation que j'ai réalisées pour répondre aux limitations des architectures existantes. En effet, elles garantissent la netteté de l'image tout en réduisant le matériel requis. Ceci est rendu possible du fait que ces architectures prennent comme entrée le modèle de tremblement physiologique et comme critère de stabilisation la métrique de qualité, présentés aux chapitres 1 et 2 respectivement.

Chapitre 1 *Modélisation de l'impact du tremblement humain sur le flou de bougé*

Pourquoi un modèle de tremblement ? Il y a au moins trois bonnes raisons à cela :

- La première raison est qu'un modèle de tremblement permet de simuler des algorithmes de stabilisation d'image. L'avantage qu'il peut avoir par rapport à un enregistrement de tremblement, c'est que le modèle peut générer un nombre infini de cas.
- La deuxième raison est qu'un modèle de tremblement réaliste permet de fournir des stimuli d'entrée réalistes aux architectures de stabilisation. En effet, il y a une très forte dépendance entre le modèle de tremblement et l'architecture de stabilisation. Par exemple, stabiliser un bougé régulier linéaire en une dimension ne sera pas le même challenge que de stabiliser un bougé chaotique dans toutes les directions. De plus, en vue de l'intégration des systèmes de stabilisation un modèle de tremblement fidèle permettra une bonne optimisation de celui-ci.
- La troisième raison est qu'un modèle de tremblement réaliste permet de générer des images ayant des flous naturels tout en connaissant l'origine de ces flous. Car lorsqu'une image floue est prise avec un appareil photo, la connaissance des origines de ce flou est perdue. Ainsi grâce à une connaissance précise du flou de bougé, des origines et de sa formation sur l'image une métrique de qualité peut être développée pour mesurer le flou perçu par l'homme.

Donc dans ce premier chapitre, nous allons nous intéresser à la modélisation du tremblement humain. En effet, le modèle de tremblement doit être le plus fidèle possible étant utilisé comme source d'excitation pour les architectures de stabilisation. Il sera un paramètre clé pour dimensionner un système réel de stabilisation et il sera aussi utilisé pour développer une métrique de qualité de flou.

Ainsi, pour modéliser finement le tremblement, nous avons besoin de connaître ses origines, mais aussi de le mesurer pour quantifier sa nature en termes d'amplitude, de fréquence et statistique. Enfin, ce modèle permettra de simuler l'impact du tremblement sur l'image acquise.

Dans un premier temps, je vais présenter un état de l'art sur le tremblement humain appliqué à la photographie. Nous en dégagerons ses caractéristiques grâce aux études physiologiques ainsi qu'aux observations de terrain photographique. Puis un état de l'art sur les différents modèles de bougé existants que sont le modèle linéaire et le modèle de marche aléatoire sera présenté pour conclure sur l'insuffisance de ces modèles.

Dans un second temps, nous détaillerons notre modèle de tremblement pour qu'il reproduise de façon fidèle les tremblements de la main et leur impact sur l'image.

Dans un troisième temps, nous regarderons l'apport de notre modèle par comparaison aux modèles existants, ainsi que sa pertinence par rapport aux observations du bougé réel validé par une approche statistique.

Enfin nous concluons sur les avantages et les points nouveaux du modèle de tremblement physiologique.

1.1 État de l'art sur le tremblement humain et sa modélisation

1.1.1 Les origines physiologiques

Le tremblement humain est défini par [15] comme une oscillation rythmique involontaire d'un ou plusieurs membres du corps. Il est un mouvement naturel désordonné du corps humain [16]. Sa nature est très complexe puisqu'il peut être physiologique ou pathologique ; visible ou invisible ; fin ou grossier ; à basse ou haute fréquence ; symétrique ou asymétrique ; se manifester au repos ou uniquement pendant une activité [17]. Toutes les personnes en bonne santé ont un tremblement dit physiologique, une littérature abondante existe à son sujet [15, 16, 17, 18, 19, 20, 21, 22, 23, 24]. Le tremblement physiologique est bénin, de haute fréquence, avec de faibles amplitudes. Il est souvent invisible à l'œil nu, mais il peut devenir visible amplifié par la tenue d'une feuille de papier dans la main ou par le pointage d'un laser sur une cible à une distance significative.

Concernant les tremblements pathologiques, de nombreuses études ont aussi été réalisées pour en connaître leurs sources, leurs caractéristiques et leurs moyens de détection afin de soigner le patient. Par exemple en étudiant directement la réponse fréquentielle du tremblement, les chercheurs arrivent à en déduire le type de pathologie [25, 26]. Malgré l'abondance de ces résultats, ces dernières études sont en dehors de notre spectre de recherche car trop spécifiques. Nous allons donc nous restreindre aux tremblements physiologiques présents dans chaque être humain.

Suivant l'état d'activité des membres, les tremblements physiologiques peuvent être classés en deux groupes [17]. Les tremblements de repos et les tremblements d'action. Les tremblements au repos sont définis lorsque le corps humain est complètement supporté, lorsqu'il est compensé par la gravité. Les tremblements d'action sont définis lorsque les muscles du corps humain sont en contractions. Ils sont différenciés en tremblement de posture, tremblement isométrique, tremblement cinétique ou tremblement de tâche spécifique :

- Le tremblement de posture est visible lorsque le membre du corps maintient sa position contre la gravité, par exemple lorsque nous étendons devant nous le bras.
- Le tremblement isométrique est présent lorsque les muscles sont en contractions contre un objet stationnaire. Par exemple lorsque nous tenons un objet dans la main.
- Le tremblement cinétique paraît lorsqu'il y a un mouvement volontaire du corps. Par exemple en déplaçant la main vers une cible.
- Le tremblement de tâche spécifique se manifeste lorsque le corps réalise une action particulière comme jouer d'un instrument.

Ces tremblements physiologiques ont des origines encore imparfaitement connues des chercheurs. Celles qui sont présentées ci-dessous par [27] en sont les principales :

- La résonance mécanique qui est une oscillation passive mécanique des membres du corps. Elle provient de la réponse d'absorption des différentes parties du corps à l'injection du sang par le cœur. Suivant la raideur du muscle, sa masse, les membres du corps vont avoir une fréquence de résonance particulière.
- Une oscillation dite centrale qui est produite à l'origine par l'activité neuronale.
- L'activité musculaire associée au fonctionnement de la boucle de contre réaction du système nerveux qui n'est pas instantanée ni parfaite. Ce tremblement est

produit pendant l'activation des muscles. Elle suit la séquence suivante : d'abord le muscle agoniste initie le mouvement, puis le muscle antagoniste freine le mouvement et ainsi de suite.

En photographie, dans le cas d'une prise de vue sans trépied, les types de tremblement mis en jeu sont le tremblement de posture et le tremblement isométrique afin de serrer et maintenir l'appareil photo dans la position souhaitée. Ces tremblements proviennent de différentes sources présentées ci-dessus.

Afin de modéliser ce tremblement, il est nécessaire d'en connaître ses caractéristiques. Elles sont présentées dans la section suivante.

1.1.2 Les caractéristiques du bougé humain

1.1.2.1 Méthodes de mesure et caractéristiques du tremblement physiologique

Différentes méthodes de mesure ont été développées pour mesurer le tremblement humain. Il existe des méthodes de notation de la force du tremblement, utilisées pour détecter une pathologie et son importance. Ainsi que des méthodes quantitatives pour mesurer finement le tremblement humain en matière d'amplitude, de fréquence...

Comme nous cherchons à modéliser précisément le tremblement humain, nous allons plutôt nous intéresser aux méthodes quantitatives. Plusieurs méthodes ont été développées par les chercheurs. Elles sont présentées ci-dessous avec leurs résultats liés aux tremblements physiologiques de la main. Ces méthodes doivent s'appliquer dans la position du corps utilisée pour prendre des photos. Cette position particulière correspond à une position de posture.

Tablette Numérique

Il existe des méthodes de mesure par tablette numérique [28]. Cette méthode consiste à enregistrer le déplacement du stylo en contact avec la tablette numérique lorsqu'une personne rédige ou dessine des formes. Puis ces déplacements enregistrés sont ensuite traités et analysés de manière numérique pour en déduire les amplitudes et fréquences du tremblement. Ce type de mesure, étant avec contact, est mal adapté pour notre application qui est un bougé de posture sans contact.

Pointeur Laser

Les méthodes basées sur des pointeurs lasers sont souvent utilisées par les chercheurs comme le témoignent un nombre important de travaux utilisant ces techniques. Une cible est utilisée pour mesurer l'amplitude du tremblement. Pour cela, il est demandé au sujet de pointer le centre de la cible avec un pointeur laser. Un exemple de relevé d'expérience similaire réalisée par [29] est présenté figure 7 pour un bougé d'une durée de 5 secondes. Cette technique de mesure est adaptée pour mesurer les amplitudes des tremblements de la main, car elle est similaire aux conditions de prise de vue avec un appareil photo.

Les travaux de [23], utilisant la méthode du pointeur laser avec la main, montrent que les déviations dues au tremblement pour une durée de 1 seconde est au minimum de $0,07^\circ$ et au maximum de $0,28^\circ$. Avec, pour chaque sujet, des moyennes comprises entre $0,09^\circ$ et $0,22^\circ$, suivant la position du bras.

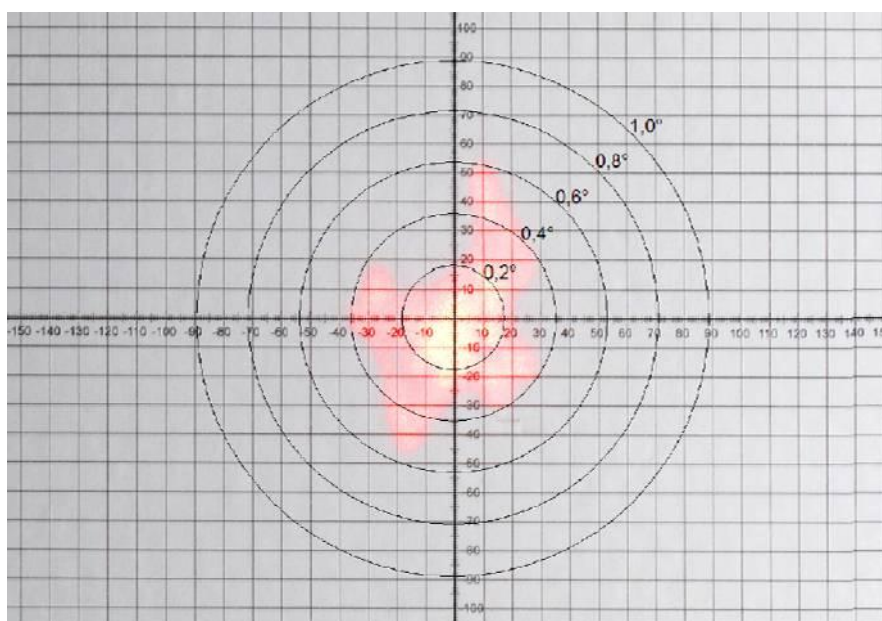


Figure 7. Exemple de relevé d'expérience de mesure de bougé par laser [29]. Cas pour un temps de pose de 5 secondes.

Ces résultats mettent en évidence que le tremblement de la main a une forte composante rotationnelle. Il nous donne aussi des amplitudes de rotation minimum et maximum pendant un bougé. Contrairement à ce que l'on peut penser, les travaux de [30] montrent que le fait de pouvoir regarder la cible et de demander au sujet de se concentrer pour réduire son tremblement va avoir l'effet inverse : l'augmentation de son tremblement. Cela est dû au fait que le sujet va avoir tendance à raidir les membres de son corps et ainsi à augmenter son tremblement.

Réflexion du laser

La méthode basée sur la mesure laser est utilisée dans l'étude du tremblement [22, 24]. Cette méthode permet une mesure précise du tremblement par la mesure de déplacement. Elle consiste à fixer une surface réflexive sur le membre à analyser (par exemple sur le bout d'un doigt). Puis à envoyer un rayon laser sur cette partie et enfin enregistrer la déviation du rayon due au tremblement du membre. Cette mesure est bien adaptée pour connaître précisément le tremblement de la main appliqué à la photographie, car elle est sans contact. Des résultats de mesures de [22] sont présentés figure 8.

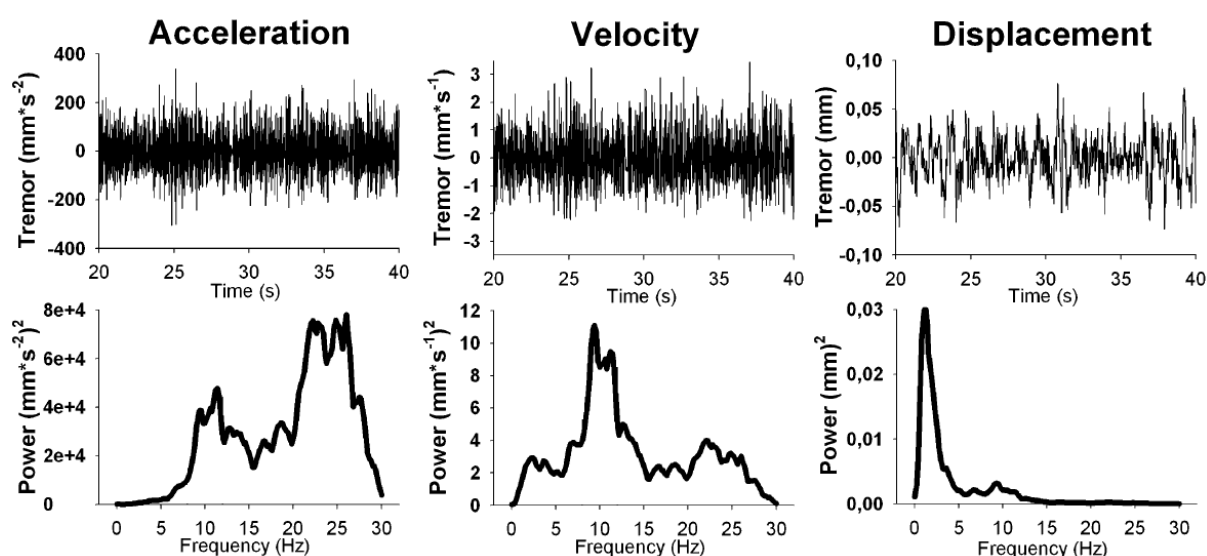


Figure 8. Exemple de tremblement physiologique de la main d'un participant de 31 ans [22] pendant une durée de 20 secondes.

Cet exemple de résultat montre d'une part l'évolution temporelle du tremblement en termes d'accélération, de vitesse et de déplacement, et d'autre part leurs réponses fréquentielles. Ces résultats montrent aussi les différentes contributions du tremblement de la main suivant les fréquences associées et leurs amplitudes. Les travaux de [22] confirment aussi que le tremblement de la main a une fréquence maximale de 30 Hz.

Accéléromètre

Les méthodes de mesure par accéléromètre comme dans [20, 19, 21, 31] ont été développées suite à l'apparition, sur le marché, d'accéléromètres intégrés, donc de faible masse et de faible encombrement. Cette méthode consiste à fixer des accéléromètres sur le membre à étudier, et à enregistrer l'accélération due au tremblement du membre. Cette mesure est adaptée, car les accéléromètres miniatures fixés sur les membres ne gênent pas le

tremblement du fait de leur faible masse. Un exemple de mesure des travaux de [20] sont présentés figure 9.

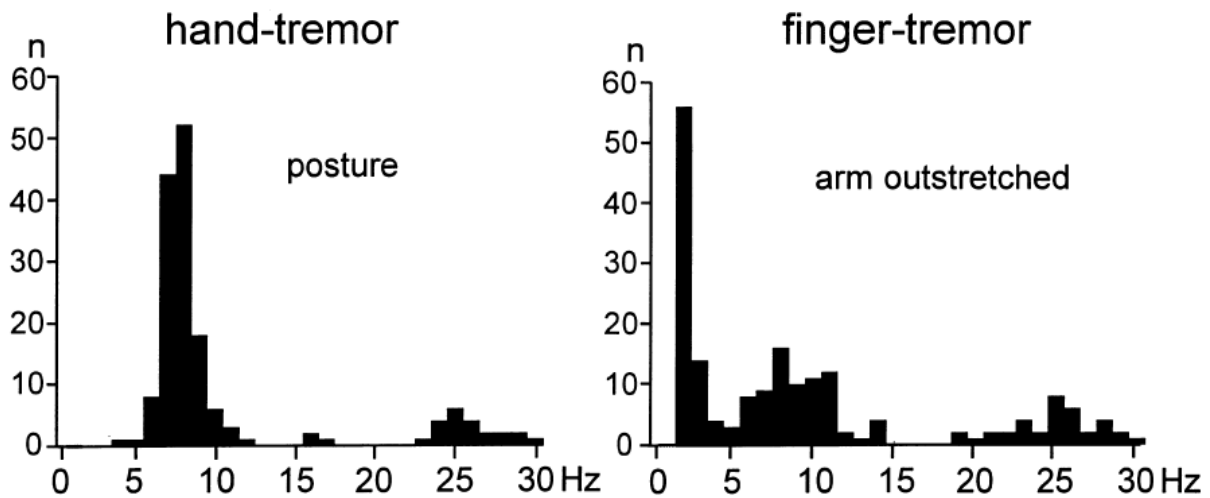


Figure 9. Exemple de relevé des histogrammes du spectre fréquentiel des déplacements du tremblement de la main et des doigts lorsque le bras du sujet est tendu.

Cette étude montre l'impact fréquentiel des différents membres du corps suivant différentes positions. Elle dévoile aussi que si la charge sur le bras augmente, il y a un déplacement de certaines fréquences vers les basses fréquences. Ajouter une charge se comporte d'une certaine manière comme un filtre passe bas sur la réponse fréquentielle. On peut remarquer une similarité de la forme de spectre des fréquences entre les mesures avec accéléromètres et celles par mesure laser.

Gyroscope

Les méthodes de mesure par gyroscope sont les plus récentes, elles sont rendues possibles grâce à la miniaturisation des capteurs appelés les MEMS pour *microelectromechanical systems*. De la même manière que pour la mesure avec accéléromètre, les gyroscopes MEMS sont directement placés sur le membre à étudier. Encore peu d'études ont été faites avec cette méthode de mesure du tremblement. Dans les travaux de [32] des données ont été extraites pour des tremblements pathologiques. Même si cette méthode de mesure semble la mieux adaptée, car elle rend compte directement des tremblements rotationnels des membres, les données pour le tremblement physiologique ne sont pas encore disponibles. Un exemple de ce type de mesure avec des gyroscopes est présentés figure 10.

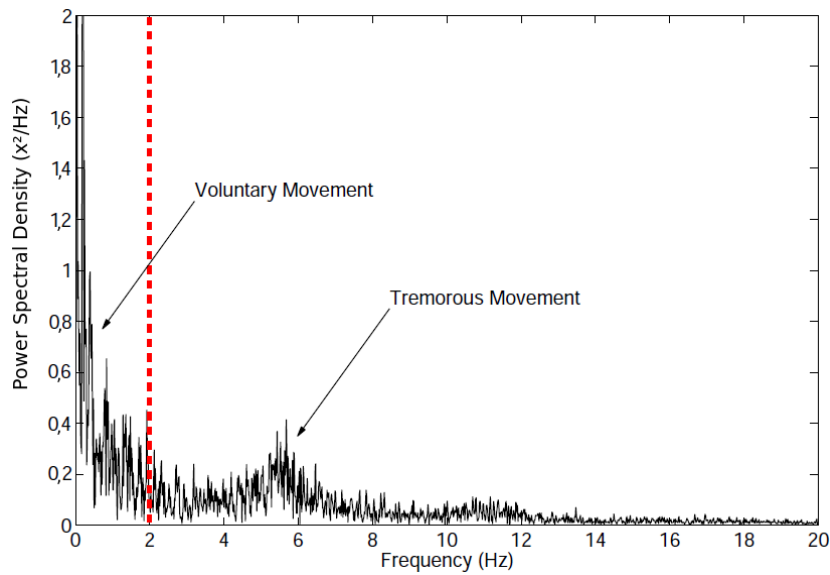


Figure 10. Densité spectrale de puissance de la rotation du poignet pendant une tâche d’aller-retour du doigt entre le genou et le nez avec un patient ayant un tremblement essentiel [32].

Ces résultats montrent une densité spectrale de puissance d’une rotation du poignet pendant une tâche d’aller-retour du doigt entre le genou et le nez, avec un patient ayant un tremblement pathologique dit essentiel. Grâce aux travaux de [33] il discrimine les mouvements volontaires des tremblements gênants du patient ayant des plages de fréquences supérieures à 2Hz. Dans la même optique, ces techniques pour détecter les tremblements grâce aux gyroscopes sont utilisées par les chercheurs pour développer des systèmes de suppression de tremblement dans les appareils de microchirurgie afin d’assister les chirurgiens pour les opérations qui demandent beaucoup de précision [34, 35, 36]. Le but de ces systèmes est de supprimer des mouvements volontaires, les mouvements parasites présentant des fréquences situées au-dessus de 2Hz.

Malgré la pertinence d’utiliser des gyroscopes pour étudier le tremblement, ces résultats ne sont pas exploitables pour notre recherche. Ils concernent en effet les tremblements pathologiques qui présentent des amplitudes bien plus importantes, avec une répartition spectrale différente des tremblements physiologiques.

Conclusion

Les résultats de ces différentes études et observations avec différentes méthodes montrent une convergence dans l’analyse spectrale du tremblement humain de la main, que ce soit avec la méthode des accéléromètres, ou avec la méthode laser. Concernant les méthodes utilisant

les gyroscopes, celle-ci ne sont pas encore disponibles pour le tremblement physiologique. Nous pouvons ainsi résumer synthétiquement que le tremblement est borné en terme d'amplitude maximale, qu'il est composé de différentes composantes spectrales qui ont de multiples origines et contributions. Les différentes contributions sont présentées figure 11 pour un tremblement physiologique de la main.

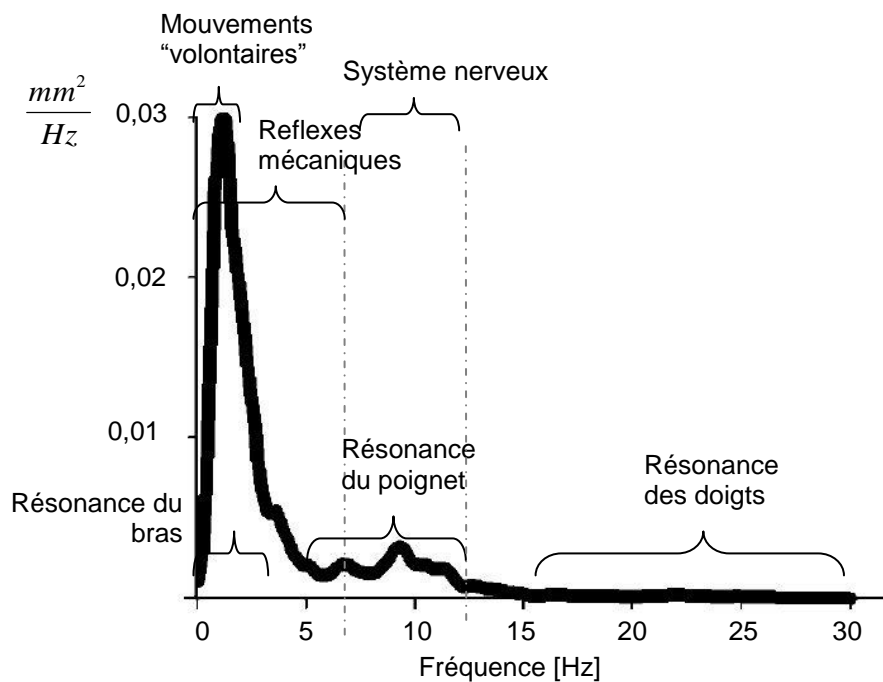


Figure 11. Exemple de densité spectrale de puissance du tremblement de la main avec ses différentes sources.

Basée sur les études physiologiques du tremblement, cette figure montre les différentes plages de fréquences de contribution des résonances des membres et de sources de tremblement. Même si les chercheurs se sont intéressés à identifier d'où proviennent ces différentes sources de résonances avec précision, elles sont difficilement séparables entre elles car plus ou moins liées les unes aux autres. Les fréquences approximativement comprises entre 1-3 Hz sont plutôt la contribution de la résonance mécanique du bras, entre 5-12 Hz celle du poignet et entre 16-30 Hz celle des doigts [37]. Les fréquences de tremblement inférieures à 7Hz sont plutôt identifiées par les chercheurs comme provenant des réflexes mécaniques [38]. Pour les fréquences comprises entre 8 et 12 Hz, les chercheurs pensent que c'est plutôt une source de tremblement due au système nerveux [39]. Ainsi, nous pouvons voir que la contribution du tremblement par les doigts en termes d'amplitude de déplacement n'est pas très importante par rapport aux autres contributions, mais par contre elle est plutôt

de haute fréquence (> 15 Hz). Les principales résonances sont celles du poignet et du bras. Comme mesurées par [33], les fréquences inférieures à 2Hz sont plutôt dues à des mouvements volontaires. Mais comme des tremblements sont encore présents alors que le sujet recherche l'immobilité, ils sont dus aux mouvements « volontaires » du corps pour ajuster la position du membre afin de le stabiliser. En matière d'amplitude ils sont importants mais de basse fréquence.

Complétant les études basées sur les accéléromètres, les études utilisant les lasers montrent que le tremblement est aussi de nature rotationnel dû aux articulations du poignet. Ainsi, le tremblement de la main est un mouvement désordonné suivant les six degrés de liberté de la main. Ces tremblements vont se propager à l'appareil photo pendant une prise de vue. Ces études de tremblement directement faites à partir de l'appareil photo sont présentées dans la section suivante.

1.1.2.2 Caractéristiques du bougé de l'appareil photo

Amplitude

Analogue à la méthode de pointage laser pour mesurer les amplitudes de tremblement de la main, [29] utilise cette méthode, mais en fixant le laser sur l'appareil photo. Un exemple de résultat est présenté figure 7. L'expérience est faite sur 6 sujets avec un temps de pose de 5 secondes dans trois conditions d'utilisation différentes. Une première condition d'expérience est réalisée avec un appareil photo reflex tenu à deux mains. Puis les suivantes sont réalisées avec un appareil photo bridge tenu à une main, l'une en regardant l'écran LCD et l'autre en regardant dans le viseur optique de l'appareil.

Les résultats montrent que les amplitudes minimum et maximum de rotation sont de 0,3-0,8 degré respectivement. Dans le cas avec l'appareil reflex tenu à deux mains, le bougé dû au tremblement est le plus faible avec un bougé moyen proche de 0,45 degré. Dans cas de la tenue de l'appareil photo bridge avec une seule main, le bougé est plus important, avec une moyenne de 0,5 degré.

Les résultats de cette expérience donnent le même ordre de grandeur que les études avec seulement le pointeur laser. Même si le temps du bougé est cinq fois plus long que dans les expériences de [23], les amplitudes ne sont pas cinq fois plus grandes. Il y a un effet de saturation au cours du temps d'exposition. Au regard du petit nombre d'expériences et de sujets, il est difficile de conclure sur une loi générale de bornage du bougé de l'appareil photo. Par contre, les ordres de grandeur peuvent être retenus.

Vitesse angulaire

Les travaux de [40] s'appuient sur des mesures de vitesse grâce aux gyroscopes. Cela permet d'étudier le bougé induit par les usagers sur les petits appareils photo, tels que les Smartphones. D'après les résultats expérimentaux, ils montrent que pour des temps de pose inférieurs à 100 ms, le bougé de la caméra peut être fidèlement approché par un mouvement linéaire. Pour les temps d'exposition plus longs, ils en dérivent une vitesse de bougé effective définie comme étant la taille de l'axe du motif du bougé le plus long divisé par le temps de pose. Cela leur permet de démontrer que la règle générale issue des formats d'appareils photo 35 mm, disant que le temps de pose doit être inférieur à l'inverse de la distance focale pour avoir une image toujours nette, n'est plus valide pour les cas d'appareils avec un faible masse ou avec une utilisation à une main. Cette étude démontre aussi que le fait de tenir l'appareil photo avec une seule main entraîne un bougé plus important qu'avec une préhension à deux mains.

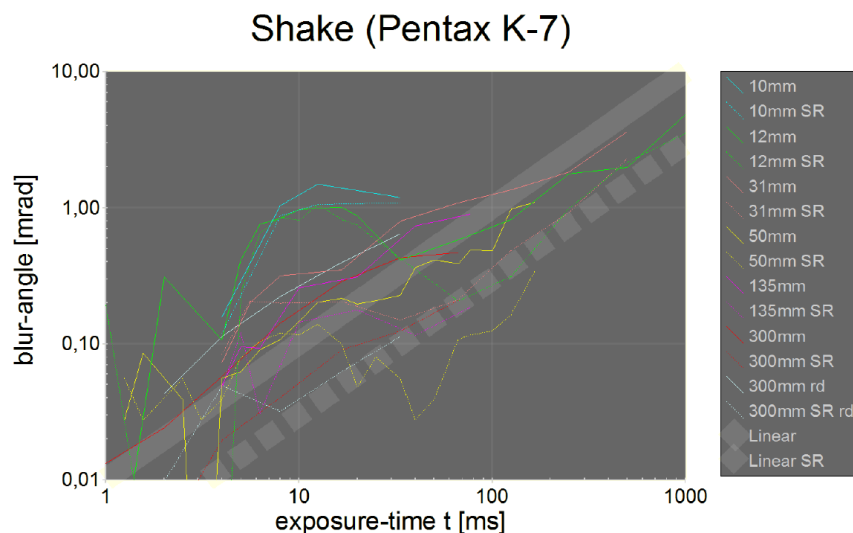


Figure 12. Mesure de bougé avec un appareil photo reflex Pentax K-7 avec *shake reduction* (SR) ou sans système de réduction du tremblement. Le bougé est approximé par un bougé linéaire.

Les travaux de [41] s'intéressent à approcher le tremblement comme un bougé linéaire pour en déduire le gain réel d'un système de stabilisation. Les observations du tremblement directement faites sur l'appareil photo [41] sont présentées figure 12.

Cette expérience, réalisée avec 4 sujets suivant 7 optiques différentes, montre l'évolution du bougé suivant le temps d'intégration. Nous pouvons remarquer que l'amplitude des rotations du bougé de l'appareil photo a tendance à être plus importante avec le réglage d'une

petite distance focale. Cela s'explique par le fait que plus la distance focale est grande, plus l'optique est lourde donc plus le tremblement va être atténué grâce à l'inertie massique de l'appareil photo.

Dans leurs observations, pour des temps d'expositions jusqu'à une seconde, le bougé est approximé par un bougé linéaire avec une vitesse moyenne de rotation angulaire constante de 10 mrad/s (0,573 °/s) dans le cas sans stabilisation. Ainsi, le bougé B_{pixel} sur l'image finale en unités de pixel est approximé comme étant l'équation linéaire :

$$B_{pixel}(t) = v \times t \times \frac{f}{T_{pixel}} \quad (2)$$

avec v la vitesse du bougé, t le temps de pose, f la distance focale et T_{pixel} la taille du pixel de l'imageur.

Statistique

Plusieurs études statistiques [10, 42] s'intéressent à l'évolution de l'écart type du bougé suivant le temps d'exposition. Dans les travaux de [10] les mesures sont faites à partir de 7 sujets photographiant une LED dans le noir suivant différents temps d'exposition avec 3 appareils photo différents. Ainsi, chaque image enregistre la forme du bougé en niveau de gris dû aux tremblements pendant le temps d'exposition présenté figure 13.

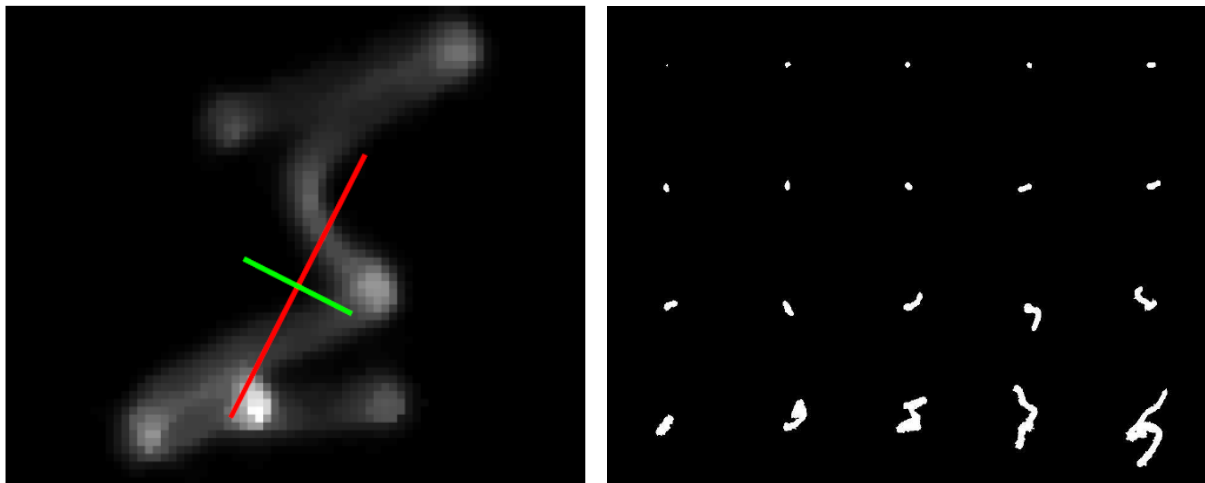


Figure 13. À gauche : bougé de l'appareil photo avec son axe long (rouge) et son axe court (vert). À droite : 20 bougés différents suivants des temps de pose allant de 0,01 à 0,8 seconde.

Ensuite pour chaque image, l'écart type du bougé en est déduit suivant l'axe court et l'axe long. L'axe court et l'axe long sont déterminés grâce à l'analyse en composantes

principales de la forme du bougé [43]. Pour chaque temps d'exposition, l'écart type a été ensuite moyenné selon toutes les prises de vues, puis approximé par une courbe en puissance. Un exemple de résultat statistique suivant l'axe long de tous les sujets est présentée figure 14.

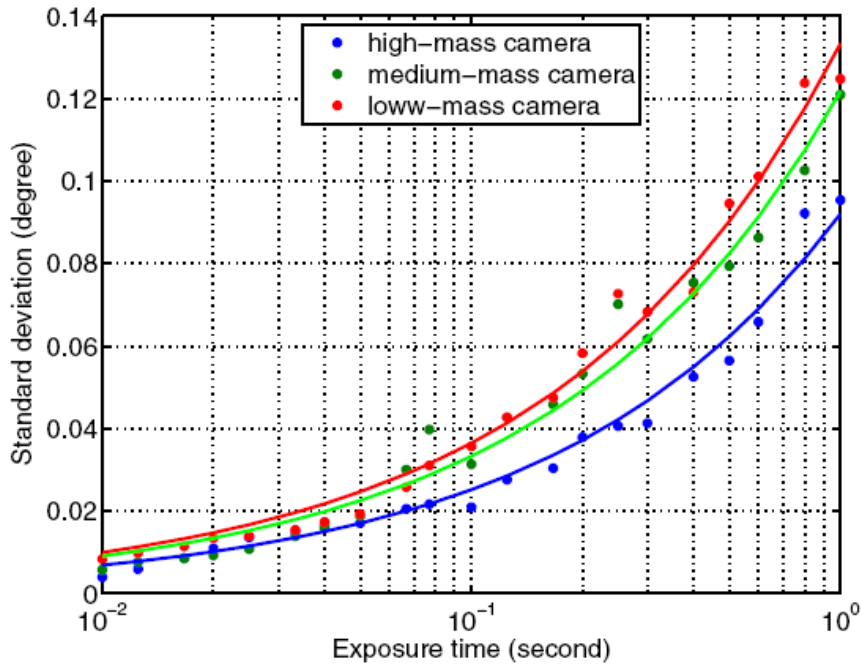


Figure 14. Exemple de résultat des travaux de [10]. Évolution de l'écart type suivant le temps d'exposition sur l'axe long pour trois appareils photo avec des masses différentes.

Les points de mesure au cours de l'intégration ont été ajustés à une loi temporelle de puissance d'équation :

$$\sigma_{Flou}(T_{Int}) = aT_{Int}^b \quad (3)$$

Avec σ_{Flou} l'écart type du flou, T_{Int} le temps d'intégration de l'image et a , b des coefficients d'ajustement. Il a été calculé que le coefficient d'ajustement optimal pour b est de 0,5632. Par contre, le coefficient a dépend fortement de la masse de l'appareil photo et de la dextérité du photographe. Cette étude met aussi en avant que, pendant les quelques premières dizaines de millisecondes, la forme du bougé est quasiment linéaire. Alors que pour des temps d'expositions plus longs, les formes de bougé évoluent vers des motifs aléatoires. Ces observations appuient la pertinence du coefficient d'ajustement b de la loi de puissance qui évolue comprise entre un déplacement linéaire ($b = 1$) et une marche aléatoire ($b = 0,5$). Ces travaux montrent aussi que plus l'appareil photo est lourd, moins le bougé est important et que plus le photographe est expert, moins il a tendance à bouger du fait de sa dextérité.

1.1.2.3 Conclusion

D'une part, le tremblement physiologique de la main est défini par les scientifiques comme une oscillation involontaire caractérisée en termes d'amplitude et de fréquence par les observations physiologiques.

D'autre part, nous avons des observations faites directement sur l'appareil photo en termes temporel, fréquentiel et statistique. Ces observations montrent que le motif du bougé est de type linéaire pour les temps d'exposition court, alors que pour les temps d'exposition long il est de type marche aléatoire.

Afin d'en avoir une représentation analytique, des modèles de bougé ont été réalisés pour simuler des flous de bougé. Ils sont présentés dans la section suivante.

1.1.3 Modèle de bougé dans la littérature

Pour simuler des flous de bougé, plusieurs méthodes peuvent être envisagées :

- Soit utiliser des modèles empiriques généralement utilisés dans la littérature telle que le déplacement linéaire ou la marche aléatoire.
- Soit à partir d'une approche mécanique.

1.1.3.1 Modélisation du tremblement avec une approche mécanique

Des modèles mécaniques du bras ont été développés par les chercheurs pour étudier les vibrations dans les membres du bras, que ce soit par modèle de simulation [44, 45, 46] ou par reconstitution de bras mécaniques [47]. Des stimuli d'excitation de nature différente sont utilisés en entrée du système mécanique pour en étudier les tremblements.

Même si le tremblement humain met à contribution le corps humain et que celui-ci peut être modélisé par un système mécanique simplifié, il est difficile d'en tirer un modèle fin de tremblement humain pour les raisons suivantes :

- Il y a un premier problème qui concerne le lieu et le type d'application de la source du tremblement. Le lieu d'application de la source de bruit n'est pas évident à définir. Par exemple réduire la source de tremblement à une agitation en bout de main ne permet pas de montrer la contribution de chaque membre pour le tremblement global. De plus le type de source d'entrée est difficile à définir car, comme nous l'avons vu au

paragraphe 1.1.2.1, ce phénomène reste mal connu même s’il est parfois modélisé par un bruit gaussien ou un sinus avec une amplitude et une fréquence variable.

- Il y a un deuxième problème qui concerne le comportement dynamique du modèle mécanique. Il paraît difficile de prendre en compte les boucles de contre-réaction du système, les séquences de muscles pour réaliser un mouvement de stabilisation de posture...

1.1.3.2 Modèle de déplacement linéaire

Le modèle de déplacement linéaire (*straight line walk*) est la plus simple modélisation d’un mouvement aléatoire. Malgré cela, il est le plus utilisé pour modéliser un bougé d’appareil photo soumis aux tremblements comme dans les travaux de [48] qui cherche un optimum entre le temps de pose, le gain ISO et la netteté de l’image. Dans la même démarche, pour avoir une image toujours nette, le choix du temps de pose est choisi suivant la distance focale. Les photographes utilisent la règle générale : d’avoir un temps de pose inférieur à l’inverse de la distance focale. Ainsi [40] approxime cette loi comme un flou de bougé linéaire maximum à ne pas dépasser.

Le bougé linéaire est alors modélisé par le tirage aléatoire d’une vitesse constante avec un angle aléatoire de départ.

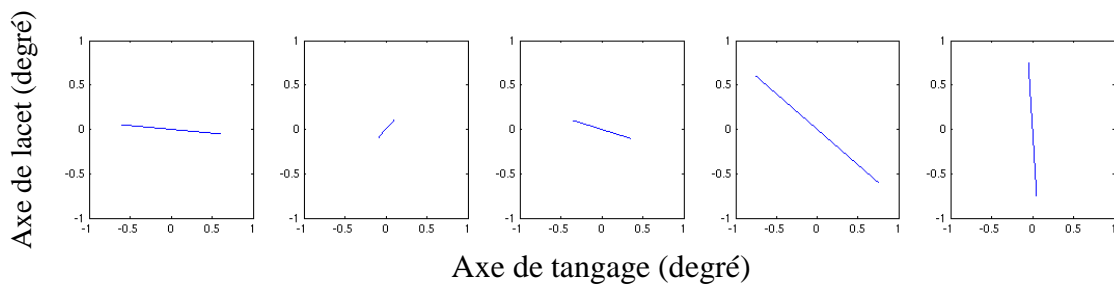


Figure 15. Cinq bougés linéaires aléatoires 2D avec des angles et vitesses différentes pour le même temps de simulation.

1.1.3.3 Modèle de marche aléatoire

Le modèle de marche aléatoire (*random walk*) est un modèle qui décrit la trajectoire d’un objet se déplaçant de manière aléatoire à chaque instant sans tenir compte du passé [49]. Un exemple de marche aléatoire est celui du mouvement Brownien [50] qui est la description mathématique du déplacement d’une particule dans un fluide. De manière analogue le

tremblement de la main, par sa nature désordonnée, peut être décrit par un modèle de marche aléatoire. Ainsi dans un premier modèle, nous avons décrit la source du bougé comme un modèle de marche aléatoire, pour modéliser un flou de bougé. Ce modèle repose sur :

- Des pas de taille aléatoire suivant une répartition gaussienne.
- Pour chaque pas, une direction aléatoire.

Voici des exemples de marche aléatoire en deux dimensions.

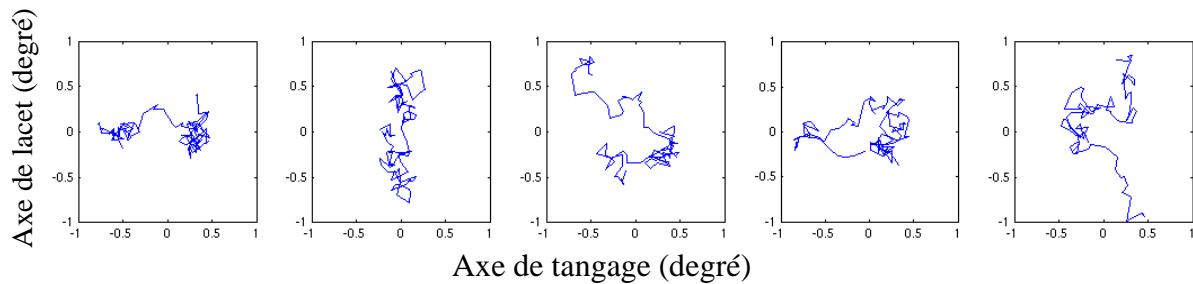


Figure 16. Cinq marches aléatoires 2D différentes avec 100 pas pour le même temps de pose.

1.1.3.4 Conclusion sur l'insuffisance des modèles

Aucun de ces modèles n'a été confronté aux observations physiologiques du tremblement humain. Néanmoins si l'on compare ces modèles aux connaissances du tremblement et aux observations faites sur les appareils photo nous pouvons en déduire que ces modèles sont limités pour reproduire fidèlement le bougé dû au tremblement :

- Du fait de leurs limites de modélisation, les modèles mécaniques de la littérature par exemple [44] n'arrivent pas à obtenir une réponse fréquentielle du tremblement similaire à ceux observés par les études physiologiques. En conséquence, leurs réponses temporelles pour modéliser le tremblement ne sont pas fidèles.
- Le modèle linéaire peut être bien approprié pour les courts temps d'intégration, car le bougé s'apparente à un bougé linéaire pour des temps inférieurs à quelques dizaines de millisecondes [10, 40]. Par contre, pour des temps d'exposition plus longs, il est clair que ce modèle n'arrive pas à reproduire un bougé avec des directions aléatoires.
- Le modèle de marche aléatoire semble plus approprié pour des temps d'exposition longs. Mais inversement, pour des temps d'exposition court, il semble trop aléatoire. De plus si l'on regarde la forme de sa réponse fréquentielle qui est du type $1/f^2$, nous

remarquons qu'il n'arrive pas à reproduire les propriétés spectrales du tremblement de la main.

En conclusion, aucun modèle existant n'arrive à reproduire de manière fidèle les observations faites par les études physiologiques et photographiques concernant le tremblement. Malgré le fait qu'un modèle reflétant précisément les caractéristiques du bougé permet de définir le cahier des charges du système de contrôle. En particulier, se passer d'accéléromètre implique d'effectuer une estimation du mouvement à partir de l'image elle-même. Ainsi la cadence d'acquisition, la manière de suivre le mouvement est très dépendante du type de bougé d'entrée. C'est pourquoi, les modèles existant ne peuvent pas répondre aux besoins précis de la problématique de stabilisation de flou de bougé en photographie. En conséquence, nous avons développé un modèle de bougé basé sur les études du tremblement physiologique.

1.2 Modèle de tremblement proposé

1.2.1 Hypothèses pour la construction d'un nouveau modèle de tremblement adapté aux besoins

La fonction d'étalement du point, ou plus connu en anglais par *point spread function* (PSF), est un filtre passe bas bidimensionnel décrivant la réponse d'un système d'imagerie à une source ponctuelle. En d'autre terme, la PSF est la « carte d'identité » du flou de l'image. La convolution de ce filtre de PSF avec une image sans bougé, appelée image idéale par la suite, génère une image floutée ayant la quantité d'information de flou de la PSF comme présenté figure 17. Elle permet donc de modéliser avec précision un flou de bougé sur une image finale. L'objectif de notre modèle de tremblement, présenté par la suite, est de reconstruire une PSF temporelle qui représente fidèlement le flou de bougé humain évoluant au cours du temps.

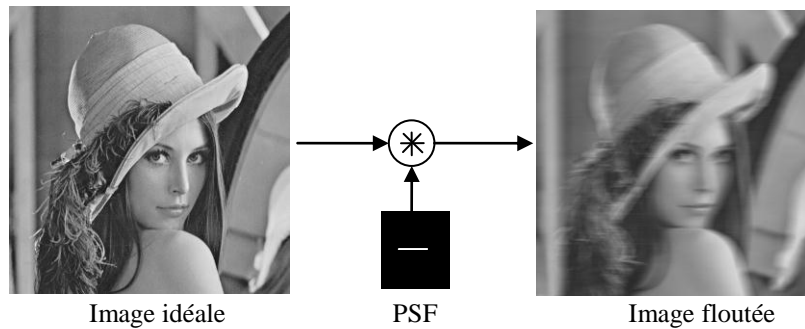


Figure 17. Exemple de PSF convoluée avec une image idéale. La PSF est la « carte d'identité » de l'image floutée.

Pour cela, nous avons besoin d'un modèle aléatoire reflétant les caractéristiques physiologiques du tremblement humain. Cela permettra d'obtenir un générateur de bougé aléatoire fidèle qui servira de vérité de terrain. Il permettra aussi d'avoir de nouvelles informations analytiques sur les caractéristiques du bougé humain. Et de simuler l'impact réel du bougé pendant une prise de vue en photographie.

Nos objectifs ont été de construire un modèle de tremblement humain et d'appareil photo qui :

- a les mêmes caractéristiques fréquentielles, temporelles et statistiques que les différentes observations faites dans la littérature ;
- a un motif de bougé aléatoire semblable à ceux qui sont observés afin de simuler un flou de bougé au cours du temps d'intégration de l'appareil photo de manière fidèle.

Pour y parvenir, notre modèle de tremblement a été réalisé sous Matlab en deux parties principales :

- Le modèle physiologique qui modélise le tremblement de la main.
- Le modèle de l'appareil photo qui tient compte des caractéristiques optiques et d'intégration de celui-ci.

Ce modèle est détaillé dans la section suivante.

1.2.2 Notre modèle de tremblement physiologique de la main

1.2.2.1 Vue d'ensemble

Le cœur de notre modèle de tremblement humain se base sur la reproduction d'un filtre permettant de simuler les comportements de tremblements humains. Nous construisons un filtre de convolution temporelle à partir de la transformée de Fourier inverse (TFI) d'une densité spectrale de puissance (DSP) d'accélération du tremblement de la main mesurée. Ce filtre émule le système mécanique du corps épaule bras main doigt. Ensuite, ce filtre est alimenté par des signaux aléatoires générant ainsi un tremblement aléatoire.

1.2.2.2 Détails

Le modèle de tremblement, basé sur les observations physiologiques, est synthétisé par le schéma figure 18.

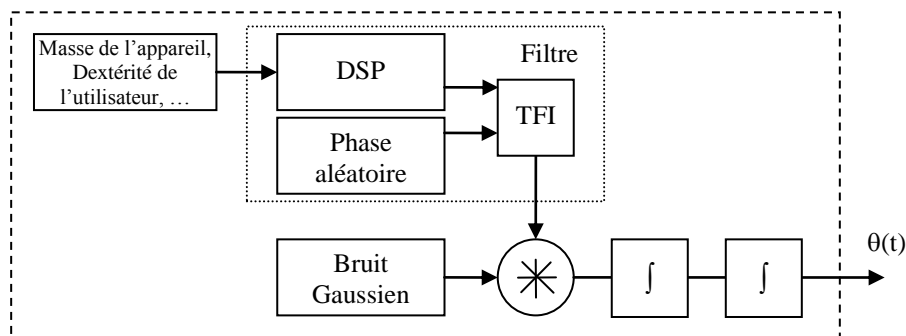


Figure 18. Schéma fonctionnel du modèle de tremblement humain physiologique.

À partir de la densité spectrale de puissance (DSP) d'accélération des études de [22] présentée figure 19, nous réalisons un filtre pour reproduire le comportement du corps humain. La construction de ce filtre consiste d'une part à reconstruire le spectre de puissance à partir de la DSP, et d'autre part à reconstruire la phase de ce filtre. Pour la reconstruction du spectre de puissance, l'amplitude des raies de fréquence sont calculées à partir de la DSP. Cette reconstruction du spectre de puissance peut être paramétrée suivant la masse de l'appareil photo, la dextérité de l'utilisateur... Par exemple comme le démontre l'étude sur le tremblement physiologique soumis à une masse [20], l'augmentation de la masse de l'appareil photo va avoir tendance à déplacer certaines raies de fréquence vers de plus basses fréquences (analogie avec un filtrage passe bas). En ce qui concerne la phase, pour chaque raie de

fréquence, nous reconstruisons sa phase en ajoutant une phase aléatoire. Cela nous permet de simuler les différents états initiaux du corps humain ainsi que d'ajouter du hasard à notre modèle de simulation. Ce choix d'ajouter du hasard à notre modèle a été motivé par les premiers résultats de bougé obtenus avec une phase constante. En effet, les bougés étaient trop similaires entre eux malgré des formes de bougé différents, ce qui n'était pas conforme aux observations des bougés observés en photographie.

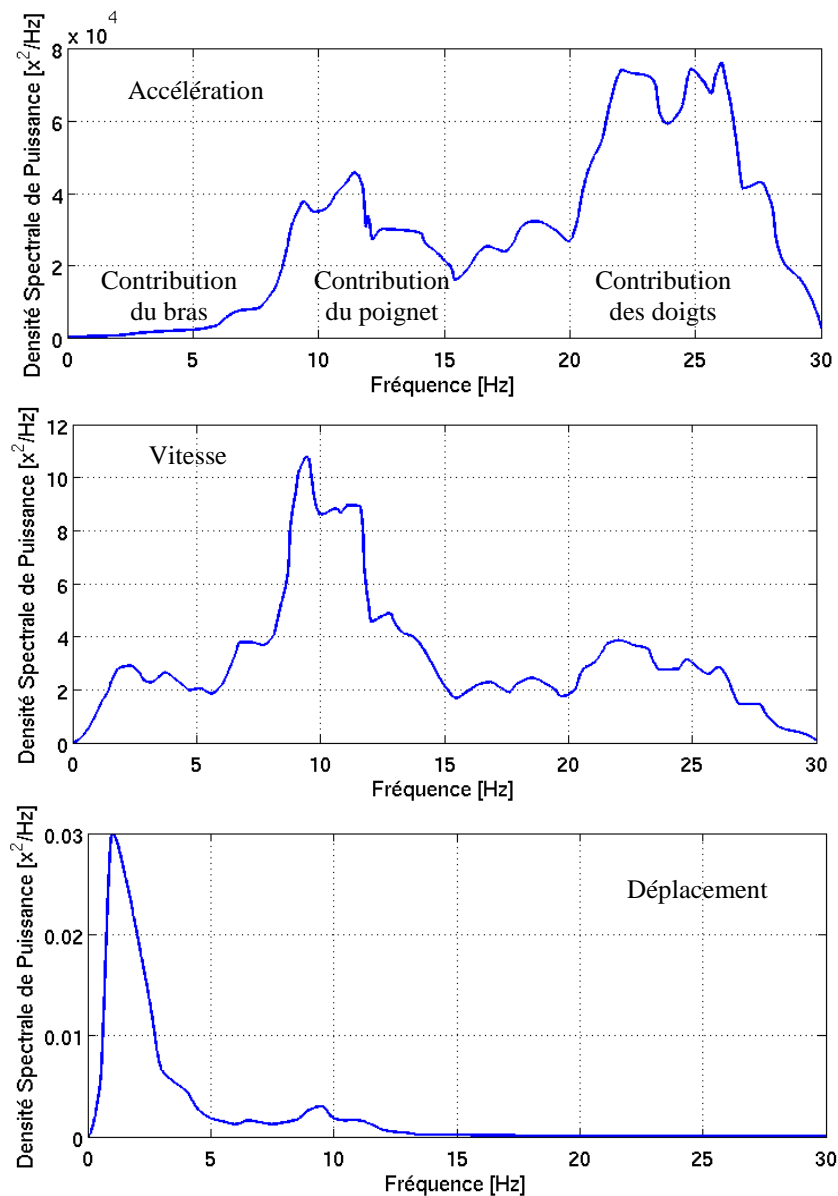


Figure 19. Exemple de densités spectrales de puissance d'accélération, vitesse et déplacement du tremblement de la main, avec les différentes contributions des membres du corps humain.

Ce filtre fréquentiel de puissance est ensuite converti en filtre de convolution temporelle. Ce filtre temporel imite la réponse du système mécanique comprenant l'épaule, le bras, le

poignet et les doigts du corps humain de l'utilisateur. Il est ensuite alimenté par un bruit blanc gaussien simulant l'activité du système nerveux des muscles [39] afin de produire en sortie un signal temporel angulaire aléatoire d'accélérations. Ce signal d'accélération est ensuite intégré deux fois au cours du temps pour générer des signaux de vitesse et de rotation respectivement. Ainsi, les signaux engendrés en sortie sont analogues à ceux des tremblements de la main.

1.2.2.3 Conclusion

Grâce à la construction de ce filtre temporel, nous pouvons reproduire le tremblement physiologique des mains. La réponse fréquentielle des signaux générés en sortie de notre modèle de tremblement de la main, est conforme à la DSP mesurée par les physiologistes.

Ainsi, grâce à ce modèle, nous disposons de signaux de tremblement réalistes permettant de simuler le comportement de la main. Ce tremblement de la main va ensuite impacter l'appareil photo en terme de bougé. Sa modélisation est présentée dans la section suivante.

1.2.3 Notre modèle photographique du bougé

1.2.3.1 Du tremblement au bougé

Lors d'une prise photographique, le tremblement de la main impacte les six degrés de liberté de l'appareil photo, comme présenté figure 20. Ce tremblement physiologique est responsable du flou de bougé sur l'image finale après intégration de l'imageur pendant le temps de pose.

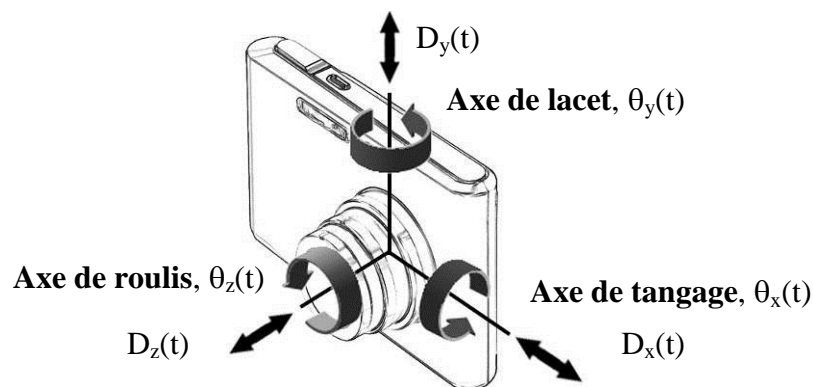


Figure 20. Les 6 degrés de liberté de l'appareil photo.

L'intensité du flou de bougé va dépendre du temps d'intégration correspondant à la durée du temps de pose. Le système optique de l'appareil photo ne va pas propager les

tremblements de la même manière sur chacun des six degrés de liberté. Ces six degrés de liberté comprennent trois rotations (θ_x , θ_y et θ_z) et trois translations (D_x , D_y et D_z).

Ainsi, nous avons besoin d'un modèle dual optique, spatio-temporel pour passer des degrés de liberté de la main à l'image de bougé final.

1.2.3.2 Modèle du système optique

Le système optique simplifié de l'appareil photo peut être schématisé comme ci-dessous :

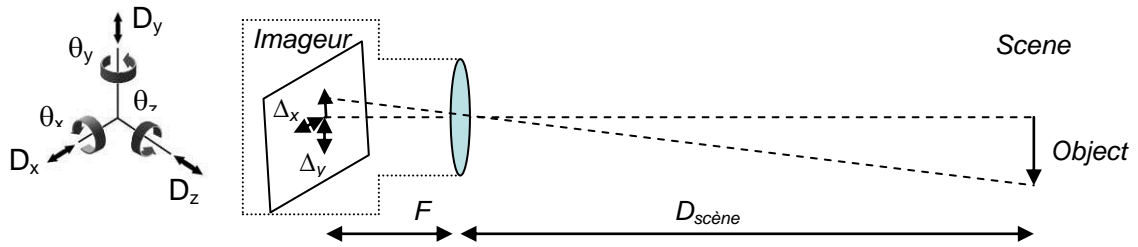


Figure 21. Système optique simplifié de l'appareil photo avec ses 6 degrés de liberté.

Le système optique de l'appareil photo compte six degrés de liberté qui ont différents impacts sur le flou généré résultant du bougé. Si nous regardons les équations du système optique de l'appareil photo dans les conditions normales d'utilisation au niveau du plan focal des pixels de l'imageur, nous avons :

$$\begin{bmatrix} \Delta_x \\ \Delta_y \end{bmatrix} = F \begin{bmatrix} \tan(\theta_x) \\ \tan(\theta_y) \end{bmatrix} + \frac{F}{D_{scène}} \begin{bmatrix} D_x \\ D_y \end{bmatrix} + f(x, y, D_z, \theta_z) \quad (4)$$

avec Δ_x et Δ_y le déplacement résultant dans le plan focal de l'imageur, F la distance focale, $D_{scène}$ la distance entre la lentille de l'appareil photo et la scène. Pour des variations données θ_z , D_z , la fonction f de transformation de la rotation et de l'homothétie dépend de la position des points en (x, y) sur le plan focal et peut-être délimité par :

$$f(x, y, D_z, \theta_z) < 2d_{x,y} \tan(\theta_z) + \frac{d_{x,y} D_z}{D_{scène} + D_z} \quad (5)$$

avec $d_{x,y}$ la distance entre le point considéré et l'axe optique. Dans des conditions typiques, $D_{scène} \gg F \gg d_{x,y}$. La distance à la scène est beaucoup plus importante que les autres distances. L'équation 4 peut donc être simplifiée en :

$$\begin{bmatrix} \Delta_x \\ \Delta_y \end{bmatrix} \approx F \begin{bmatrix} \tan(\theta_x) \\ \tan(\theta_y) \end{bmatrix} \quad (6)$$

De plus, comme les tremblements de bougé observés par [23, 29, 40, 41] sont de quelques degrés au maximum, nous pouvons simplifier l'équation 6 par :

$$\begin{bmatrix} \Delta_x \\ \Delta_y \end{bmatrix} \approx F \begin{bmatrix} \theta_x \\ \theta_y \end{bmatrix} \quad (7)$$

Cette équation illustre bien pourquoi il a été largement observé que suivant les six degrés de liberté de l'appareil photo de la figure 20, seuls les axes de tangage et de lacet dominant le flou de bougé dans les conditions standard d'utilisation. Cela a été par exemple expérimenté par [51] en cherchant à utiliser une caméra pour mesurer des rotations et des translations. Dans cette expérience il a été observé que même pour des translations de plusieurs centimètres, les translations sur le capteur étaient trop faibles pour être mesurées. Ainsi pour des faibles translations de l'appareil photo, dans les conditions normales de prise de vue, les translations n'ont aucun impact sur l'image. Par conséquent, dans notre modèle, nous avons réduit les six degrés de liberté de l'appareil photo en deux degrés de liberté.

Si maintenant nous récrivons l'équation 7 pour ramener les déplacements Δ_x et Δ_y en déplacements pixeliques nous obtenons :

$$\begin{bmatrix} \Delta_x \\ \Delta_y \end{bmatrix} \approx \frac{F}{T_{pix}} \begin{bmatrix} \theta_x \\ \theta_y \end{bmatrix} \quad (8)$$

avec T_{pix} , la taille des pixels. C'est pourquoi pour le même tremblement initial, plus la taille des pixels sera petite, plus le nombre de pixel impacté par le tremblement sera important. C'est pour cela que la réduction de la taille des pixels augmente le flou intrinsèque de l'image.

1.2.3.3 Modèle d'intégration du bougé sur le capteur

C'est le modèle de l'appareil photo qui est responsable de la transformation des tremblements angulaires en translation sur l'imageur. La PSF est le résultat de cette intégration du bougé au cours du temps. Enfin, la PSF est utilisée pour flouter les images à partir d'une scène idéale de référence et ainsi simuler la dégradation de l'image au cours de son intégration sur l'imageur. Le schéma synthétique du modèle de simulation de l'appareil photo est présenté figure 22.

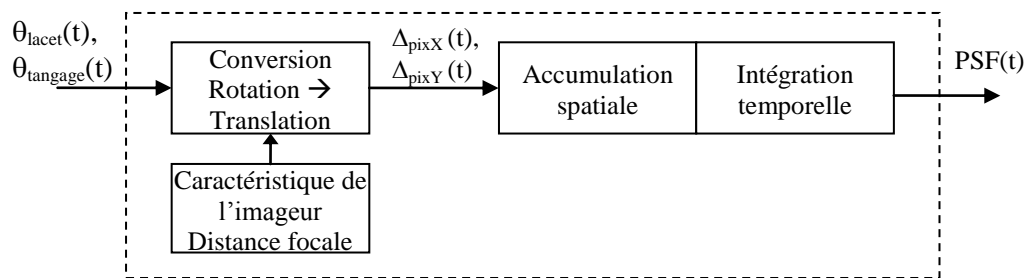


Figure 22. Modèle de l'appareil photo utilisé dans notre simulation.

Comme montrée sur cette figure 22, une paire de signaux aléatoires de rotation angulaire, provenant du modèle de tremblement de la main, est convertie en bougé de translation sur l'imageur. Cette conversion est calculée grâce à l'équation 6. Elle est paramétrée suivant les caractéristiques de l'imageur, telles que la taille des pixels et la distance focale. À cet instant, les signaux de translation sont encore dans le domaine continu. Ils doivent être échantillonnés pour rendre compte de la pixellisation de l'imageur. Un accumulateur spatial est responsable de cette opération permettant de convertir un flux de pixels entier en rapport de pixels suivant la surface de recouvrement impacté par le tremblement. Ce rapport d'illumination est alloué aux quatre pixels voisins recouverts par le déplacement sous pixel comme l'illustre la figure 23.

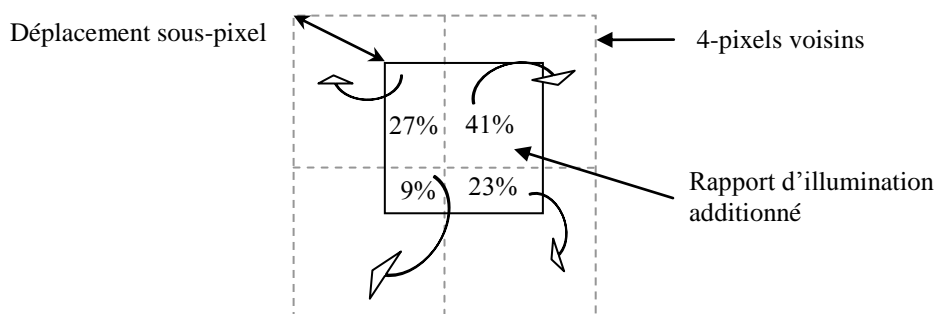


Figure 23. Exemple de notre modèle d'accumulateur spatial pour un déplacement sous pixel. Ligne pleine : déplacement sous pixel, ligne pointillée : position actuelle de la matrice de pixel.

Finalement, la PSF est obtenue par l'intégration temporelle de ces rapports d'illumination. Une image floutée par le bougé est ensuite obtenue par une simple convolution 2D à partir de l'image idéale de référence et de la PSF.

La convolution de celle-ci par une scène idéale donne une image floutée naturelle en cours d'intégration comme présentée figure 24. L'image de sortie finale simule au niveau pixel les valeurs des tensions des photodiodes au cours d'une prise de vue.

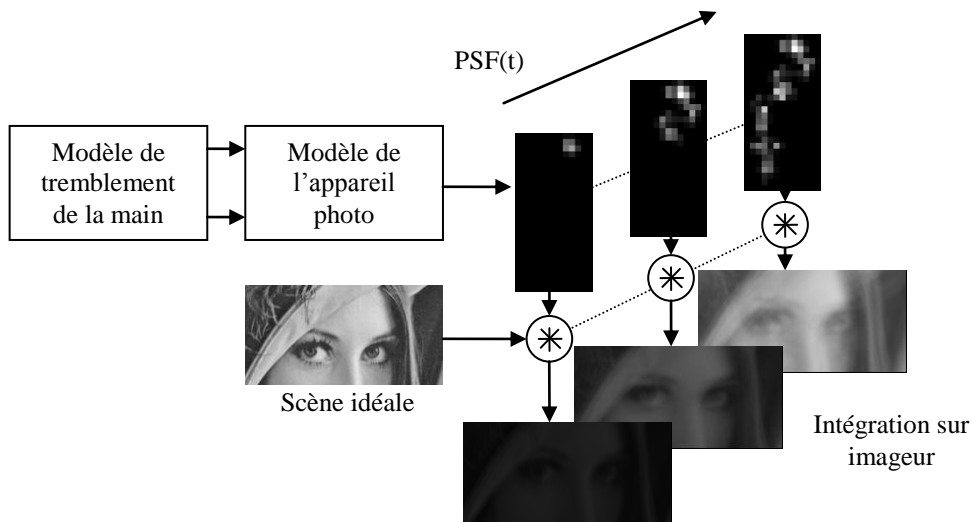


Figure 24. Modèle de simulation de flou de bougé pendant une prise de vue.

La PSF temporelle résultante de la sortie de notre modèle permet de décrire mathématiquement le flou de bougé qu'aurait subi une source ponctuelle. La PSF est ainsi la carte d'identité du flou, la réponse du système à la perturbation du bougé.

Le modèle de simulation du tremblement humain présente différents types de sorties. Elles permettent de simuler différents niveaux dans un système final. Un exemple de ces sorties est présenté figure 25.

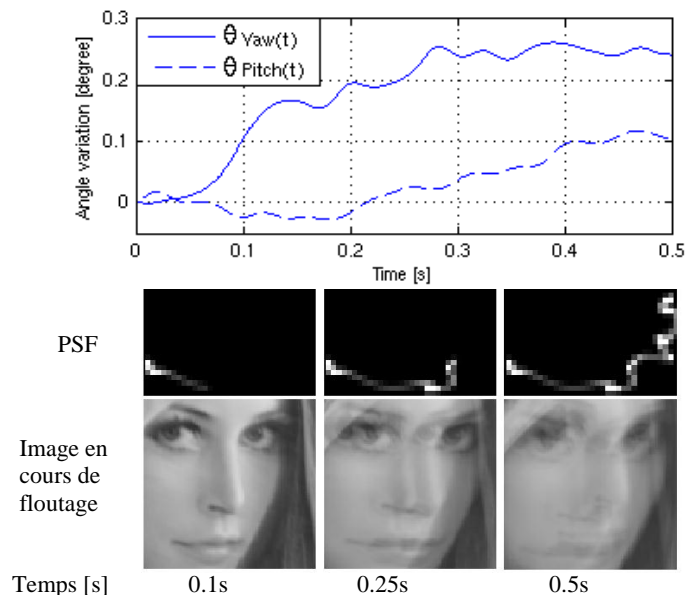


Figure 25. Exemple de sortie du modèle de génération de flou de bougé. Bougé de rotation angulaire (haut), Matrices de PSF (milieu), Trois images résultantes à 100% de l'éclairement (bas).

Grâce à la génération de la PSF avec notre modèle de bougé, nous pouvons simuler des flous naturels provenant des tremblements physiologiques de la main prenant en compte les caractéristiques de l'appareil photo. Pour valider la pertinence de ce nouveau modèle proposé, dans la section suivante, nous allons le comparer aux modèles existant ainsi qu'aux observations faites dans la littérature.

1.3 Comparaison et validation

1.3.1 Comparaison subjective aux autres modèles

Dans un premier temps, nous avons comparé notre modèle de façon subjective. D'une part avec les autres modèles existants que sont le modèle de bougé linéaire et le modèle de marche aléatoire. D'autres par avec les formes des sorties enregistrées des gyroscopes pendant une prise de vue. Cela nous permet une première validation subjective de la pertinence de notre nouveau modèle proposé.

1.3.1.1 Formes de bougés

Dans cette partie, nous comparons notre modèle à des modèles plus simples qui sont généralement utilisés pour simuler le bougé d'un appareil photo. Comme nous l'avons vu dans l'état de l'art de ce chapitre, le tremblement de bougé se comporte comme quelque chose d'aléatoire mais qui reste entre un mouvement proche du mouvement linéaire pour un temps d'exposition court et d'un mouvement aléatoire pour un temps d'exposition long.

Pour comparer ces modèles de manière subjective de la meilleure façon, nous avons choisi de fixer l'écart-type des trois modèles. Ce calcul d'écart-type est réalisé suivant l'axe long reprenant la démarche de calcul de [10]. Cela nous permet d'avoir une juste référence de comparaison malgré le fait que ces modèles présentent des formes de sorties très différentes.

La figure 26 s'intéresse à comparer les différents modèles de simulation de flou pour le cas d'un temps d'exposition court.

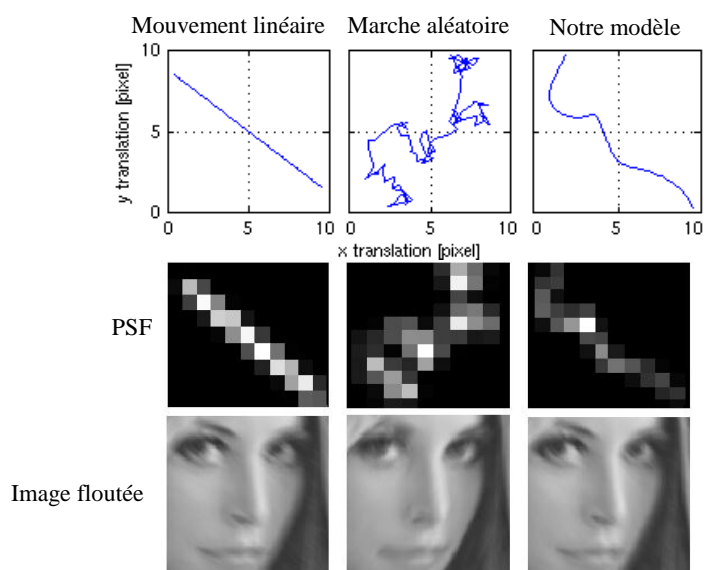


Figure 26. Comparaison des différents modèles pendant un temps d'exposition court (0,1 s).

Si l'on regarde la figure 26, nous pouvons remarquer que le modèle de marche aléatoire montre une très forte activité de déplacement sous pixel et n'arrive pas à reproduire un déplacement proche du type flou linéaire qui est largement observé pour des temps d'expositions court. A l'inverse, notre modèle reproduit ce type de bougé linéaire. Si l'on regarde les images floutées finales, nous pouvons remarquer que le modèle de flou linéaire et notre modèle reproduisent un flou observé pour des temps d'expositions faibles. Alors que le flou final du modèle de marche aléatoire génère un flou trop étalé pour le temps de pose.

Pour des temps d'exposition longs, un exemple de résultat est présenté figure 27.

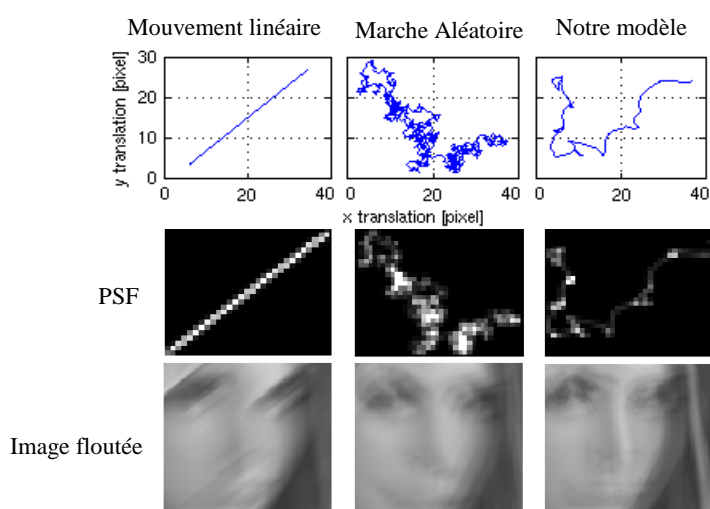


Figure 27. Comparaison des différents modèles pour un temps d'exposition long (0,8 s).

Les résultats montrent clairement que pour un long temps d'intégration, le modèle simple de bougé linéaire produit un flou non naturel, trop directionnel. La forte dispersion de la marche aléatoire génère une trop forte dispersion du flou de bougé. L'image floutée semble beaucoup trop agitée dans toutes les directions pour un flou de bougé naturel. Alors que dans ce cas, notre modèle arrive à reproduire un flou naturel.

Seul notre modèle arrive à reproduire de manière subjective des flous pour tout type de temps d'exposition.

1.3.1.2 Comparaison aux sorties gyroscopes

Une autre manière subjective est de comparer les sorties des gyroscopes d'appareils photo aux sorties de notre modèle du tremblement angulaire de la main pendant un temps de pose.

Dans le brevet de [52], un relevé de sortie de gyroscope d'un appareil photo du type reflex pendant un temps de pose de 1 seconde est présenté figure 28 à gauche. Nous le comparons avec les sorties de notre modèle à droite pour une variation angulaire aléatoire.

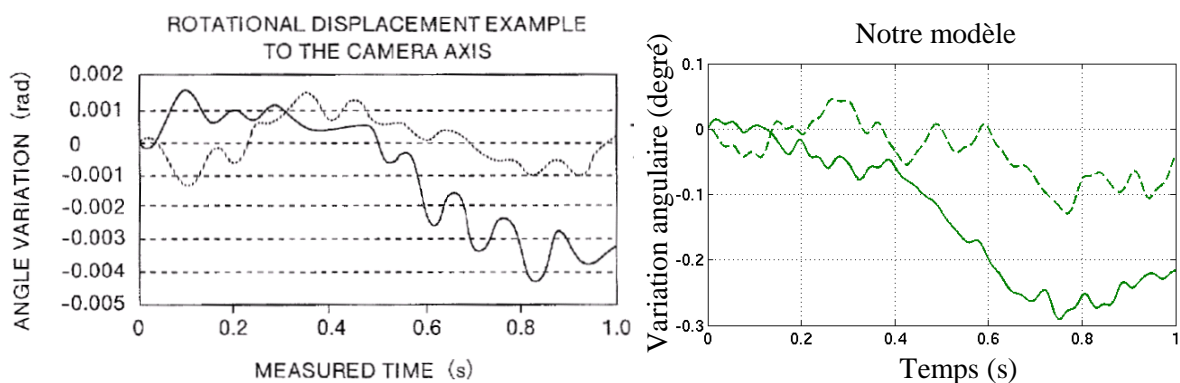


Figure 28. Comparaison de sorties gyroscopes [52] (gauche), avec notre modèle (droite)

Les résultats de simulation de notre modèle montrent une forte ressemblance avec le relevé des mesures gyroscopiques faites par [52]. Pendant un long temps d'intégration, notre modèle arrive à reproduire les formes des variations angulaires du bougé.

Une autre comparaison subjective est réalisée figure 29 pour des relevés avec des temps d'exposition court. Si l'on compare les sorties gyroscopes 2D de [40] faites sur un temps d'exposition de 100 ms avec un appareil photo de masse légère, nous pouvons observer une

forte similarité entre les mesures gyroscopiques et les sorties 2D du modèle de simulation. Ces mouvements peuvent effectivement s'apparenter à des mouvements quasi linéaires.

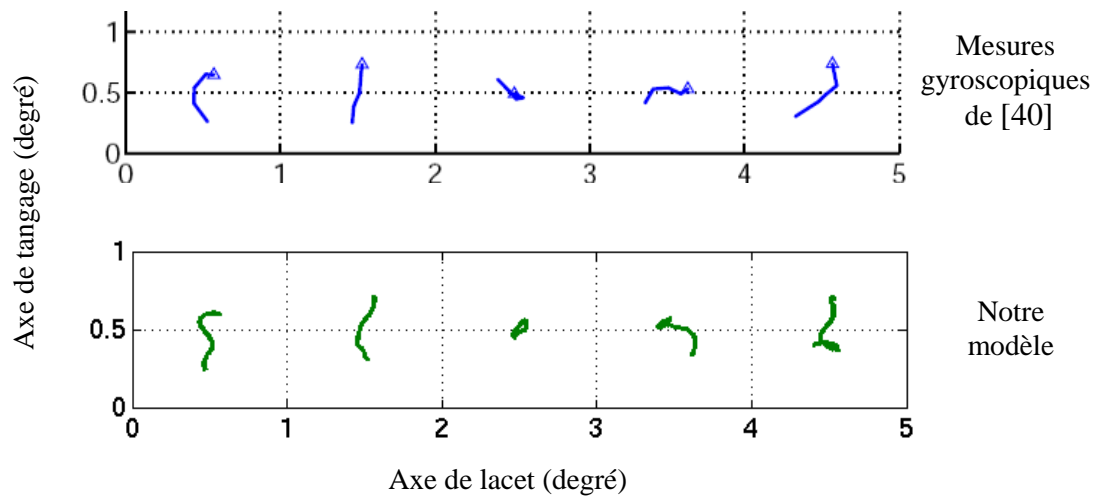


Figure 29. Sorties gyroscopiques 2D de [40] comparées à notre modèle de tremblement.

Ces résultats subjectifs montrent que le modèle proposé pour simuler le flou de bougé semble reproduire d'une manière assez fidèle les formes des sorties des gyroscopes d'appareils photo. Même si ces résultats sont bons, ils ont besoin d'être validés par des mesures plus objectives. Ces mesures objectives sont proposées dans la section suivante.

1.3.2 Validation statistique

Une étude statistique a été conduite pour comparer notre modèle de tremblement physiologique aux résultats statistiques de l'étude terrain de [10]. Premièrement, nous nous sommes remis dans les conditions de l'expérience menée par Silverstein en matière de distance focale et taille de pixel. Par simulation Matlab, nous avons reproduit son protocole d'expérience mais avec notre modèle de tremblement physiologique. Pour cela, nous avons simulé 100 PSF de flous de bougés différents, puis nous en avons calculé leur écart-type suivant les grands et petits axes au cours de leur temps d'intégration. Puis à chaque temps d'intégration, nous avons moyenné tous les résultats provenant des 100 grands axes des PSF pour construire l'évolution de l'écart-type suivant le temps de pose.

Nous avons aussi réalisé ce calcul pour les modèles de bougé linéaire et de marche aléatoire afin de les comparer. Les résultats sont présentés figure 30.

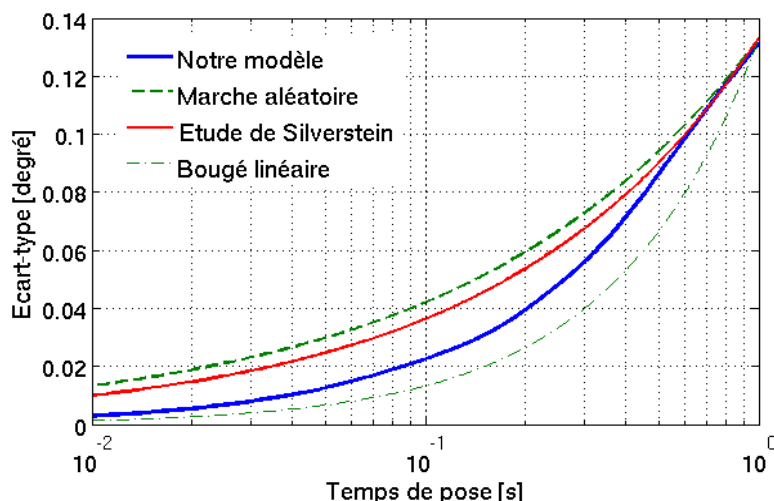


Figure 30. Comparaison de l'évolution des écarts-types suivant le temps de pose entre la courbe d'équation d'ajustement de [10], le bougé linéaire, la marche aléatoire et notre modèle.

Nous pouvons remarquer dans la figure 30 que l'équation d'ajustement de [10] ainsi que notre métrique est bornée par l'écart-type du flou de bougé linéaire et de marche aléatoire. Pour des temps d'exposition court, notre modèle est plus proche de la statistique du flou linéaire que l'équation d'ajustement de [10]. Cela peut s'expliquer par le fait que le coefficient d'ajustement b vaut 0,5632 qui est très proche de la marche aléatoire de valeur 0,5. Donc pour des temps d'exposition court, notre modèle montre un plus faible écart-type que les résultats de [10]. Mais si l'on regarde plus en détail leurs points de mesures, on peut voir que l'ajustement de la courbe tend à sur estimer l'écart-type dans ces temps courts. De plus, du fait de la relative imprécision de la mesure due au diamètre de la tâche de la LED, il semble difficile de mesurer avec précision l'écart-type pour des temps d'exposition très courts.

Inversement pour des temps d'exposition de longue durée (c'est-à-dire supérieur à 400 ms), l'équation d'ajustement et notre modèle convergent vers une statistique de marche aléatoire.

Seul notre modèle reproduit correctement les observations statistiques tout au long de son temps d'intégration.

1.4 Conclusion sur notre modèle de génération de flous de bougé

1.4.1 Avantages et points nouveaux

Un modèle de génération de flous de bougé a été présenté dans ce chapitre. À notre connaissance, c'est le premier modèle de bougé qui réunit les observations faites dans les domaines statistique, temporel et fréquentiel. Nous avons démontré que notre modèle de tremblement physiologique permet de reproduire fidèlement le comportement du tremblement humain et son impact sur l'appareil photo comme synthétisé dans le tableau 2.

Tableau 2. Tableau comparatif sur la pertinence des différents modèles de bougé : Bougé linéaire, marche aléatoire et notre modèle proposé.

| | <i>Temps de pose</i> | <i>Bougé linéaire</i> | <i>Marche Aléatoire</i> | <i>Tremblement Physiologique</i> |
|--|----------------------|-----------------------|-------------------------|----------------------------------|
| Ressemblance au spectre du tremblement de la main | - | ✗ | ✗ | ✓ |
| Ressemblance subjective avec les sorties gyroscopiques | Court | ✓ | ✗ | ✓ |
| | Long | ✗ | ✗ | ✓ |
| Ressemblance au flou naturel | Court | ✓ | ✗ | ✓ |
| | Long | ✗ | ✗ | ✓ |
| Validité statistique | Court | ✓ | ✗ | ✓ |
| | Long | ✗ | ✓ | ✓ |

Ce tableau montre que notre modèle de bougé est bien plus riche et fidèle que les modèles généralement utilisés pour simuler ce type de flou. Notre modèle est valable en temps continu, quel que soit le temps de pose jusqu'à 1 seconde. Dépassé ce temps notre modèle ne peut pas prendre en compte toutes les contre-réactions volontaires du photographe face à sa propre dérive de bougé. Mais d'une certaine manière notre modèle reste valable, car la partie de tremblement physiologique de l'homme sera toujours présente quel que soit le temps de pose.

Ainsi, cette bonne connaissance et modélisation du flou de bougé nous permet d'avoir des entrées réalistes dans nos systèmes de stabilisation d'image présentés dans le chapitre 3.

1.4.2 Conclusion sur le modèle

En utilisant une approche unifiée des études physiologiques, temporelles, fréquentielles et statistiques, notre modèle nous permet de générer des images avec des flous réalistes. Par contre, une métrique de flou de bougé a besoin d'être développée pour quantifier l'impact visuel du flou de bougé sur l'image finale. Et ainsi d'une part, mesurer le gain visuel de ces systèmes de stabilisation et, d'autre part, développer des algorithmes de stabilisation prenant en compte la qualité finale visée.

Chapitre 2 *Métrique de qualité d'image pour le flou de bougé*

La prédiction de qualité d'image est un champ de recherche complexe puisqu'il s'appuie sur l'hypothèse qu'il est possible de prédire la qualité des images d'une manière objective, alors que chaque personne perçoit les images de façon subjective provenant de l'interprétation de l'image par l'homme [53]. De nombreuses études ont été faites sur la qualité des images. La complexité de la prédiction provient des nombreux paramètres qui entrent en jeu dans la perception de la qualité :

- La nature intrinsèque de l'image (scène naturelle, image virtuelle, type de scène, image personnelle...).
- La composition de l'image (cadrage, composition de la scène, personnage principal...).
- La dégradation de la chaîne de l'image (optique, capteur, condition de prise de vue, support d'enregistrement, support de visualisation...).

Dans notre étude, nous avons restreint notre champ de recherches sur la qualité des images au cas de la dégradation de la chaîne de l'image, spécifiquement dans le cas des images subissant un flou de bougé pendant une prise de vue. Ce choix a été motivé dans le but d'isoler l'impact du bougé de l'appareil photo sur la qualité de l'image. De ce fait dans ce deuxième chapitre tous mes travaux sur l'évaluation de la qualité des images sont concentrés sur la quantification du flou de bougé sur les images, quels que soient la nature et les détails de l'image.

Dans un premier temps, je présenterai un état de l'art sur la qualité des images. Nous regarderons comment les images sont perçues par l'homme, et quelles sont les différentes méthodes de mesures psycho-visuelles développées par les chercheurs. Puis nous examinerons les métriques de flou existantes. Ma conclusion sera qu'elles ne sont pas adaptées à la quantification du flou de bougé.

Dans un second temps, je détaillerai comment est construite la métrique de qualité que je propose pour quantifier le flou de bouger.

Dans un troisième temps, nous verrons les expériences visuelles que j'ai menées pour valider la métrique de qualité. Une première expérience a été mise en place pour valider la métrique dans le cas particulier du flou de bougé linéaire, puis une deuxième expérience pour étendre cette validation aux flous quelconques. Ces flous sont générés grâce à notre modèle de tremblement physiologique. Puis nous verrons l'application de la métrique au flou de bougé lors d'une prise de vue photographique.

Enfin, nous concluons sur la pertinence de notre métrique pour quantifier le flou de bougé.

2.1 État de l'art sur la qualité d'image

2.1.1 Mesure liée à la perception visuelle

2.1.1.1 L'homme comme seul juge

L'objectif d'une image est de reproduire une scène par l'intermédiaire d'un support pour transmettre une information visuelle à l'homme. Ainsi, la mesure de la qualité d'image revient à mesurer la satisfaction de l'homme lorsqu'il regarde le rendu de cette image. Cette satisfaction de l'homme peut varier suivant la difficulté de reproduire la scène. Par exemple des archives de vieilles images, des images de l'espace, des images prises sur le vif ou insolites... Dans tous les cas c'est l'homme qui sera juge de la qualité finale de l'image. Pour les dégradations de type floutage, l'action de rendre une image flou, le système mis en jeu pour juger de la qualité de l'image repose sur 3 composantes :

- L'homme : avec sa vision, ses intérêts pour l'image, ses attentes de l'image, son interprétation de l'image...
- L'image : le type de scène, le support de visualisation, les dégradations (netteté dans notre cas)...
- La distance de visualisation, la distance entre l'homme et l'image.

Une image nette, est définie comme une image sans flou, qui a un piqué important caractérisé par des contours francs avec une forte variation de la luminance dans les zones

texturées de celle-ci. Si l'on prend exactement la même scène mais avec plusieurs nettetés différentes, l'image jugée la meilleure sera celle qui aura le piqué le plus fin, celle qui pourra rendre les plus petits détails de l'image. Inversement une image floutée sera celle perçue comme ayant des contours lissés ainsi qu'ayant subi une perte de détails. Donc, la mesure de flou va tenter de caractériser cette dégradation de perte de netteté.

Dans une situation donnée (ou l'homme, l'image et la distance sont fixés), il existe un seuil de perception qui permet de distinguer le « flou » du « net ». Cela arrive lorsque le plus petit point de l'image discernable par l'œil est parfaitement compris dans un cercle, on l'appelle le cercle de confusion (CdC) qui a été défini par [54].

Le cercle de confusion

Le CdC est défini comme le point équivalent le plus petit sur une image discernable par l'œil humain. Le rayon du CdC dépend de :

- L'acuité visuelle : la capacité de résolution spatiale du système visuel, qui permet à l'œil de voir les détails les plus fins. Elle est limitée par la diffraction, les aberrations et le nombre de photorécepteurs contenus dans l'œil [55]. Cette acuité visuelle pour les personnes dites normales, avec une vision de 10/10, est de 1/60 degré [56], 1 arc minute. Pour des observateurs entraînés avec une excellente vision (20/10) sur un fond très lumineux, l'acuité maximale visible peut être jusqu'à 0,5/60 degré [57]. Mais en général, 1 arc minute est retenu pour les différents calculs mettant en jeu l'acuité visuelle humaine.
- La distance de visualisation : la distance entre l'observateur et le support de l'image, qui joue un rôle d'agrandissement ou de rétrécissement lorsque cette distance est réduite ou augmentée respectivement. Le fait de faire varier la distance d'observation va faire varier la taille du CdC.
- Du support de visualisation : le support de visualisation prend en compte le fait que par rapport à l'image originale, il peut y avoir un agrandissement ou un rétrécissement. Si l'agrandissement est trop important, c'est ce facteur qui limitera la taille du CdC et inversement pour un rétrécissement. Par exemple pour une image avec une très forte résolution, lorsqu'elle sera affichée à l'écran, celle-ci sera rétrécie. C'est la résolution de l'écran qui limitera la taille du CdC. Le contraste du support de visualisation rentre aussi en jeu, plus il sera contrasté, plus il sera facile de discerner un point de petite taille.

Ces différents paramètres sont importants à renseigner lorsque l'on parle de qualité d'image. Ainsi, les différentes expériences psycho-visuelles qui tentent de mesurer le jugement de la perception humaine doivent prendre en compte ces données de distances, de supports et d'acuité visuelle.

Sensibilité au flou (Luminance vs chrominance)

La sensibilité au flou vis-à-vis de la couleur ou de la luminance (c'est-à-dire des images en niveau de gris) a été étudiée par [58, 59] en s'intéressant à la sensibilité de notre système visuel aux contrastes spatiaux. En effet, un flou sur l'image va avoir tendance à réduire les hautes fréquences spatiales de celle-ci. Les images pouvant se décomposer suivant les informations de luminance et de chrominance, nous pouvons nous poser la question : est-ce la luminance ou la chrominance qui porte les informations de netteté d'une image ? En d'autres termes, d'un point de vue de la netteté de l'image : est-ce qu'une image en niveau de gris va être perçue de la même manière que si elle était en couleur ?

Si l'on regarde la réponse en luminance, présentée figure 31.a proposée par [58], nous pouvons observer que l'œil est sensible aux fréquences spatiales intermédiaires. Alors que si l'on regarde la réponse de la chrominance, présentés 31.b proposé par [59], nous observons que notre œil est sensible aux basses fréquences spatiales. Ainsi nous pouvons en déduire que lorsqu'il y aura un flou sur l'image, c'est au niveau de la luminance que la dégradation va être gênante. En effet, le flou va impacter les hautes fréquences spatiales ainsi que les fréquences intermédiaires.

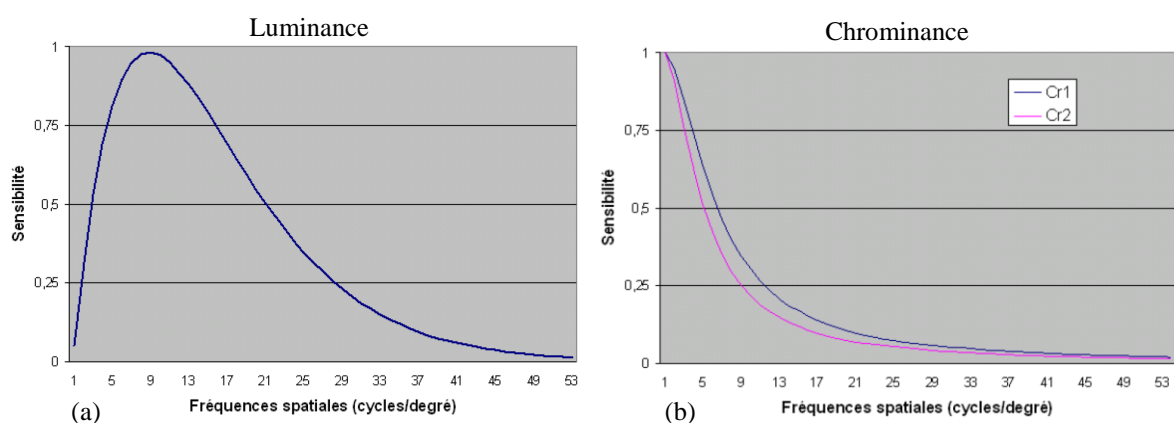


Figure 31. Réponses normalisées de la perception visuelle vis-à-vis des contrastes spatiaux. (a) Réponse de la luminance. (b) Réponse des composantes couleurs (notées C_{r1} et C_{r2}).

Cette propriété a aussi été observée et validée par expérience visuelle dans [9]. Cette expérience a consisté à confronter les résultats de perception d'image floutée en couleur avec celle en niveau de gris. Les résultats montrent une très forte similarité entre les deux types d'image. De manière analogue, aucune différence n'a été observée au niveau de l'attention visuelle des images que ce soit avec les stimuli en couleur ou en niveau de gris [60].

Comme nous pouvons le constater sur la figure 32, la perception du flou est identique pour des images en couleur ou en niveau de gris. Ainsi pour simplifier mes travaux sur le développement d'une métrique de qualité de netteté, nous nous sommes limités à la perception des images en niveau de gris.



Figure 32. La perception du flou est similaire pour les images en couleurs (a) ou en niveaux de gris (b).

2.1.1.2 Les mesures psycho-visuelles développées

Le développement d'une métrique de qualité d'image demande au préalable des méthodes de mesure de qualité d'image par l'homme. Pour cela, deux principaux critères ont été développés : le *mean opinion score* (MOS), ou note d'opinion moyenne et le *just noticeable difference* (JND), ou de différence juste perceptible en français.

Expérience de « mean opinion score »

Les mesures psycho-visuelles de MOS sont les plus répandues pour évaluer la qualité des images. Ces méthodes d'évaluation ont été formalisées dans [61, 62].

La méthode MOS consiste à présenter des images appartenant à un ensemble de jeux de stimuli. Ceux-ci comportent les images ayant été modifiées par différentes dégradations à partir d'images de référence. Les images sont présentées une à une sans la présence de l'image de référence. Puis, il est demandé à l'observateur d'évaluer l'image en donnant une note à l'image présentée. La note peut être de type numérique ou textuelle tel que : excellente, bonne, moyenne, mauvaise... Puis pour chaque dégradation, une moyenne est réalisée entre les différents observateurs. Ainsi est obtenue la qualité de l'image suivant la dégradation de celle-ci. De par la simplicité de la méthode, un grand nombre de bases de données d'images ont été développées. Ces bases de données comprennent un grand nombre d'images ayant des dégradations différentes avec leurs notes MOS associées. Les plus connues sont accessibles en libre téléchargement : LIVE [63], TID [64], CSIQ [65], IVC [66], A57 [67], Toyama [68], WIQ [69].

Pour l'observateur, la tâche est assez directe et rapide, mais peut poser quelques problèmes de jugement du fait de n'avoir aucune image de référence. Par contre, grâce à cette rapidité, il est possible de noter un grand nombre d'images par un grand nombre d'observateurs. De la même manière que [62], les résultats sont obtenus par le calcul de la moyenne des notes données par les observateurs pour chaque image. Ces scores reflètent une tendance relative de préférence, mais ne donnent pas de réponse absolue sur la valeur de la qualité des images. De plus la méthode MOS, par l'absence d'image de référence présente 3 problèmes :

- Il est difficile de comparer les différentes bases de données MOS réalisées par les chercheurs [70].
- L'observateur va avoir tendance à répartir sa notation d'image sur toute la gamme de notation qui lui est proposée quel que soit le jeu d'images présentées [71]. Ainsi, la même image avec la même dégradation notée dans deux jeux différents n'aura pas la même note.
- L'observateur va avoir tendance à noter l'image suivant l'image précédente [71]. Ainsi, la même image avec la même dégradation affichée à deux moments différents n'aura pas la même note.

Pour ces raisons d'autres méthodes d'évaluation de la qualité ont été développées pour introduire une référence dans le système de notation.

Expérience de « just noticeable difference »

Tel que présenté par [9], JND est une manière naturelle pour quantifier la qualité des images pour des observateurs. Le protocole d'expérience est défini par [9, 72]. Il est basé sur un vote forcé de comparaison entre deux images, pour choisir l'image qui semble de meilleure qualité. À partir du pourcentage de bonnes réponses pour un couple d'image donné, une note en JND est calculée. Le standard propose un intervalle d'un JND entre deux images comme étant le résultat de 75 % de bonnes réponses. Ce choix correspond exactement à la séparation de deux ensembles de 50 %, ceux qui répondent en sachant qu'ils ont raison (50 % de bonnes réponses) et ceux qui répondent en devinant (50 % des autres réponses). En effet, dans les 50 % de réponse devinée, il y a 25 % de mauvaises réponses (les personnes n'ont pas fait exprès de choisir la mauvaise réponse) et donc par symétrie, il y a 25 % de bonnes réponses devinée. Pour récapituler, il y a donc 75 % de bonnes réponses dont 25 % des réponses ont été devinées (les autres 50 % ont été consciemment choisies) et 25 % de mauvaises réponses. Ce choix proposé par le standard est aussi appelé 50 % JND.

Le calcul des réponses en JND est basé sur l'hypothèse que la perception visuelle ressemble à une fonction de répartition, mais avec une queue finie. Elle est définie par l'équation 9.

$$JNDs = \frac{12}{\pi} \sin^{-1}(\sqrt{p}) - 3 \quad (9)$$

où p est le pourcentage de bonnes réponses pour un résultat entre un couple de comparaison de deux images et $JNDs$ le résultat de cette comparaison. La courbe de l'équation 9 est présentée figure 33. En pratique, le résultat de bonne réponse est compris entre 50 et 100 %. Lorsque l'observateur n'a aucun problème pour différencier les images, il choisira toujours la bonne réponse et obtiendra 100 % de bonnes réponses. Dans le cas inverse, un observateur n'arrivant pas à différencier les images choisira de manière indéterminée celle qui lui semble de meilleure qualité. En conséquence statistiquement après un grand nombre de votes il aura choisi autant de bonnes que de mauvaises réponses soit un score de 50 % de bonnes réponses. C'est pour cela que l'abscisse est présentée figure 33 entre 50 et 100 %.

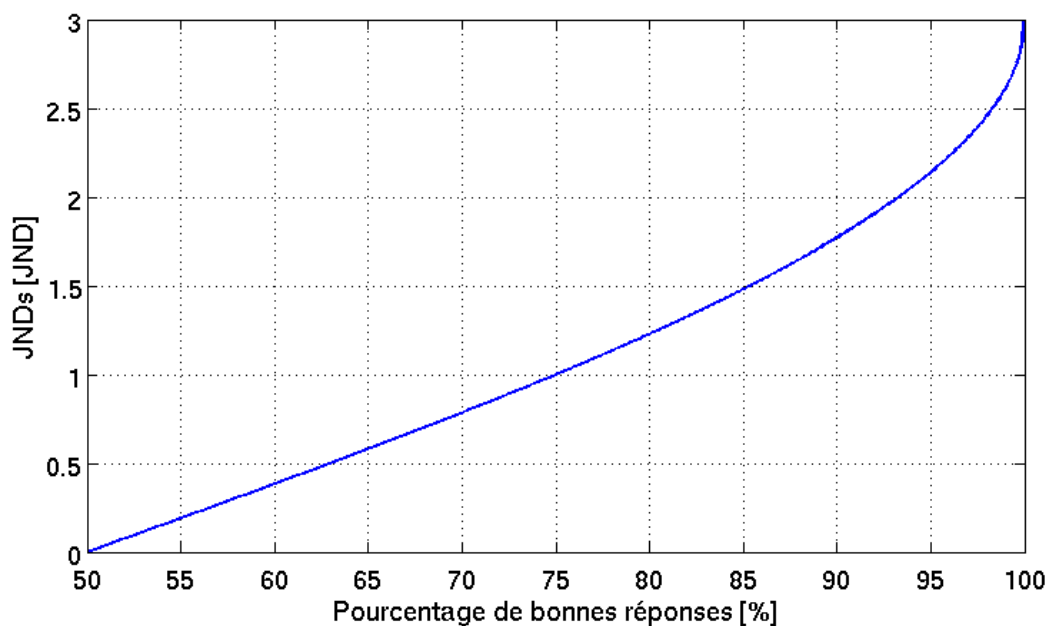


Figure 33. Courbe de l'équation 9 présentant les résultats en JND suivant le pourcentage de bonnes réponses.

L'intégration de plusieurs JNDs tout au long d'une large gamme de dégradation permet d'obtenir l'évolution de la qualité d'image au long de cette dégradation. Une nuance est utilisée entre JNDs et JND. JND fait référence à la mesure de détectabilité de l'apparition d'une différence entre les deux images au niveau de la qualité globale de l'image. Alors que JNDs fait référence à la détectabilité d'un seul attribut de dégradation comme la netteté, un bruit, un contraste...

Expérience de Règle de Qualité

L'expérience JND est très adaptée pour la mesure de perception d'une faible différence de qualité, mais ne s'applique pas pour les grandes amplitudes de différence de qualité. Pour répondre à ce besoin, les expériences de règles de qualité ont été développées pour mesurer de manière absolue la qualité des images. Cette méthode a été à l'origine présentée par [9] puis standardisée comme ISO [72]. Cette norme a été révisée suite aux améliorations proposées par [73] concernant l'implémentation de la règle de qualité d'image. Dans ce type d'expériences, une base d'images calibrées répartie tout au long d'une gamme est présentée à l'observateur avec un jeu d'images à évaluer en matière de qualité. Cette base d'images calibrées provient de l'expérience de JND. L'image à évaluer est mise en vis-à-vis avec une image de référence calibrée modifiable. L'observateur doit ainsi ajuster de la meilleure façon l'image calibrée qu'il fait varier vis-à-vis de l'image à évaluer. Cela est facilité par

l'utilisation d'un ascenseur qui permet de faire varier la qualité de l'image suivant la dégradation. Lorsque la correspondance de qualité est maximale, connaissant la note de l'image de référence en JND, la même note est attribuée à l'image à évaluer. L'image de référence ajustable est en réalité un grand nombre d'images de référence espacées de manière proche en niveau de qualité, ceci permet ainsi de mesurer une large gamme de qualité d'image. Le faible espacement entre les images de référence permet à l'observateur d'évaluer l'image avec un fort taux de confiance. Du fait de la monotonie de la règle de référence, le risque d'effet de gamme ainsi que d'inversion, comme discuté précédemment, est très fortement réduit. En effet, lorsque l'on note un nombre d'images important avec des qualités différentes, la règle de qualité permet de garantir des résultats étalés sur la totalité de la gamme de qualité.

Grâce à l'introduction d'une référence absolue, cette méthode analogue à la règle classique de mesure linéaire permet ainsi de pallier des problèmes des méthodes MOS mentionnés par [71].

2.1.1.3 Conclusion et limitations des mesures psycho-visuelles

Les deux méthodes MOS et JND présentées précédemment sont comparées dans le tableau 3. Les avantages et les inconvénients de ces méthodes en sont ainsi récapitulés.

Tableau 3. Avantages et inconvénients des deux différentes méthodes des mesures psycho-visuelles.

| <i>Méthode</i> | <i>MOS</i> | <i>JND et Règle de Qualité</i> |
|----------------|--|---|
| Avantages | Facile à implémenter Bases de résultats disponibles | Précision de résultats Pas d'effet de gamme ni d'inversion. Mesure absolue (avec référence) |
| Inconvénients | Donne seulement une tendance Problème d'effet de gamme et d'inversion Mesure relative (sans référence) | Difficile à mettre en place Base de résultat peu disponible |

La méthode que j'ai choisie est celle basée sur les JND, car par l'introduction d'une référence dans la mesure, les problèmes d'effets de gamme et d'inversion en sont supprimés. Ainsi, cette méthode permet d'obtenir des résultats précis sur un type de dégradation d'image, celui du flou et plus particulièrement le flou de bougé. De plus cette méthode me permet par

la suite de valider avec précision la métrique de qualité d'image développée pour le flou de bougé.

Du fait de la complexité du système visuel humain ainsi que de ces interprétations, ces méthodes de mesure de qualité ne sont malheureusement pas directement implémentables. Puisque ni les algorithmes et ni les modèles peuvent être réalisés, les chercheurs ont développé des métriques spécialisées basées sur des paramètres objectifs en les confrontant ou non aux résultats d'expériences psycho-visuelles.

2.1.2 Mesures objectives

2.1.2.1 Positionnement du problème

Les mesures objectives ont été développées pour mesurer la qualité des images à partir de grandeurs mesurables, sans passer par l'œil humain. Néanmoins, ces mesures objectives ont besoin d'être confrontées à la perception humaine, car elles permettent de valider ou non leur pertinence. L'avantage de ce type de mesure est de pouvoir caractériser les résultats du ressenti humain sans avoir besoin de le modéliser directement. Ces types de méthode sont normalisés sous forme mathématique ou algorithmique que l'on appelle des métriques. Ces métriques suivant leur nature utilisent différentes entrées pour calculer la valeur du flou. Elles peuvent être rassemblées sous trois types différents :

- Les métriques dites sans référence (SR) (*No-reference*) se basent seulement sur l'image finale. Puisqu'elle utilise l'image finale, ce type de métrique n'est pas prédictible.
- Les métriques dites avec référence (AR) (*Full-reference*) se basent non seulement sur l'image finale, mais aussi sur l'image originale n'ayant pas subi de dégradation. De même ce type n'est pas prédictible.
- Les métriques dites référence réduite (RR) (*Reduced-reference*) ne se basent ni sur l'image finale ni sur l'image originale, mais sur d'autres sources d'informations externes de l'image. Elles sont du type : taux de compression, linéarité du capteur ou de l'optique, sources de bruit extérieur... Ce type de métrique indépendant de l'image est donc d'une certaine manière prédictible.

Un état de l'art sur les métriques de qualité d'image s'intéressant de près ou de loin aux problématiques de quantification du flou est présenté ci-dessous. Toutes ces métriques ont été

développées pour répondre à des besoins particuliers que nous vous présentons une à une, regroupées par leur origine de développement.

2.1.2.2 Métriques orientés autofocus

Pour différents types d'appareils (appareil photo, microscope...) utilisant des optiques, une mise au point est nécessaire afin d'avoir une image finale parfaitement nette. Des métriques d'autofocus ont été développées pour rendre automatique cette mise au point. Ces métriques s'intéressent à la quantification du flou en cherchant le maximum de netteté pour renseigner le système de la mise au point ou non. Les métriques présentées qui suivent ont été réalisées dans ce but précis.

- Métrique basée sur la variance [74] : cette métrique calcule la variance sur toute l'image. Lorsque le flou augmente sur l'image, les bords de celle-ci deviennent de plus en plus lisses, ce qui implique que la transition entre les niveaux de gris de l'image diminue. De ce fait, la variance diminue aussi. Plus l'image sera nette, plus la métrique aura une forte valeur.
- Métrique basée sur un seuil de fréquence [75] : cette métrique calcule la somme des amplitudes de toutes les fréquences au-dessus d'un certain seuil. Ce seuil est déterminé de façon expérimentale. En augmentant le niveau du seuil, plus de bords vont être pris en compte dans le calcul de la métrique, mais en même temps, la métrique sera plus sensible aux bruits. Plus l'image sera nette, plus la valeur de la métrique sera grande.
- Métrique basée sur histogramme avec seuil [75] : cette métrique se base sur le poids de la somme des raies d'un histogramme suivant un certain seuil. Les poids de l'histogramme sont considérés comme les niveaux de gris de l'image. Le niveau du seuil est généralement choisi comme celui qui est proche de la moyenne du niveau de gris de l'image. Une image nette comprendra un plus grand nombre de hauts niveaux de gris, donc l'histogramme aura un nombre de raies plus important. Ainsi plus l'image sera nette, plus la métrique sera forte.
- Métrique basée sur l'autocorrélation [76] : dérivée de la fonction d'autocorrélation, cette métrique utilise la différence entre valeurs d'autocorrélation à deux distances différentes suivant les directions horizontales et verticales. Si l'image est floutée, donc que les bords sont lissés, la corrélation entre pixels voisins sera forte. Par

conséquent l'autocorrélation augmentera, donc la somme des différences diminuera. Plus l'image sera nette, plus la métrique sera forte.

- Métriques basées sur la dérivée [76] : ces métriques utilisent deux dérivateurs, un au premier ordre (Gradient) et un autre au second ordre (Laplacien). Ces métriques marchent comme des filtres passe-haut dans le domaine fréquentiel. En conséquence plus l'image sera nette, plus l'image sera composée de hautes fréquences et donc plus la métrique aura une forte valeur.
- Métrique basée sur l'entropie [77] : l'entropie est la quantité d'informations contenues dans une image. Si la probabilité de chaque niveau de gris est faible, l'entropie est forte et inversement. Les probabilités sont calculées en normalisant l'histogramme obtenu de l'image. Les images nettes auront un grand nombre de niveau de gris, donc chaque niveau de gris aura une probabilité d'apparition faible. Alors l'entropie sera grande.

Dans toutes ces métriques orientées autofocus, les seuls tests visuels réalisés pour valider leurs métriques ont été faits sur le maximum de netteté et non sur la variation de la métrique en fonction de la perception. Cela est dû au fait qu'ils s'intéressent plutôt à la recherche de la position de mise au point optimale, avec la plus grande rapidité et la meilleure précision. Comme ces métriques utilisent seulement l'image finale pour leurs calculs, elles appartiennent aux métriques sans références, elles sont donc par construction non prédictives.

2.1.2.3 Métriques orientées compression d'image

Lorsque des images sont compressées, des détails de l'image sont lissés pour augmenter le taux de compression, ceci entraîne un flou sur l'image. Ce flou est plus ou moins important suivant le taux de compression. Ces métriques développées pour quantifier le flou de compression ont été dérivées pour être utilisées comme métriques généralistes de quantification de flou des images. Elles sont présentées ci-dessous.

- Métrique basée sur l'histogramme des coefficients de la *discret cosinus transform* (DCT) [78] : cette métrique se base sur l'histogramme d'apparition des coefficients de DCT pour toutes les macros-bloc de 8×8 pixels de l'image (un poids plus important est donné aux coefficients de DCT dans la diagonale). La métrique de flou est calculée par le comptage des coefficients de DCT proche de zéro. Plus l'image est nette, moins il y a de coefficient de DCT proche de zéro, plus la métrique est petite.

- Métrique basée sur l'estimation de l'épaisseur des bords [79, 80] : cette métrique utilise en premier lieu un détecteur de bord. Ensuite, chaque colonne est scannée pour localiser les bords. Puis les épaisseurs des bords sont calculées, et moyennées entre elles. Plus l'image sera nette, plus les épaisseurs de ces bords seront petites, plus la métrique sera petite.
- Métrique Kurtosis [81] : le Kurtosis est une mesure statistique d'aplatissement ou de pointicité d'une distribution (moment d'ordre 4). Plus la distribution est étroite, par exemple en forme de pique, plus le Kurtosis sera fort et inversement. Celui-ci est utilisé dans la métrique pour mesurer la netteté de l'image. Plus l'image sera nette, plus le Kurtosis sera faible, plus la métrique sera petite.
- Métrique basée sur les contrastes locaux [82] : cette métrique se base sur les contrastes locaux des images. Plus l'image sera nette, plus elle aura de forts contrastes locaux. Une mesure isotropique de contrastes locaux est utilisée dans cette métrique grâce à une combinaison de filtres orientés. Plus l'image sera nette, plus la métrique sera grande.
- Métrique basée sur la taille des bords orientés [83], [84] : Cette métrique utilise un détecteur de bord (Canny) pour détecter les bords. Pour chaque bord pixélisé l'extremum local (c'est-à-dire le minimum et le maximum de la taille du bord), suivant la direction que donne le gradient, est estimé pour connaître la taille du bord qui s'étale des deux côtés du front. La moyenne des différentes tailles de bord est calculée puis injectée dans une équation exponentielle pour déterminer le résultat de la métrique. Plus l'image sera nette, plus les bords seront petits, plus la métrique sera petite.

Dans toutes ces métriques orientées compression d'image, les auteurs ont fait au mieux des tests visuels de type MOS, avec plus ou moins de sujets. Il est par conséquent difficile de juger de la précision de ce type de métrique. De plus ces métriques, utilisant seulement l'image finale, appartiennent aux métriques sans références. Elles sont de ce fait par construction non prédictives.

2.1.2.4 Métriques généralistes

Afin que quantifier la qualité des images de manière automatique, des métriques généralistes ont été développées. Ces métriques tentent de donner la qualité d'image ayant

subi différents types de dégradation combinés ou non tel que : de la compression, un flou, du bruit, une distorsion... Ces métriques généralistes sont présentées ci-dessous.

- Métriques basées sur le *peak signal-to-noise ratio* (PSNR) : parfois ce type de métrique est utilisé pour prédire la qualité des images et donc pour prédire le flou d'une image. Comme synthétisée dans l'étude de [85], la métrique de PSNR pour prédire la qualité des images (excepté dans le cas du bruit) ne donne pas de bons résultats. Cela a encouragé les chercheurs à développer d'autres métriques de qualité généraliste présentées ci-dessous.
- Métrique de *structural similarity* (SSIM) [86] : cette métrique se base sur l'hypothèse que les images sont structurées, du fait que les pixels sont fortement dépendants des uns des autres. La métrique va comparer par corrélation les informations de luminance, de contraste et de structure entre l'image de référence et l'image finale. Puis en calculer sa valeur de qualité finale. Plus l'image finale sera nette (proche de l'image initiale), plus la métrique sera faible.
- Métrique de *visual information fidelity* (VIF) [87] : cette métrique se base sur la perte d'information que peut subir une image lors d'une dégradation. Pour cela, elle fait appel aux transformées en ondelette pour utiliser la théorie des champs aléatoires des sous-bandes de l'image. La quantité d'information perdue par rapport à l'image de référence donne la perte de la qualité visuelle. Plus l'image est nette, moins elle a perdu de qualité, en conséquence plus la métrique a une valeur faible.
- Métrique de *visual signal-to-noise ratio* (VSNR) [88] : cette métrique utilise les transformées en ondelette pour calculer la distorsion de l'image au niveau du contraste ainsi qu'au niveau des perturbations globales de l'image. Le VSNR (un calcul homologue au SNR) est ensuite calculé avec ces deux paramètres. Plus l'image est nette, plus la valeur de la métrique est forte.
- Métrique de *singular value decomposition* (MSVD) [89] : cette métrique se base sur la décomposition en valeur singulière en deux dimensions des images. Un calcul de distance entre les valeurs singulières de l'image de référence et de l'image finale est réalisé pour déterminer la qualité des images. Plus l'image est nette, plus la valeur de la métrique est petite.
- Métrique basée sur les niveaux de coefficient de transformé en ondelette [90] : cette métrique utilise les différents niveaux de sous-bande de la transformée en ondelette d'une image. Ces coefficients d'ondelette sont pondérés suivant leurs

niveaux pour calculer la métrique. En effet, leurs différents niveaux n'impactent pas de la même façon la qualité de l'image. Plus l'image est nette, plus la métrique est petite.

Toutes ces métriques généralistes pour la qualité des images sont évaluées grâce aux bases d'images MOS ainsi leurs performances peuvent être comparées. Par contre, ces métriques utilisent l'image de référence ainsi que l'image finale donc elle appartiennent aux métriques avec référence et sont donc par construction non prédictives. De plus elles demandent une forte puissance de calcul compromettant leur intégration.

2.1.2.5 Métriques spécialisées dans le flou

Le flou étant une forte source de dégradation d'image, des métriques dédiées à la quantification du flou ont été développées par les chercheurs. Elles sont présentées ci-dessous.

- Métrique de netteté immunisée au bruit [91] : cette métrique repose sur les propriétés de régularité dite de Lipschitz, permettant de séparer le signal utile du signal de bruit par l'application d'une transformée en ondelette dynamique. Puis en s'appuyant sur la métrique de netteté de [80]. Cette métrique a été développée pour être robuste au bruit même avec un faible rapport signal sur bruit. Plus l'image est nette, plus la métrique à une valeur petite.
- Métrique basée sur le rapport de sortie de filtre [92] : cette métrique est basée sur le rapport des amplitudes de signal entre les fréquences passe-haut divisé par les fréquences passe-bande des détails locaux de l'image. Ces détails sont extraits par seuillage grâce à un filtre passe-bande. Plus l'image est nette, plus ce rapport est important, donc plus la métrique est élevée.
- Métrique basée sur le spectre de puissance [93] : cette métrique calcule le spectre de puissance pondéré normalisé de l'image grâce à la *modulation transfert function* (MTF). Plus l'image est nette, plus la métrique est grande.
- Métrique basée sur la probabilité de détection du flou [94] : dans un premier temps cette métrique détecte les bords grâce à l'opérateur de Sobel sur des macros-bloc de 64×64 pixels puis mesure la taille des bords détectés. Dans un second temps la métrique va pondérer l'impact du bord par une notion psycho visuel de *just noticeable blur* pour chaque bord. Ainsi, des bords floutés mais non visibles par l'œil vont être diminués dans le résultat global de la métrique et inversement. Plus l'image sera nette, plus la métrique sera petite.

- Métrique basée sur le rayon de la PSF estimée [95] : cette métrique se base sur quatre étapes. Une première de détection de bord, puis une de sélection de bord. Il vient ensuite une étape de dérivation de LSF pour estimer une PSF et enfin le calcul du rayon de la PSF estimée. Plus l'image sera floue, plus le rayon de la PSF augmentera, plus la métrique sera grande.
- Métrique basée sur l'estimation du flou gaussien [96] : cette métrique se base sur une détection de bord par filtrage de l'image grâce à un filtre quadratique. Ensuite regardant la forme des bords détectés, la métrique cherche à retrouver un profil gaussien pour en déduire son sigma. Ainsi plus l'image sera nette, plus le sigma sera petit, plus la métrique sera petite.

Dans toutes ces métriques dédiées à l'estimation du flou, les auteurs ont réalisé des tests visuels de type MOS avec plus ou moins de sujets. Il est donc difficile de comparer les résultats ainsi que de juger de la pertinence de précision de ce type de métrique. De plus ces métriques, utilisant seulement l'image finale, appartiennent aux métriques sans références et sont de la sorte par construction non prédictives.

2.1.2.6 Métriques orientées optiques de l'œil

Dans le domaine d'optométrie, des métriques ont été modélisées pour prédire la qualité de netteté de l'œil. Cette prédiction d'acuité visuelle est réalisée suivant différents problèmes visuels (astigmatisme, myopie...) qui contribuent à une dégradation de netteté. Ce type de métrique s'applique à une modélisation sphérique de l'œil. Elle peut par exemple être modélisée comme une PSF sphérique. Ainsi, les métriques proposées dans [97, 98] vont tenter de prédire une note d'acuité visuelle suivant la dégradation.

Ces métriques étant basées en partie sur la PSF font partie des métriques de type références réduites. Elles sont donc prédictibles. Le problème de ces métriques réside dans le fait qu'elles sont intrinsèquement limitées à des PSF particulières simples (sphériques) et donc sont limitées pour prédire des flous plus complexes. De plus l'évaluation de ces métriques par rapport à la perception a été faite dans un but d'optométrie. Elles ont été simplement validées par le floutage de caractères alphabétiques et non par des images naturelles.

2.1.2.7 Métriques orientées chaîne de l'image de l'appareil photo

Pour caractériser les objectifs des appareils photo ou le système global de la chaîne de l'image des appareils photo, des métriques de netteté ont été développées pour connaître quels sont les appareils les plus fidèles, de meilleure qualité. Ces métriques sont présentées ci-dessous.

- Métrique basée sur la *modulation transfert function* (MTF) [99] : cette métrique est très largement utilisée dans le domaine de caractérisation des objectifs d'appareils photo, car elle permet de connaître le pouvoir de résolution de l'objectif ou de la chaîne globale de l'image de l'appareil photo. Plus le résultat de cette métrique sera élevé, plus l'appareil sera performant. Cette métrique est efficace pour les PSF simples (type gaussien, disque) mais n'est pas assez précise pour quantifier les flous de bougé. De plus seul des évaluations de type MOS ont été réalisées par les chercheurs pour confronter les résultats de celle-ci à la perception visuelle.
- Métrique BxU [100] : cette métrique propriétaire de DxO Labs se base sur les informations de la PSF. Elle est proportionnelle à la variance de la LSF. Ainsi plus l'image est nette, plus la métrique a une valeur petite. Elle a été évaluée par quelques expériences de type MOS et est souvent utilisée dans le monde de la photo pour caractériser les objectifs d'appareils photo en matière de rendu de netteté. Par contre, cette métrique n'est pas accessible librement car elle est protégée par un copyright.
- Métrique basée sur le *root means square* (RMS) du déplacement résiduel de l'optique de l'appareil photo [101] : cette métrique est proposée pour mesurer les performances des stabilisateurs d'image optique en mesurant la valeur efficace du déplacement résiduel de l'optique par rapport à une position initiale fixe. Elle pourrait être éventuellement dérivée pour prédire la qualité des images du bougé résiduel. Mais pour l'instant il n'existe pas de lien entre la perception visuelle et la valeur RMS du déplacement.

2.1.2.8 Conclusion et limites sur les métriques objectives existantes

De nombreuses métriques ont été développées par les chercheurs pour mesurer la qualité des images, nous les synthétisons dans le tableau 4. Les approches ont été diverses comme le

montre le grand nombre de métriques existantes. À chaque fois, ces métriques répondent à un besoin particulier. Concernant nos besoins réels, pour modéliser une métrique de flou de bougé, nous avons les contraintes suivantes :

- Être une métrique prédictible ; nous voulons pouvoir prédire la qualité de l'image finale quelle que soit la scène afin de réagir en conséquence pour supprimer le flou de bougé.
- Être de faible complexité ; afin de l'intégrer facilement pour tout type d'application embarquée.
- Être une métrique précise, fiable et fidèle à la perception humaine.

Nous observons que les métriques existantes, même si parfois elles peuvent répondre de manière partielle, ne répondent pas à la totalité de nos besoins. En conséquence, nous avons besoin de développer une nouvelle métrique de qualité.

Tableau 4. Etat de l'art sur les métriques de qualité d'image pouvant servir à mesurer le flou sur une image. Légende : D = Déplacement de l'appareil ; Opt = Optique de l'œil ; SR = Sans référence ; ? = Non disponible.

| Métrique | Variance | Seuil de fréquence | Histogramme avec seuil | Auto corrélation | Dérivée | Entropie | Histogramme des coefficients de la DCT | Estimation de la taille des bords | Kurtosis | Contraste local | Taille des bords orientés | PSNR | SSIM | VIF | VSNR | MSVD | Niveaux de coefficient de transformé en ondelette | Immuniée au bruit | Rapport de filtre | Spectre de puissance | Probabilité de détection du flou | Rayon de la PSF estimée | Estimation du flou gaussien | Optométrie | MTF | BxU | RMS du déplacement résiduel | | |
|--------------------------------------|----------------|--------------------|------------------------|------------------|---------|----------|--|-----------------------------------|----------|-----------------|---------------------------|------|----------------|--------------------|------|------|---|-------------------|-------------------|----------------------|----------------------------------|-------------------------|-----------------------------|------------|------|-------------------|-----------------------------|----------|-----|
| Référence | [74] | [75] | [75] | [76] | [76] | [77] | [78] | [80] | [81] | [82] | [83] | [85] | [86] | [87] | [88] | [89] | [90] | [91] | [92] | [93] | [94] | [95] | [96] | [98] | [99] | [100] | [101] | | |
| Orientation | Autofocus | | | | | | Compression | | | | | | Généraliste | | | | | | Flou | | | | | | Opt | Chaine de l'image | | | |
| Entrée | Sans référence | | | | | | Sans référence | | | | | | Avec référence | | | | | | Sans référence | | | | | | PSF | SR | PSF | D | |
| Prédictive | Non | | | | | | Non | | | | | | Non | | | | | | Non | | | | | | oui | | | | |
| Complexité | + | ++ | + | + | + | + | ++ | + | ++ | ++ | ++ | + | + | ++ | ++ | ++ | ++ | ++ | + | ++ | + | + | + | - | ++ | ? | - | | |
| Comparaison à la perception visuelle | Aucune | | | | | | Type MOS | MOS | | | | | | Base de donnés MOS | | | | | | Type MOS | Non | Type MOS | MOS | Non | MOS | non | MOS | Type MOS | non |

2.2 Métrique de qualité proposée pour le flou de bougé

2.2.1 Limite des méthodes de l'état de l'art

Nous avons d'une part, les expériences visuelles MOS qui ne permettent pas d'obtenir des mesures objectives précises de la perception du flou, et d'autre part des métriques qui tendent de prédire l'impact visuel du flou, mais qui ne satisfont pas nos besoins comme discuté précédemment. De ce fait, je propose une nouvelle métrique pour l'évaluation du flou de bougé validée par des expériences de JND et des règles de qualité. Pour cela, nous nous basons sur les hypothèses de la section suivante.

2.2.2 Hypothèses pour la construction d'une métrique de bougé

Une métrique basée sur la PSF du flou de bougé

Comme nous l'avons montré dans le premier chapitre, le flou de bougé d'une image peut être modélisé par sa matrice de PSF. En d'autres termes, la PSF est la carte d'identité du flou de bougé de l'image. Contrairement aux flous linéaires, les flous arbitraires ont une forme complexe de PSF avec une taille et des valeurs de coefficient variables. Une façon intuitive de modéliser l'impact de ces flous arbitraires est de considérer la forme de cette PSF avec ses différentes valeurs.

Les paramètres prépondérants

Pour étudier l'impact de la forme de la PSF sur une image, nous avons développé dans des études préliminaires, des PSF avec des formes particulières pour mieux connaître l'évolution de la dégradation en fonction de sa forme et de ses valeurs. Les formes de PSF particulières utilisées ont été : des disques uniformes avec des rayons variables, des gaussiennes avec des sigmas variables, des flous linéaires, des points distincts avec des tailles rayons différents reliés entre eux ou non... Nous en avons déduit que l'impact de la PSF sur le flou varie suivant au moins trois tendances distinctes mais corrélées :

- L'étalement de la PSF : plus la PSF sera étalée, plus l'image sera floue et inversement.

- La valeur des points de la PSF : plus les points de la PSF seront uniformes, plus le flou sera uniforme et inversement.
- Le rapport des valeurs entre les points de la PSF qui contribuent à l'image principale, par rapport aux points de la PSF qui contribuent au flou de l'image.

Plus la PSF sera petite et concentrée sur un seul point, plus l'image sera nette. C'est ce que les systèmes de stabilisation d'image tendent à faire : de réduire la PSF à une matrice avec une seule valeur 1 ou une matrice de PSF totalement incluse dans le CdC.

Tentative d'une Métrique basée sur le Cercle de Confusion

Dans nos premières tentatives de recherche de métrique de qualité d'image exploitant la PSF, nous nous sommes orientés vers une métrique basée sur le CdC. Nous sommes partis de la propriété que si la PSF est entièrement comprise dans le CdC quel que soit le bougé à l'intérieur du CdC, l'image serait vue par l'observateur comme idéale. Ainsi nous avons pris l'hypothèse que pour une image floue, donc avec une PSF plus grande que le CdC, nous pouvions séparer la PSF en deux composantes. La première composante serait la partie de la PSF comprise dans le CdC correspondant à l'image nette ou en d'autres termes l'image principale, alors que la deuxième composante serait la partie de la PSF extérieure au CdC correspondant à l'image floutée. Ainsi la deuxième composante est une PSF résiduelle définie comme étant la PSF de bougé initiale où l'on a retiré son CdC. Le calcul de la métrique se faisait à partir de la PSF résiduelle en prenant en compte la valeur des éléments de cette PSF ainsi que la distance d'écartement au centre du CdC retiré. Le choix de la position du CdC à retirer de la PSF pour construire la PSF résiduelle était basé sur le critère de minimisation de la métrique. Plus l'image était floue, plus la métrique était importante. Lors des tests, cette approche n'a pas été concluante au regard des résultats finaux de la métrique. Même s'il est difficile d'en conclure une loi générale, nous constatons qu'il est impossible de séparer la partie de l'image nette avec l'image de flou à partir des informations de la PSF. En effet, le flou final de l'image est bien la résultante des deux composantes et non d'une seule.

L'obtention de la PSF de bougé

La PSF peut être déduite soit directement à partir de l'intégration des sorties du gyroscope de l'appareil photo, soit d'un modèle de simulation de bougé. Notre métrique utilise les données de la PSF pour calculer directement le flou ressenti par l'utilisateur sur l'image finale. Comme détaillé par [85] ce type de métrique de qualité peut être classé dans les métriques de référence réduite, car la seule information fournie à la métrique est la PSF

d'entrée, les informations de flou de la future image. Ma métrique a donc la particularité d'être prédictive.

Le protocole de validation de la métrique

Pour valider la métrique par rapport à la perception visuelle du flou ressenti, j'ai mis en place cette démarche :

- Premièrement, l'expérience basée sur les JND m'a permis d'obtenir la variation de la perception visuelle en fonction d'un flou de bougé simple, le flou linéaire.
- Deuxièmement, j'ai étalonné la métrique par rapport à la variation de flou linéaire bien définie.
- Troisièmement, j'ai étendu les résultats de ce type de flou particulier à des flous quelconques mais représentatifs des flous de bougé en utilisant l'expérience de mise en correspondance grâce à la règle de qualité ajustable.

C'est dans ce contexte et ces hypothèses que j'ai développé la métrique de qualité d'image pour le flou de bougé. Les détails de son implémentation sont présentés dans la section suivante.

2.2.3 Fonctionnement de notre métrique

Les différentes étapes utilisées pour réaliser la métrique sont les suivantes :

- 1) Prenant comme entrée une PSF de bougé, les coordonnées de chaque point de la matrice au centre de gravité de celle-ci sont calculées.
- 2) Puis chaque valeur de la PSF est pondérée par leur distance au centre de gravité précédemment calculé.
- 3) Tous ces coefficients pondérés de la PSF sont ensuite sommés pour obtenir un écart-type spatial.
- 4) La valeur 1 est ajoutée à ce résultat intermédiaire puis il est transformé par la fonction logarithmique. Nous nous attendions à une fonction logarithmique dans notre métrique car les études physiologiques sur le système visuel humain [102, 103, 104] démontrent une variation logarithmique de la vision humaine. Nos résultats expérimentaux vont aussi confirmer ces observations. Ainsi, ce décalage de 1 nous permet de calculer les PSF sans bougé.

- 5) Puis, deux coefficients d'ajustement sont utilisés pour calibrer la métrique avec la perception visuelle.
- 6) Enfin, pour les très faibles bougés globalement compris dans la notion de CdC, un seuil est utilisé afin d'avoir des valeurs toujours positives ou égales à zéro quelles que soient les valeurs de la PSF.

L'équation finale de la métrique est la suivante :

$$M_0 = -a \cdot (\ln(STD_{2D}(PSF) + 1)) + b$$

$$M = \begin{cases} M_0 & , M_0 \leq 0 \\ 0 & , M_0 > 0 \end{cases} \quad (10)$$

où M est la métrique de netteté en JND, 'a' et 'b' sont des coefficients d'ajustement.

2.3 Expériences visuelles de validation de la métrique

Dans cette partie, deux expériences visuelles ont été menées pour d'une part étalonner la métrique et d'autre part, valider ses performances de prédiction de qualité d'image. La première expérience consiste à quantifier le flou de bougé linéaire afin d'étalonner la métrique par rapport à ce type de dégradation. Cette expérience se base sur la méthode de mesure psycho-visuelle de JND. Puis une deuxième expérience est conduite pour valider la métrique dans sa capacité à prédire la qualité d'image de flou quelconque. Cette expérience se base sur la méthode de notation par mise en correspondance à l'aide de l'utilisation de la règle de qualité. Ces deux expériences sont présentées dans les deux sections suivantes.

2.3.1 1^{ère} expérience de « *just noticeable difference* »

Les objectifs de l'expérience 1

La première expérience visuelle que j'ai conduite a deux principaux objectifs. D'une part de quantifier un flou linéaire simple pour créer une règle de qualité. Cette règle permettra de quantifier le flou de bougé afin de mesurer des flous quelconques. Et d'autre part de calibrer la métrique de qualité d'image suivant un type de variation recouvrant ainsi une grande amplitude de variation. Pour atteindre ces objectifs, nous nous sommes basés sur les

recommandations de la norme [72] comme mentionné dans l'état de l'art section 2.1.1.2 pour implémenter l'expérience de vote de différence juste perceptible (JND).

La base d'image de référence

La norme recommande d'utiliser plusieurs images afin de ne pas rendre dépendant l'expérience suivant un seul type d'image. L'optimum défini par la norme est de six scènes différentes. J'ai pu sélectionner avec attention six images provenant de base de données libre représentant une variété de scène typique présentée ci-dessous.



Figure 34. Sélection de scènes typiques pour les différentes expériences visuelles.

Cette base de données inclut un portrait (Lena), une image d'animaux (Les perroquets), un paysage (Le phare), une scène d'intérieur (La cuisine), une scène d'extérieur de ville (Les bâtiments) et une scène d'action (Course de motos).

L'interface de l'expérience

Une interface graphique a été développée sous Matlab présentée ci-dessous.

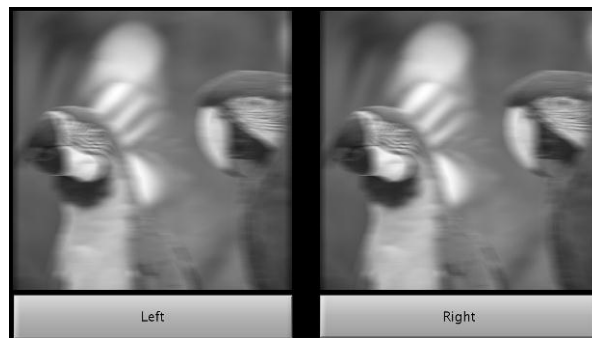


Figure 35. L'interface graphique de l'expérience de JND.

Cette interface permet de présenter deux images sur le même écran. Pour chaque observateur, la même base d'image floutée est présentée de manière aléatoire avec un couple d'images différentes mais proches en matière de dégradations. Ce couple d'images est affiché à gauche ou à droite de manière aléatoire. L'algorithme implémenté sous Matlab est présenté synthétiquement figure 36.

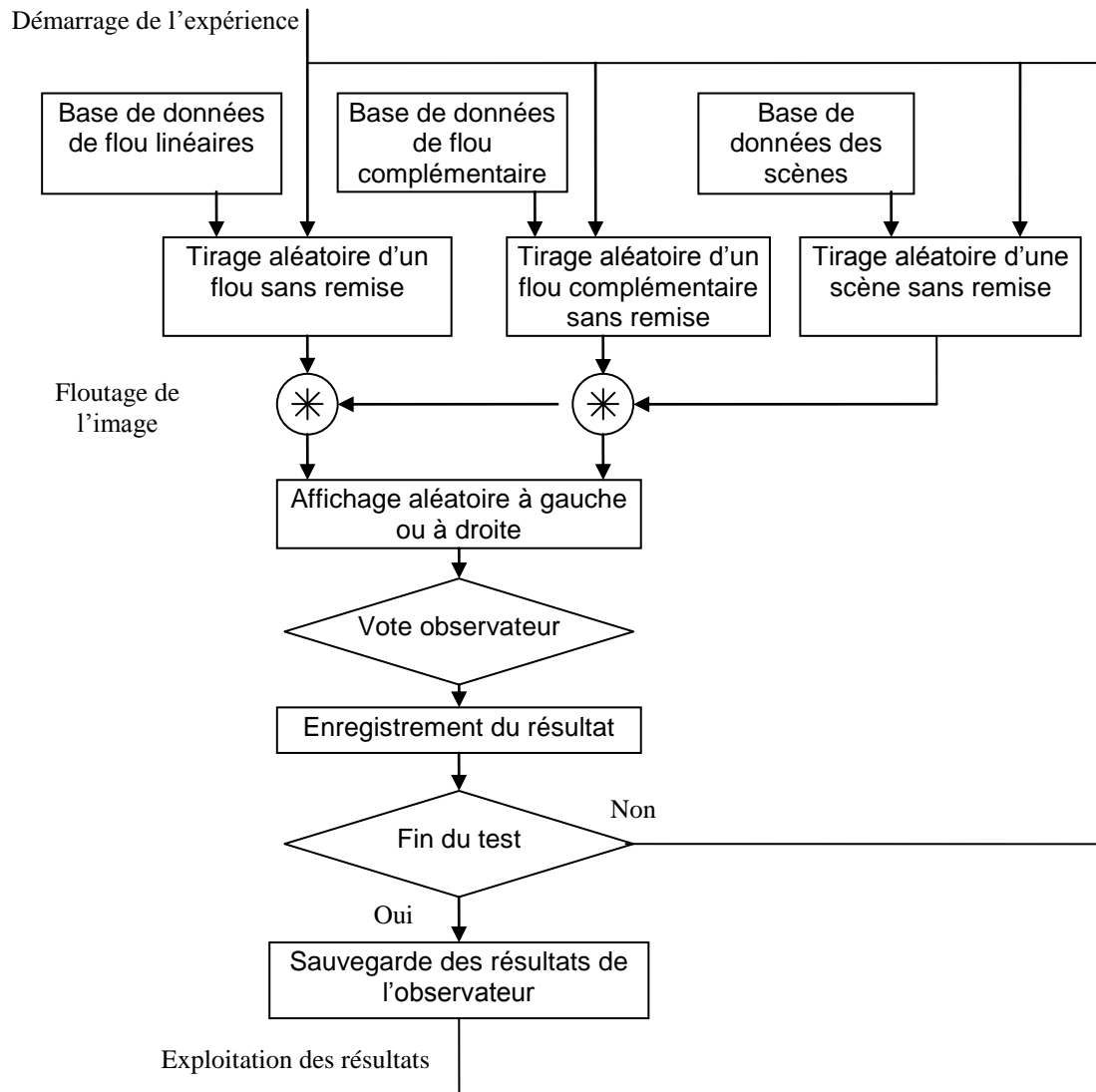


Figure 36. Schéma bloc représentant l’algorithme implémenté dans Matlab pour réaliser l’expérience de JND.

Dans notre expérience, 16 couples de flous différents ont été répartis tout le long de la dégradation pour couvrir une large gamme de dégradations. Cela a été fait grâce à des expériences préliminaires. Chaque couple d’images est présentée 30 fois de manière aléatoire pour chaque observateur. Ce nombre de répétitions nous permet de calculer l’évolution de la perception pour chaque observateur suivant la dégradation. Pour chaque sujet et pour chaque couple de flou, le pourcentage de bonnes réponses nous permet de calculer l’évolution de la perception en JND. Ainsi, chaque personne a comparé 480 couples d’images. Au final, la base totale de réponses comprend 9 600 réponses de vote.

Les conditions de visualisation

Les conditions d'affichage des images sont les suivantes. Des images affichées de taille de 512×512 pixels avec un rapport 1:1 sur un écran LCD HP 1825 de 18.1 pouces (1280×1024). L'écran a un pas pixel de 0.281 mm, un pic de luminosité de 250 cd/m² et un rapport de contraste de 250 :1. Afin de s'assurer que l'écran n'utilise pas de correction dynamique d'amélioration de netteté, des photos photographiant les pixels de l'écran ont été prises pour s'assurer de la linéarité des réponses de pixel, particulièrement dans les zones de transition de couleur.

Nos expériences étant basées sur des simulations de flou de bougé, les artefacts de simulation numérique doivent être évités afin d'avoir des résultats non biaisés. Un exemple de ce type d'artefact est présenté figure 37.

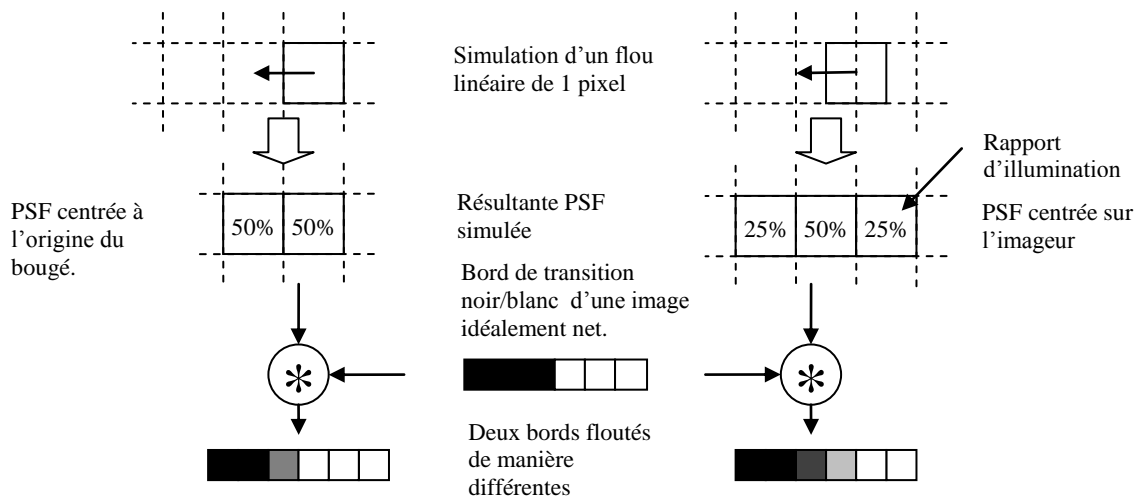


Figure 37. Deux manières différentes de générer un flou de 1 pixel dans le domaine numérique.

Cet exemple montre deux manières de générer un flou linéaire de bougé d'un seul pixel avec deux références de générations différentes. Ces flous ayant une même longueur devraient être perçus de la même manière. À une distance typique d'observation où 1 pixel est égal à l'acuité visuelle de 1/60°. Ces deux flous de même taille vont être perçus au final avec une qualité d'image différente. Afin d'éviter ce type de problème, la distance d'observation des images a été augmentée de manière à ce que des personnes ayant une acuité visuelle normale ne puissent plus distinguer de différence. La distance a été déterminée par des expériences avec un observateur ayant une bonne acuité visuelle (environ de 12/10). Le protocole d'expérience utilisé pour déterminer cette distance critique a été la même que pour

l'expérience de vote pour l'image de meilleure qualité. L'observateur devait choisir en comparant l'image de meilleure qualité entre un couple de deux images ayant le même bougé mais générées de plusieurs manières différentes. Plusieurs simulations de longueurs de bougé ont été testées pour connaître l'influence de la force du bougé avec le cas limite comme présenté figure 37. Ce cas limite consiste à générer un flou de bougé soit centré sur l'origine du bougé, soit centré sur la PSF finale de floutage. Plusieurs longueurs de bougé ont été testées comme le bougé d'1 pixel sur la figure 37. Ils sont [0,05 ; 0,1 ; 0,25 ; 0,5 ; 0,75 ; 1 ; 1,5 ; 2,5 ; 3] pixels. Pour chaque couple d'images, le vote a été répété de manière aléatoire pendant 30 fois. Puis la distance d'observation a été augmentée progressivement par paliers, jusqu'à atteindre environ 50 % de réponse pour chacun des couples. Ainsi, ce pourcentage de 50 % correspond au fait qu'il est impossible pour l'observateur de dire quelle image est la meilleure impliquant donc que les images ont la même qualité. Les résultats du cas dit limite d'un bougé d'un pixel sont présentés dans le tableau 5.

Tableau 5. Résultats partiels de l'étude de l'influence de la distance pour deux PSF différentes simulant le même type de flou de 1 pixel.

| | | | | | | | | | | |
|------------------------------------|----------------|-----|----------------|----|-----------------|----|-----------------|----|-----------------|----|
| Angle de vision d'un pixel (degré) | $\frac{1}{60}$ | | $\frac{1}{90}$ | | $\frac{1}{120}$ | | $\frac{1}{150}$ | | $\frac{1}{180}$ | |
| Distance de visualisation (m) | 0,97 | | 1,46 | | 1,94 | | 2,43 | | 2,91 | |
| Matrice de flou (PSF) | | | | | | | | | | |
| Réponse (%) | 0 | 100 | 3 | 93 | 22 | 78 | 37 | 63 | 47 | 53 |

Nous avons ainsi déterminé une distance nécessaire de 2,91 m pour éviter ce problème d'échantillonnage. A cette distance, il est impossible de voir la différence pour des personnes ayant une acuité visuelle normale quelle que soit la manière de générer les images floues. Pour le reste des expériences visuelles, cette distance de visualisation a été fixée constante pour tous les observateurs en utilisant une chaise fixe.

Les observateurs

L'expérience de vote forcé par comparaison d'images a été réalisée avec 20 personnes, 6 femmes et 14 hommes dont l'acuité visuelle, corrigée ou non, est supérieure ou égale à 10/10 avec les deux yeux comme recommandé par [72]. L'âge des observateurs est compris entre 21

et 45 ans. Cinq des vingt observateurs se sont déclarés avoir une expérience dans le domaine de la qualité d'image, mais la moyenne des résultats des observateurs dit expérimentés a montré que ce groupe était à peine meilleur que les observateurs dit normaux. Nous en avons déduit qu'il y a une faible influence des observateurs expérimentés sur les résultats en JND des observateurs.

Optimisation de l'expérience

Pour mener à bien l'expérience de vote, j'ai dû faire face à deux contraintes fortes rappelées dans la norme [72] :

- Limiter le temps d'expérience malgré un grand nombre de votes par observateurs afin d'avoir des résultats statistiquement exploitables. C'est-à-dire pour chaque observateur, que l'expérience ait une durée inférieure à une heure pour éviter leur fatigue. Mais aussi que chaque couple de comparaison soit répété et voté au moins 30 fois pour avoir un nombre de données statistiques assez grand.
- Avoir des résultats de vote des observateurs exploitables. C'est-à-dire que leur pourcentage de bonnes réponses soit compris entre 58 % et 86 %. Soit des valeurs en JNDs compris dans l'intervalle : $0,3 \leq \text{JNDs} \leq 1,5$. En effet pour des valeurs supérieures à 1,5 JNDs, la réponse de la fonction entre dans une zone de saturation. Pour les valeurs inférieures à 0,3 JNDs, il sera difficile d'en déduire ensuite une pente. Ainsi dans ces zones, une erreur de mesure entraînera une erreur importante dans le résultat final pour le couple d'images.

Pour remédier à ces limitations, j'ai mis en place une première expérience de calibration afin de m'adapter au comportement visuel des observateurs pour optimiser les résultats de l'expérience. Pour cela, trois différentes bases de données de flou ont été créées avec chacune des couples de flou différent. Une première dite « calibre facile », pour les personnes qui auraient plus de mal à juger de la différence entre les images, composées de flous avec des différences importantes. Une deuxième dite « calibre normal », pour les personnes qui sont dans la moyenne. Et une troisième dite « calibre difficile », avec des petites différences de flou pour ceux qui ont une forte aptitude à différencier les flous. Des exemples de résultats de l'équation 9 avec deux calibres différents sont présentés figure 38.

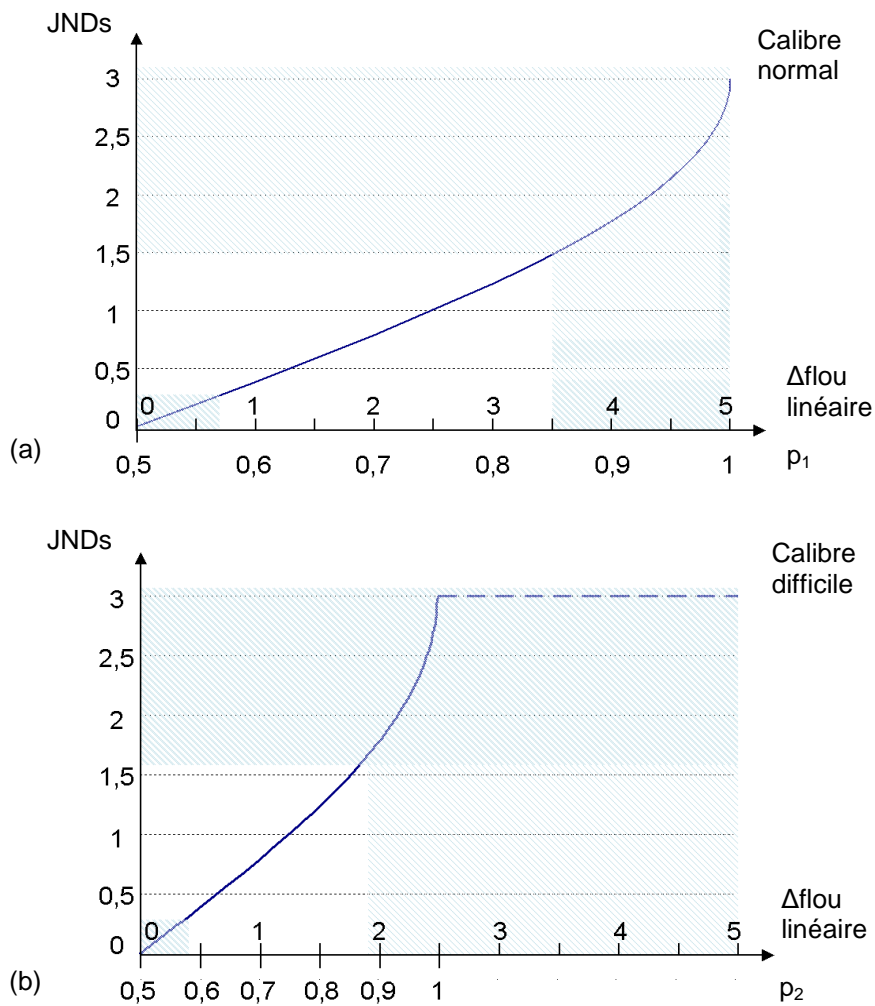


Figure 38. Exemple de résultats de deux observateurs différents ayant utilisé deux calibres différents. (a) Résultats p_1 de couple d'images suivant leur $\Delta\text{flou linéaire}$, d'un observateur avec une vision normale. (b) Résultats p_2 de couple d'images suivant leur $\Delta\text{flou linéaire}$, d'un observateur avec une bonne vision.

Prenons un exemple : un premier observateur a une vision « normale » avec des résultats de bonnes réponses p_1 présentés figure 38a. Donc pour le cas d'un couple d'images séparé de 2 $\Delta\text{flous linéaires}$, il a obtenu 70 % de bonnes réponses. Maintenant, un deuxième observateur a une bonne vision « deux fois meilleure » que le premier. Ainsi pour le deuxième observateur le 2 $\Delta\text{flous linéaires}$ est équivalent à un flou 2 fois plus grand que pour le premier observateur, soit 4 $\Delta\text{flous linéaires}$. Le deuxième observateur obtiendra alors 90 % de bonnes réponses dans ce cas. 90 % de bonnes réponses n'est pas un résultat exploitable, car il se trouve dans la zone de saturation. Pour éviter ce problème de saturation, on lui donne un calibre deux fois plus difficile présenté figure 38b avec des $\Delta\text{flous linéaires}$ divisés par deux. Dans notre même exemple, il obtiendra seulement 70 % de bonnes réponses pour 1 $\Delta\text{flou linéaire}$ car le calibre

est deux fois plus difficile. Ainsi grâce à cette calibration, le deuxième observateur obtiendra lui aussi des résultats exploitables.

Suite à cette méthode, j'ai fait passer un premier test de calibrage à tous les observateurs afin de choisir la base de flou la mieux adaptée à leur acuité visuelle pour maximiser les performances du test JND. Dans ce test de calibration, seuls trois couples d'images de taille de flou différents ont été présentés 2 fois pour chaque image, soit un total de 12 répétitions par couples de Δ flou. Puis le pourcentage de réponses correctes a été calculé pour ces 36 votes. La moyenne des résultats corrects inférieurs à 60 % ou supérieurs à 80 % démontre respectivement que le test de calibration était trop dur ou trop facile. Dans ces cas, la base de données plus ou moins facile suivant le cas a été sélectionnée pour l'observateur. Au final, neuf personnes ont passé le calibre dit difficile et les autres le calibre dit normal.

Comme recommandé par [72], pour chaque observateur l'expérience globale incluant les explications, le test de calibrage et le test final ont été inférieurs à une heure. La partie seule du test final a été de 38 minutes en moyenne. Si elles le souhaitent, les personnes étaient autorisées à prendre une pause de quelques minutes pour éviter les problèmes de maux de tête dus à la concentration et à la répétition d'images floues.

Le protocole de l'expérience

Les différentes étapes successives du protocole de tests implémentés sont présentées ci-dessous :

1. L'acuité visuelle a été testée pour chaque personne à l'aide d'un test type ophtalmo afin de nous assurer que les observateurs ont une vision normale ou supérieure à 10/10 avec ou sans correction. Pour cela, le sujet devait lire des lignes de lettres affichées à l'écran à une distance de 5 m correspondant à une note de perception visuelle jusqu'à ce que la personne face une faute dans la ligne.
2. Une première session d'entraînement a été implémentée pour prendre en main l'interface graphique, le mode de vote et les différentes recommandations. Les explications données à l'observateur sont en annexe 1. La consigne de jugement donnée aux observateurs a été de choisir, entre les deux images présentées, l'image de meilleure qualité. Et que ce jugement soit plutôt donné en fonction de la première impression du ressenti de la personne afin d'éviter un long temps de scrutation des détails de l'image. Ce qui augmenterait d'une part le temps total de l'expérience et

d'autre part qui biaiserait l'expérience car l'observateur aurait tendance à se fixer sur des détails de l'image et non plus sur la qualité globale de l'image.

3. Ensuite, le test rapide de calibration a été présenté à l'observateur pour sélectionner la base de flou la mieux adaptée à sa perception visuelle.
4. Puis le test principal de l'expérience des 480 votes a été passé avec le bon calibre, sans en informer la personne pour ne pas l'influencer dans ses votes. Les résultats des paires de comparaison ont été convertis en JNDs par chaque observateur grâce à l'équation 9. Comme recommandé par [9], les valeurs en dehors de [0,3-1,5] JND ont été enlevées pour éviter les problèmes de précision. Pour chaque sujet, l'évolution de la qualité de l'image suivant la variation du flou a été calculée grâce aux équations 11 et 12 suivantes.

$$Q(l) = Q_R - \int_{l_R}^l \frac{dl'}{\Delta l_j(l')} \quad (11)$$

avec l la longueur du flou linéaire, $Q(l)$ la qualité d'image en JND à la longueur l , l_R la longueur de référence qui est ajustée à la qualité d'image référent Q_R . Nous avons choisi comme référence de qualité $Q_R(0) = 0$ comme la référence parfaite de qualité. $\Delta l_j(l')$ correspond à la pente des JND de la fonction d'incrémentaire suivant l'équation 12.

$$\Delta l_j(l') = \frac{|l'_x - l'_y|_j}{JNDs_j} \quad (12)$$

Où l'_x et l'_y sont les longueurs des flous linéaires des couples d'images et $JNDs_j$ le résultat du vote de ce couple en unités de JND. Les différentes étapes de traitement des données pour un observateur sont présentées figure 39. Ainsi que sa perception visuelle en fonction de la dégradation linéaire en JND figure 40.

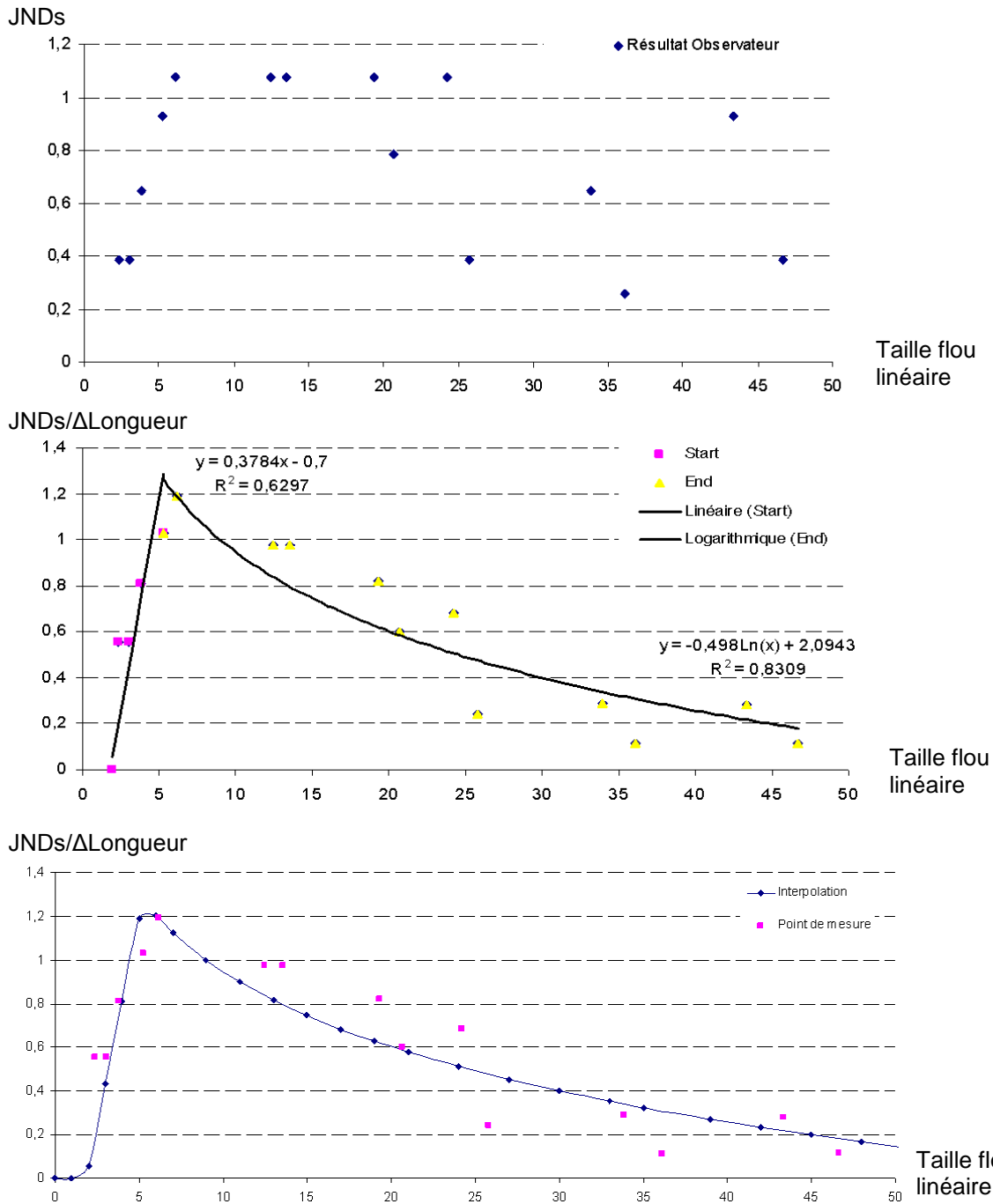


Figure 39. Exemple d'exploitation de résultats pour un observateur.

- Enfin pour connaître l'évolution moyenne de la perception humaine, tous les résultats de chaque observateur ont été moyennés entre les 20 observateurs, présentés figure 41.

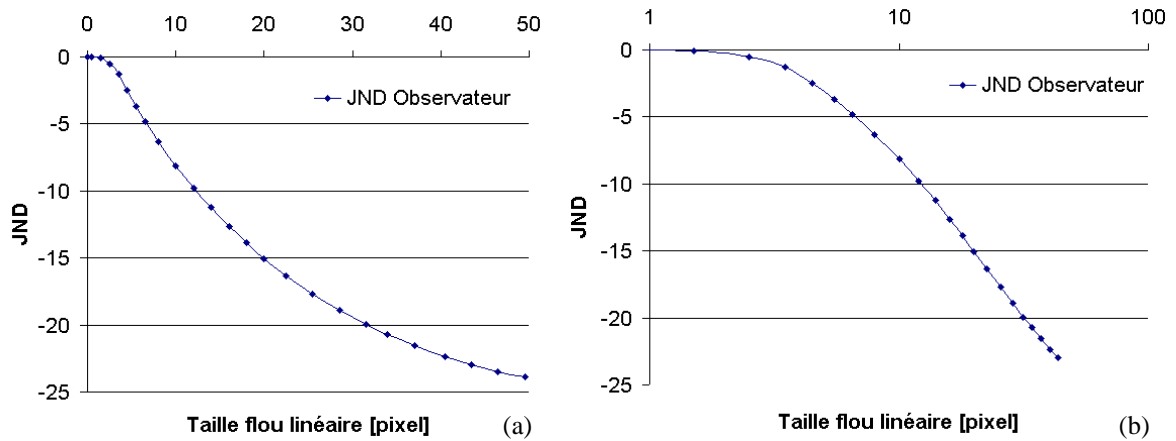


Figure 40. Résultats de l'exemple en JND de l'observateur. (a) Echelle linéaire. (b) Echelle logarithmique.

2.3.1.2 Résultats de la 1^{ère} expérience JND

Deux types de résultats sont présentés suite aux résultats de l'expérience présentée figure 41. Les résultats d'observation qui concernent la perception visuelle humaine et les résultats de la métrique de qualité de flou de bougé développée.

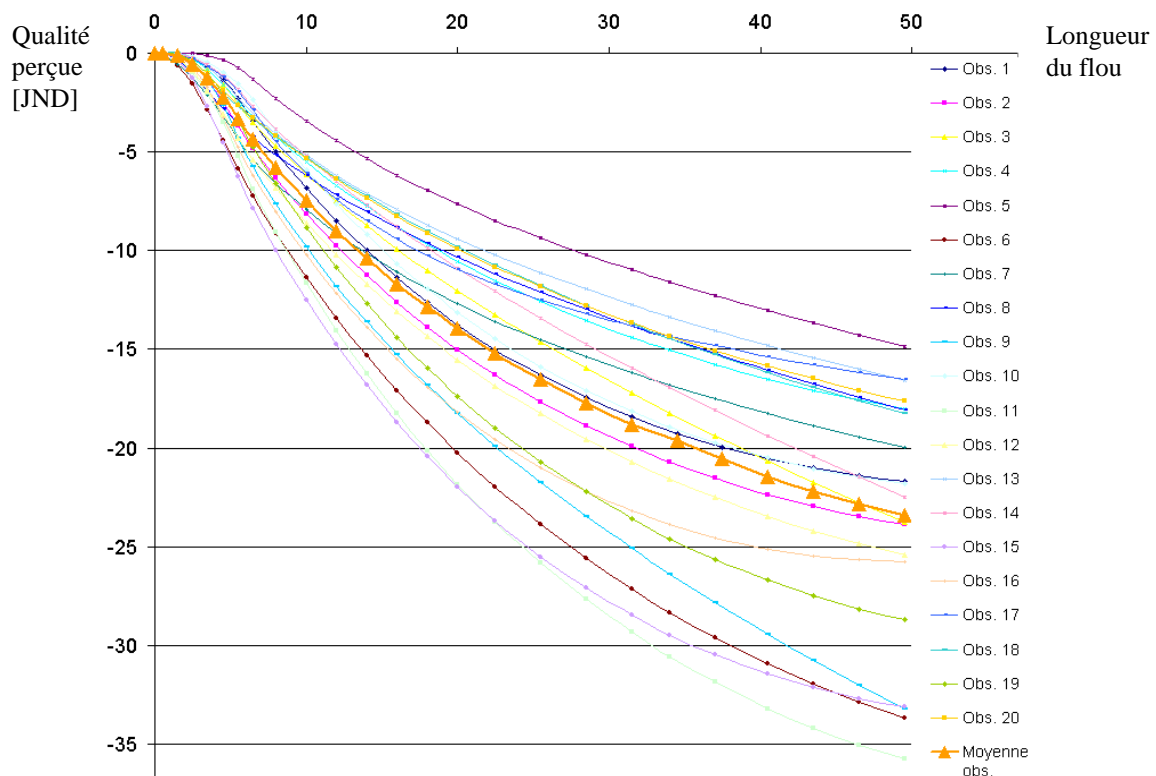


Figure 41. Évolution de la perception visuelle pour chaque observateur suivant une dégradation de type flou linéaire. En orange la valeur moyenne de la perception visuelle des 20 observateurs.

Résultats d'observation

Différentes observations peuvent être déduites de ces résultats :

- Globalement nous observons une forte variabilité des résultats traduisant le caractère subjectif d'une telle expérience. Néanmoins, il existe une tendance commune aux différents observateurs. Comme espéré plus le flou linéaire est long, plus il est désagréable pour l'observateur.
- Pour des faibles flous (inférieurs à 3 pixels), la qualité perçue par l'œil est quasiment constante et proche de la netteté parfaite de 0 JND. Cela montre l'impact de l'acuité visuelle. Lorsque le flou est inférieur au cercle de confusion, le flou est difficilement discernable par les observateurs.
- Puis lorsque le flou augmente, la pente des JND augmente pour atteindre un maximum autour de 10 pixels, puis elle rediminue. Cela montre que pour des flous de taille moyenne, le système visuel humain a une forte discrimination pour dissocier des images de flous différents.
- Lorsque le flou devient important, la discrimination du flou devient de plus en plus difficile. D'après la figure 40.b, si nous observons les résultats sur une échelle logarithmique, pour des tailles de flou important nous obtenons une droite (c'est-à-dire un logarithme en échelle linéaire). Cette tendance de discrimination logarithmique a aussi été observée par [48] pour des expériences de JND similaires.
- Nous avons pu observer que la force de différenciation de deux images de différentes qualités n'est pas directement liée à l'acuité visuelle. C'est-à-dire que si l'on regarde en détail, la note d'acuité des observateurs réalisée par le test ophtalmo, les personnes avec une très bonne acuité visuelle (>14/10) n'aurons pas toujours les meilleurs résultats même si en moyenne ils seront légèrement meilleurs que les personnes ayant une acuité dite normale (10/10).
- Du fait de l'intégration, pour générer cette évolution de la perception en fonction du flou linéaire, l'écart-type entre les différents observateurs augmente avec la force du flou.

Ces premiers résultats de vérité de terrain nous ont permis d'ajuster notre métrique de qualité et de valider notre métrique dans le cas du flou linéaire.

Nous avons aussi déduit de cette variation linéaire, une règle de qualité d'image analogue à la règle décimètre classique. Cette règle nous servira dans la deuxième expérience à mesurer des flous arbitraires.

Calibration de la métrique par rapport au flou linéaire

L'expérience de vote forcé nous permet d'avoir une vérité de terrain pour les flous de bougé particuliers que sont les bougés linéaires. Cela nous a permis de calibrer la métrique de qualité d'image en ajustant les coefficients a et b de la métrique. Nous avons alors appliqué l'équation 10 sur des flous linéaires en ajustant a et b. Le meilleur ajustement trouvé est $a = 12.2$ et $b = 10$. La prédiction de la métrique dans le cas linéaire est présentée figure 42.

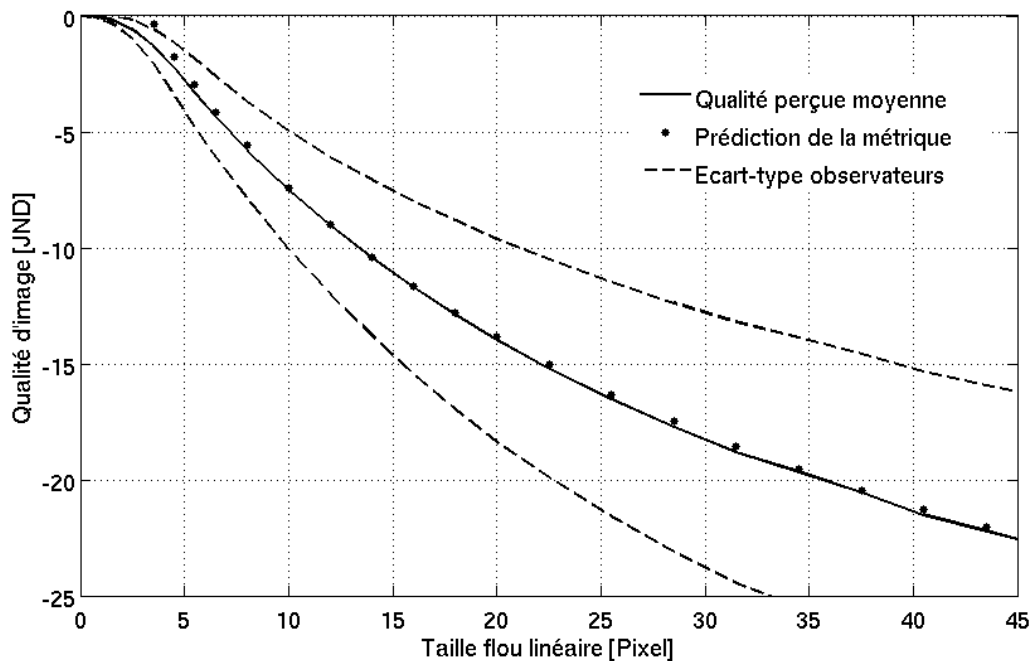


Figure 42. Prédiction de la métrique de qualité d'image (points) par rapport à la vérité de terrain pour l'évolution des flous linéaires (courbe).

Comme il peut être remarqué sur la figure, grâce à l'utilisation de ces coefficients d'ajustement la métrique arrive à prédire les flous linéaires de façon très précise sur une grande gamme de flous linéaires.

2.3.2 2^{ème} expérience de règle de qualité ajustable

Dans cette deuxième expérience, nous allons étendre notre mesure de flou linéaire à tout type de flou. Cela permettra de valider notre métrique pour des flous de bougé quelconques.

Génération de la règle de qualité

Pour mesurer la qualité des images avec des flous quelconques, j'ai utilisé les résultats de l'expérience de JND pour générer une règle étalonnée de qualité d'image. Une règle de qualité a été créée pour chaque image de la base de données. Cette règle a été réalisée par la même gamme de noyaux de bougé linéaire allant de 0 pixel à 45 pixels de bougé. La règle de qualité a été créée pour couvrir une gamme de 0 à -24 JND avec un pas d'espacement de 1 JND. Ainsi, chaque image a une large gamme de dégradations linéaires avec de faibles espacements en matière de qualité d'image. Toutes ces dégradations d'image progressives réparties sur toute la gamme créent la règle de qualité d'image. Cette règle de qualité d'image peut donc mesurer le flou d'une image en JND, analogue à la règle décimètre qui mesure une distance en cm.

Génération de flous de bougé quelconques

Pour générer les flous quelconques, nous avons utilisé notre modèle physiologique de tremblement humain [105] pour simuler une variété de temps d'exposition et de condition d'utilisation. Dix-huit matrices de PSF aléatoires ont été générées grâce à différents temps d'exposition recouvrant une grande gamme de temps d'exposition depuis des faibles temps d'exposition ($<1/64$ s) jusqu'à des temps longs d'exposition ($1/4$ s). Nous avons dérivé ces matrices de PSF en utilisant ces hypothèses : 200 g. pour la masse de l'appareil photo, une distance focale de 105 mm, $5 \mu\text{m}$ comme taille de pixel.

Les différentes PSF générées par notre modèle sont présentées figure 43.

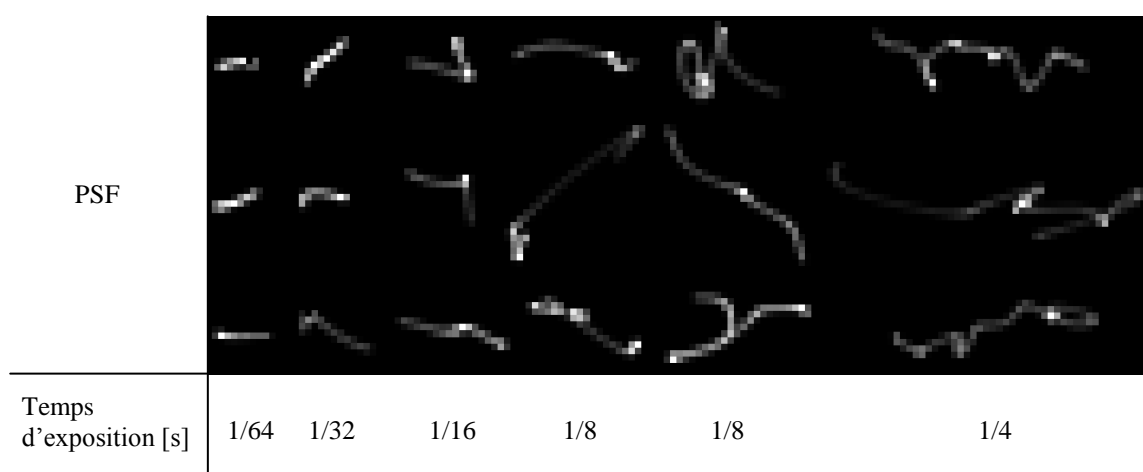


Figure 43. Dix-huit différentes PSF obtenues avec notre modèle de tremblement suivant différents temps d'exposition.

Cette base de données de flou nous donne différents scénarios de condition de prise de vue avec différents temps d'exposition. Ces flous quelconques vont être évalués par rapport à la vérité de terrain des flous linéaires en utilisant une expérience d'ajustement de règle de qualité d'image. Cette deuxième expérience nous mène à construire une deuxième vérité de terrain pour des flous quelconques.

Nouvelle interface graphique

La qualité des images a été évaluée grâce à la règle de qualité d'image présentée précédemment. Les conditions de visualisation ont été les mêmes que dans l'expérience de vote précédente. L'évaluation d'un flou de bougé quelconque a été mesuré par la présentation de deux images adjacentes présentées à l'observateur. La première image, celle qui est à évaluer, est dégradée par une PSF de flou quelconque. La seconde image, générée en temps réel grâce à l'ajustement de la taille du flou linéaire, est ajustée en netteté pour obtenir une équivalence en termes de qualité d'image. Pour chaque PSF de bougé quelconque, nous avons demandé au sujet d'ajuster le flou linéaire de façon à trouver la meilleure équivalence par rapport à la première image. Par cet ajustement, connaissant la valeur du flou linéaire grâce à la règle de qualité, il nous est possible de donner une note au flou quelconque évalué. Pour accomplir cette expérience, une deuxième interface graphique Matlab a été implémentée. Celle-ci est présentée figure 44 avec son ascenseur qui permet de faciliter l'ajustement de la règle de qualité en temps réel.

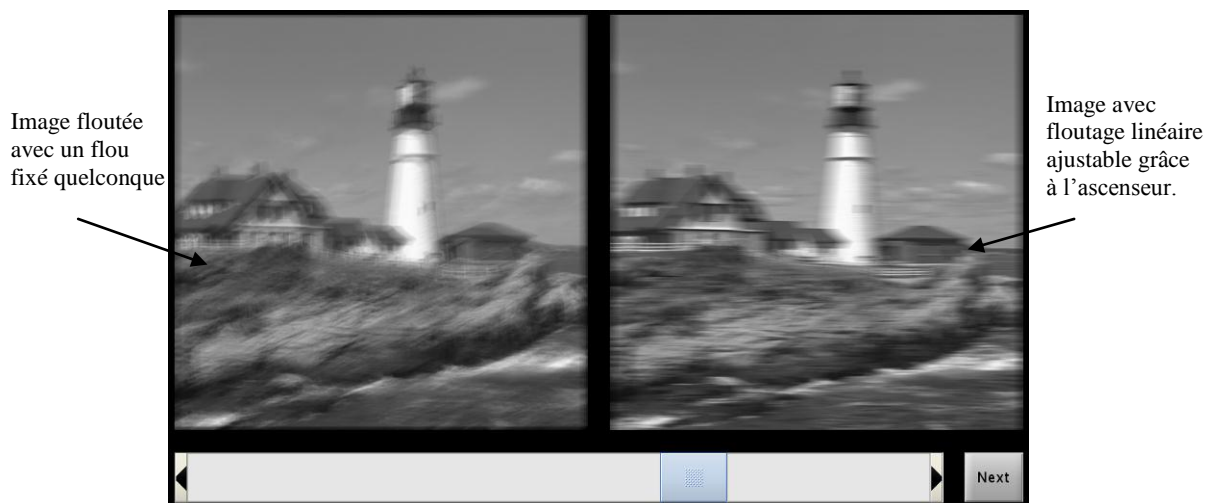


Figure 44. Interface graphique de l'expérience de règle de qualité d'image ajustable.

Nouveau groupe d'observateur

La même base de données d'images floutées a été évaluée par 20 observateurs. Les observateurs sont des adultes volontaires comptant 5 femmes et 15 hommes avec des âges répartis entre 21 et 45 ans. La plupart des observateurs n'ont aucune expérience en notation de qualité d'image. Chaque observateur a évalué les dix-huit matrices de PSF de flou quelconques par rapport au flou linéaire sur les six images de notre base de données. Nous avons alors obtenu une base totale de mesure de 2160 résultats. Le temps de l'expérience a été inférieur à 45 minutes comme recommandé par [72] afin de prévenir des risques de fatigue. Chaque ajustement a été en moyenne fait en moins de 15 secondes, ce qui correspond à un temps normal si on regarde d'autres expériences avec le même type de mesure [106]. Enfin, pour connaître l'impact de chaque flou quelconque, nous avons calculé leur note moyenne suivant les six scènes et des 20 observateurs.

Protocole de la seconde expérience

Dans cette expérience, nous avons utilisé ces différentes étapes :

1. L'acuité visuelle a été testée pour chaque observateur afin de s'assurer que leur vision soit normale ou supérieure à 10/10 corrigé ou non.
2. Un court exemple avec quelques images test a été présenté à l'observateur pour lui présenter l'interface graphique, son fonctionnement ainsi que les différentes recommandations présentées en annexe 2. Il a été demandé à l'utilisateur d'ajuster du mieux que possible les images en matière de qualité globale et non par rapport à une zone de l'image ou à un type de dégradation.
3. Ensuite, le test complet comprenant la base des 108 images floutées a été présenté à chaque observateur.
4. Les résultats ont été moyennés suivant les 6 scènes et les 20 observateurs afin de déterminer la moyenne spécifique à chaque matrice de PSF quelconque.

Cette nouvelle expérience nous permet d'en déduire une nouvelle vérité de terrain pour des flous quelconques. Les résultats sont présentés dans la section suivante.

2.3.2.2 Résultats de l'expérience et de la performance de prédiction de la métrique

La deuxième expérience utilisant la règle de qualité nous donne une vérité de terrain sur les dix-huit flous arbitraires. Pour évaluer les performances de notre métrique de qualité, à

prédire la qualité des images, nous avons appliqué l'équation 10 à chacune des dix-huit PSF. Ainsi, notre métrique de flou nous donne la qualité d'image prédite pour chaque scénario de flou. Ces résultats de prédiction sont comparés à ceux des observateurs dans la figure 45.

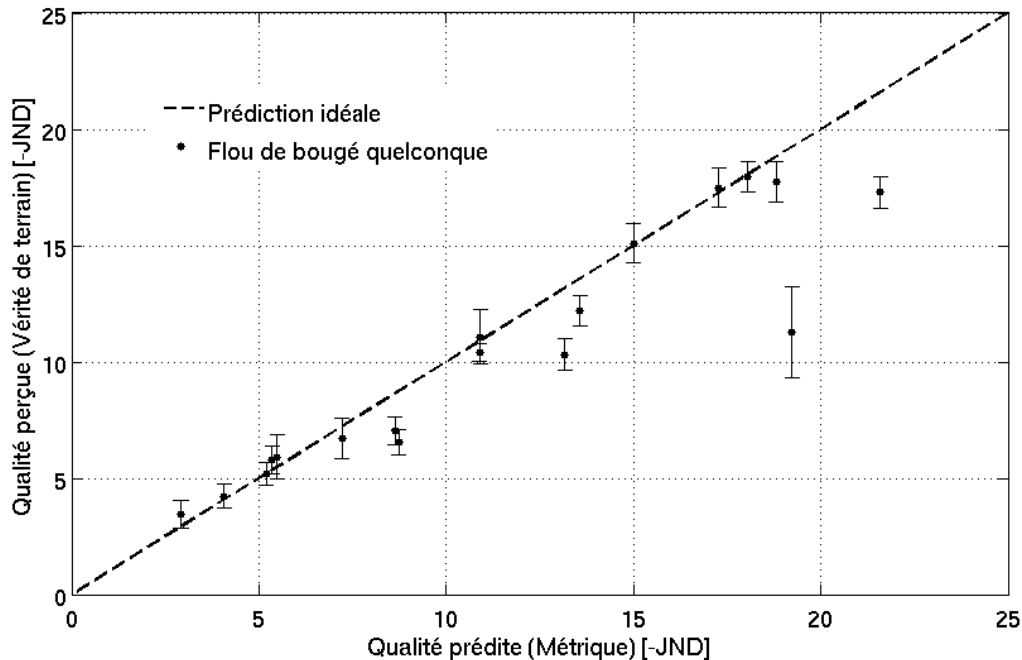


Figure 45. Comparaison de la prédiction de la métrique (abscisses) avec la vérité de terrain issue de l'expérience d'ajustement de la qualité des images (ordonnées). L'écart-type entre observateurs pour chaque flou est représenté par la barre d'erreur.

Cette deuxième expérience nous mène vers deux types de résultat : ceux qui sont propres à la règle de qualité et ceux qui sont propres à la métrique.

Résultats de la règle de qualité

Si nous regardons dans un premier temps les résultats du point de vue des expériences visuelles, nous pouvons observer une variabilité dans les résultats. Cela est normal pour une expérience subjective, mais nous remarquons les deux choses suivantes :

- Plus le flou ressemble à un flou linéaire, plus la mise en correspondance sera facile et donc similaire entre observateurs, de la sorte l'écart-type entre les mesures sera faible.
- Plus le flou a une forme exotique, plus il sera difficile pour l'observateur de faire correspondre un flou linéaire en termes de qualité. Alors au final l'écart-type sera fort. Le pire cas observé est lorsque l'image est de forme image dédoublée.

Résultats de la métrique

Les résultats montrent que la métrique a une très bonne aptitude de prédiction de la tendance globale. La métrique donne une bonne prédiction pour la majorité des points avec un écart de précision inférieur à 3 JND.

Pour quelques points, la métrique n'arrive pas à bien prédire la qualité des images. Elle sous-estime leur qualité, c'est-à-dire que l'observateur la perçoit finalement de meilleure qualité. Cela vient du fait que ce type de dégradation d'image est assez particulier, car elle génère une image de forme dédoublée comme présenté figure 46. Dans ces cas précis, nous pouvons remarquer que cette mauvaise prédiction est aussi corrélée à l'augmentation de l'écart-type de notation entre utilisateurs. Du fait de sa sous-estimation en matière de qualité, la métrique garantit dans le pire cas au moins la qualité minimum que représente une image pour la vision humaine.



Figure 46. Exemple d'image floutée, cas de l'image dédoublée.

2.3.2.3 L'influence des conditions de visualisation

Dans la première expérience visuelle de la section 2.3.1, j'ai placé les observateurs à 2,97 m de l'écran afin d'éviter le problème de sous résolution de l'écran comme expliqué précédemment. Maintenant en pratique, la distance de visualisation n'est pas si importante, elle est plus proche de 2 à 3 fois la hauteur de l'écran [106]. De plus la tendance actuelle est de les regarder soit directement sur l'écran de l'ordinateur ou encore sur la télévision et les

cadres numériques [3]. Suivant le support de visualisation de l'image, la distance de visualisation de celle-ci ne sera pas la même. Par exemple pour la télévision, la distance recommandée sera de 3 fois la hauteur de l'écran pour les téléviseurs HD et 6 fois pour ceux avec une définition standard [107]. Ou encore dans le cas des images imprimées sur des supports papier, la distance de visualisation va varier de 2 à 3,4 fois la hauteur de l'image suivant la taille de la photo [9]. Si l'on regarde ces distances de visualisations par rapport à la taille des points de l'image visualisée, nous pouvons observer que finalement cette distance correspond généralement à la distance pour laquelle 1 point de l'image est regardé par un angle de 1 arc minute, soit l'acuité visuelle. Ceci correspond à la distance pour laquelle il n'est plus possible de discerner un point de l'image. Excepté pour les expériences visuelles, nous nous basons sur cette hypothèse de distance de visualisation : que l'image est observée à la distance pour laquelle 1 point est regardé avec un angle d'un arc minute, soit l'acuité visuelle. Le flou ressenti dans les images n'étant pas perçu de manière linéaire comme le rappelle la figure 47.a, il y a donc une transformation à faire pour ramener la distance de visualisation où un pixel est regardé à 1 arc minute. Pour réaliser cette transformation, la courbe figure 47.a de la qualité suivant la dégradation est utilisée comme abaque car se rapprocher de l'écran d'un facteur 3, revient à augmenter la taille du flou de ce même facteur. Le résultat de cette transformation est présenté figure 47.b.

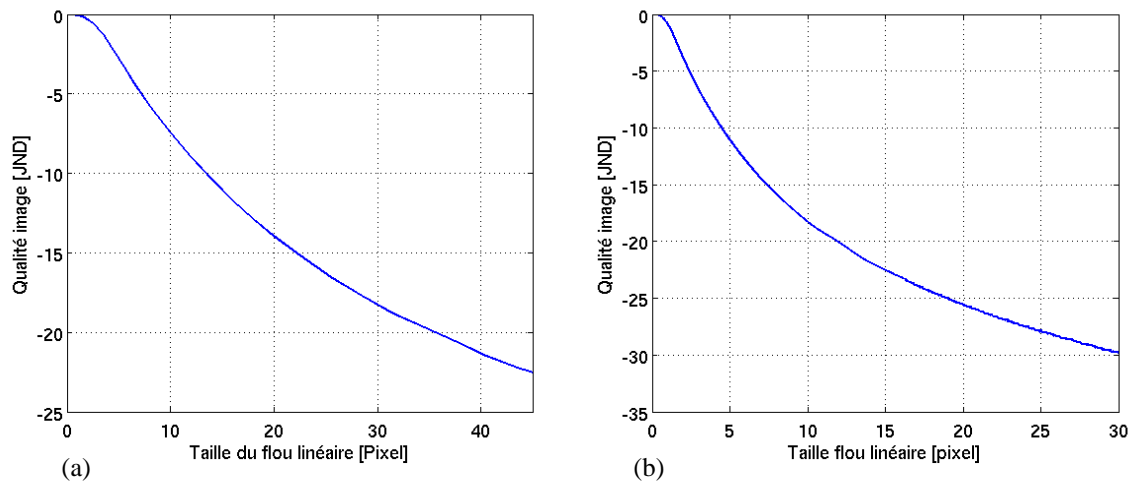


Figure 47. Résultats de la 1^{ère} expérience visuelle de la section 2.3.1.2 montrant l'évolution de la qualité d'image perçue suivant la taille du flou linéaire. (a) Pour une distance de visualisation à 2,97 m. (b) Pour une distance de visualisation à 0,97 m (1 pixel vu à l'acuité visuelle).

Qualité intrinsèque d'une image

Lorsque l'image vient d'être numérisée sur l'imageur, dans un premier temps, nous pouvons nous ramener à une image qui sera visualisée à 100 % de sa taille imageur. C'est-à-dire que chaque pixel de l'imageur correspondra à 1 point de l'afficheur. Nous appellerons cela la qualité intrinsèque de l'image. Comme vu précédemment, dans le cas où cette image est affichée, sa distance de visualisation sera la distance pour laquelle 1 point de l'image correspondra à l'acuité visuelle de l'œil humain.

En général, les images sont rarement regardées à 100 % de leur taille car leurs supports de visualisation ont une résolution plus faible [106]. Du coup, nous pouvons nous poser la question : « Quelle est leur qualité après réduction suivant le support de visualisation ? ». De manière analogue, regarder les images à une distance plus importante est la même chose que de réduire l'image d'un point de vue détail et taille de l'image. Ainsi, nous pouvons utiliser les courbes présentées figure 47 comme abaque pour transformer la qualité intrinsèque d'une image à sa qualité observée comme présentée figure 48.

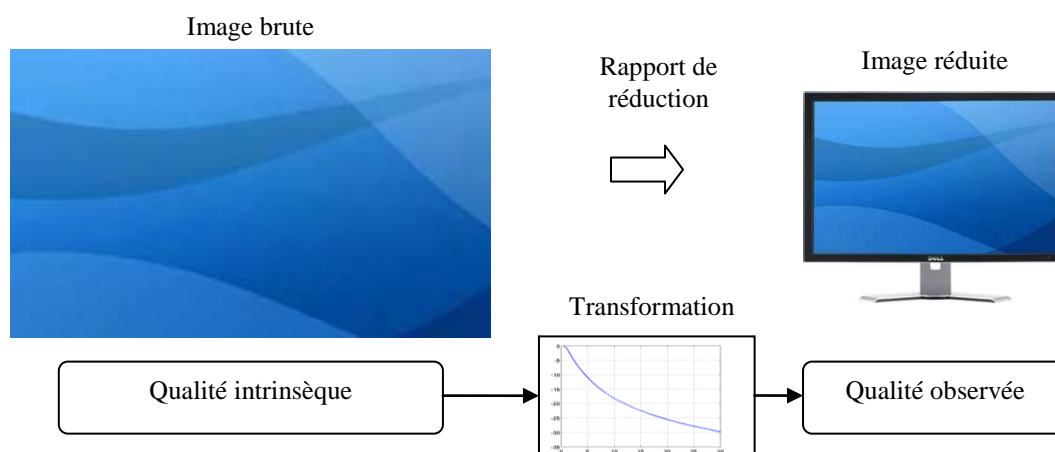


Figure 48. Ajustement de la qualité d'image après réduction de celle-ci.

En conclusion le support et la distance de visualisation sont déterminants lorsque l'on parle de qualité d'image. Suivant les applications, les besoins ne seront pas les mêmes. Pour faciliter le raisonnement dans la suite de mes travaux, de façon arbitraire, nous ne parlerons que de qualité intrinsèque. C'est-à-dire des images affichées à 100 % de leur taille pixel avec une distance de visualisation où 1 point image est vu à l'acuité visuelle soit 1 arc minute d'angle.

2.3.3 Application de la métrique au flou de bougé lors d'une prise de vue

La métrique nous permettant de quantifier le flou de bougé, nous pouvons donc nous intéresser à la variation de la qualité suivant le temps de pose pour prendre une photo. Nous avons vu dans le chapitre d'introduction que le temps de pose est fixé par la luminosité de la scène ainsi que par les paramètres de l'appareil photo. En résumé moins la scène sera lumineuse, plus le temps de pose sera long, alors plus le flou risquera d'être visible sur l'image suivant la dextérité du photographe. Suivant la qualité d'exigence de l'utilisateur ce flou visible sur l'image sera plus ou moins gênant.

2.3.3.1 Ce que souhaite l'utilisateur en terme de qualité d'image

L'exigence de qualité augmente au cours de temps

Nous avons vu que la mesure de qualité en JND permet de noter la qualité des images de manière objective. Mais en général, les utilisateurs finaux souhaitent se rattacher à des critères plus subjectifs pour qualifier une image telle que : excellente, bonne, moyenne, pauvre, ne vaut pas la peine d'être gardée. De plus comme nous l'avons vu en introduction cette exigence de qualité augmente au cours du temps rappelé figure 49.

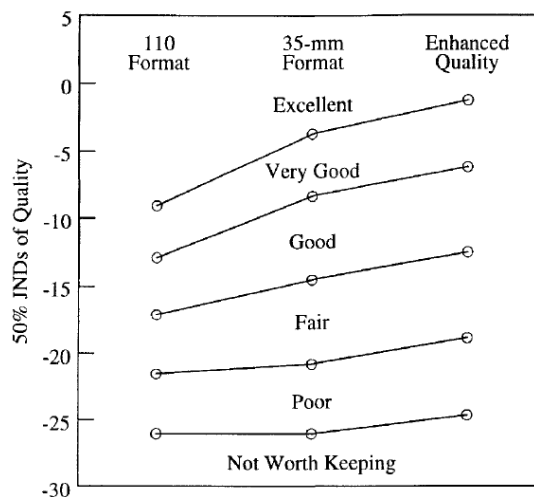


Figure 49. Évolution de l'exigence de la qualité des images au cours du temps exprimé en JND par [9]. 1979 : format 110 ; 1990 : format 35-mm ; 2002 : Qualité améliorée.

La tendance des besoins aujourd'hui

Comme les derniers résultats de l'étude de [9] présentée figure 49 date de 2002, au jour de la rédaction de ce manuscrit, 10 années ce sont écoulées. En conséquence nous avons encore « subi » une augmentation de l'exigence de qualité. Afin de mettre à jour ces données, j'ai réalisé une expérience de notation type MOS à partir d'images ayant un JND connu pour les placer dans cinq cas différents : « excellent », « Bonne », « Moyenne », « Pauvre » et « Ne vaut pas la peine d'être gardé ». Les images utilisées ont été les mêmes que celle des expériences précédentes avec 180 flous linéaires (6×30) allant de 0 à -29 JND avec un intervalle de 1 JND. Les conditions de visualisation ont aussi été les mêmes que précédemment. Les résultats de l'expérience sont présentés figure 50.

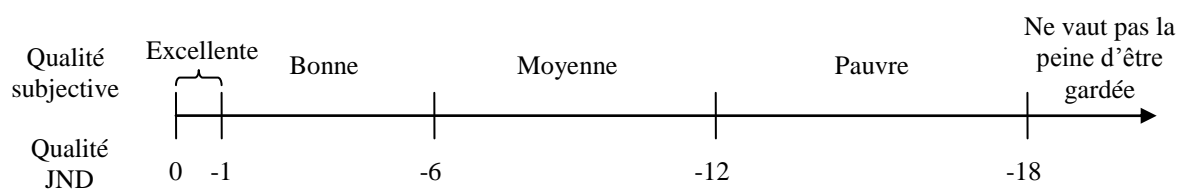







Figure 50. Qualité subjective observée suivant la valeur de qualité objective en JND.

Ces résultats montrent que l'exigence de qualité continue à augmenter. Néanmoins, les écarts entre les différents groupes sont conservés. Le tableau 6 présente des exemples d'images ayant subi un flou linéaire avec leur note subjective et objective. Ces images ont été préalablement ajustées de manière à pouvoir les imprimer au moins en 180 dpi pour les juger à 48,5 cm de distance. Nous pouvons conclure de cette expérience que l'exigence de qualité est encore de plus en plus importante.

Tableau 6. Exemple d'images ayant subi un flou de bougé linéaire avec leurs qualités objectives et subjectives respectives. Ces images ayant une définition de 180 dpi, une fois imprimée en 180 dpi ou plus, devront être visualisées à 48,5 cm afin de retrouver les résultats inscrits dans le tableau.

| <i>Image avec flou linéaire</i> | <i>Qualité objective</i> | <i>Qualité subjective</i> |
|---|--------------------------|------------------------------------|
|  | 0 JND | Excellente |
|  | -3 JND | Bonne |
|  | -9 JND | Moyenne |
|  | -15 JND | Pauvre |
|  | -21 JND | Ne vaut pas la peine d'être gardée |

2.3.3.2 Du temps de pose au bougé sur l'imageur

Comme nous l'avons vu dans le chapitre 1 figure 24, en entrée de perturbation du système de l'appareil photo, nous avons le tremblement physiologique de la main. Ainsi au final suivant les paramètres de l'appareil photo, nous avons une image floutée sur le capteur de celui-ci. Maintenant, grâce à la métrique de qualité d'image il est possible d'anticiper la qualité de l'image prise. Nous allons définir une valeur dite de bonne qualité, qui sera utilisée comme référence, dans la suite des travaux. Cette note est défini à -3 JND pour les raisons suivantes :

- L'image doit être subjectivement perçue comme étant de bonne qualité.
- L'écart entre 0 et -3 JND correspond à une limite statistique dans la définition des JND. Cette définition est la suivante : si l'on présente deux images de qualités différentes où 100 % des personnes peuvent enfin faire la différence, alors ces images sont espacées de 3 JND.

Ainsi pour la suite de ce manuscrit, lorsque nous parlerons d'une image de bonne qualité, ce sera une image ayant -3 JND comme qualité subjective.

Qualité suivant le temps de pose

Pour un temps de pose fixé, il reste 3 paramètres qui jouent sur la qualité intrinsèque de l'image ; la force du tremblement de la main, la taille des pixels de l'imager et la distance focale de l'appareil photo. La formule de transformation de l'impact du tremblement de la main est rappelée ci-dessous exprimée en déplacement pixelique :

$$\begin{bmatrix} \Delta_x \\ \Delta_y \end{bmatrix} \approx \frac{F}{T_{pix}} \begin{bmatrix} \tan(\theta_x) \\ \tan(\theta_y) \end{bmatrix} \quad (13)$$

dans laquelle Δ_x et Δ_y représentent le déplacement du bougé exprimé en pixel. F , la distance focale et T_{pix} , la taille des pixels en m. θ_x et θ_y le tremblement de la main en degré. Ainsi les deux paramètres F et T_{pix} , fixent un ratio d'amplification du tremblement physiologique de la main.

En fixant ce ratio d'amplification, deux applications directes pour la photographie sont présentées :

- L'évolution de la qualité moyenne suivant le temps d'exposition figure 51.
- Le pourcentage de bonnes images (Qualité > -3 JND) suivant le temps d'exposition figure 52.

Ces courbes ont été réalisées grâce à la métrique de qualité, en notant les images floutées à partir du modèle de tremblement physiologique. Cent différents bougés pour trois types d'appareils photo ont été simulés fixant ainsi une distance focale (équivalent 35 mm) et une taille de pixel :

- Un Smartphone (28 mm, 1,4 μm).
- Un appareil photo compact en mode portrait (100 mm, 2 μm).
- Un appareil photo bridge avec un zoom (280 mm, 2,8 μm).

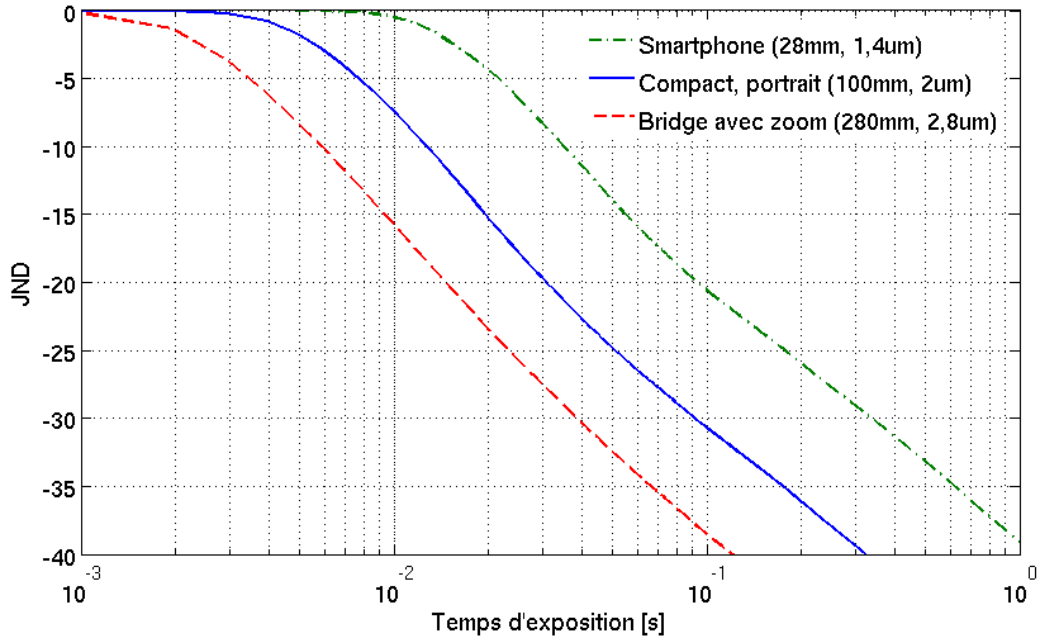


Figure 51. Évolution de la qualité de netteté en JND suivant le temps de pose pour trois types d'appareils photo.

Les résultats présentés figure 51 montrent que plus le temps de pose augmente, moins les images sont de bonne qualité en termes de netteté. Ceci est amplifié par l'augmentation de la distance focale ainsi que la réduction de la taille des pixels comme prévu par l'équation 13.

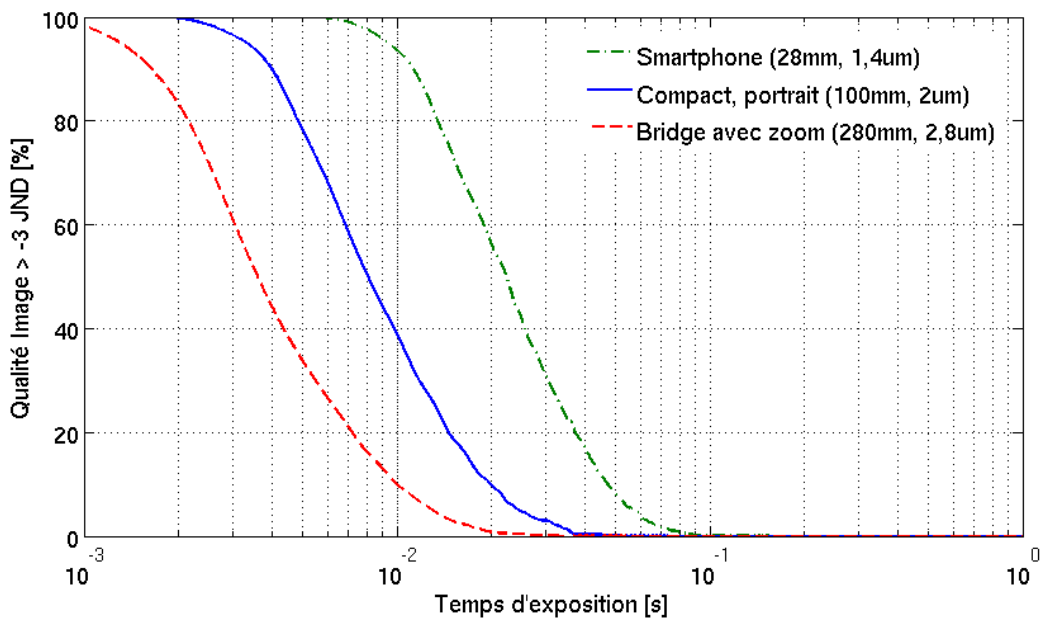


Figure 52. Evolution du pourcentage de bonne image (Qualité > -3 JND) suivant le temps d'exposition pour 3 types d'appareils photo.

En photographie, un principe de base est utilisé pour les formats 35 mm avec les appareils reflex, comme rappelé dans les travaux de [40]. Ce principe consiste à ne pas dépasser un temps de pose inversement à la distance focale. Par exemple dans le cas d'un appareil avec une distance focale de 28 mm, le temps de pose maximal que préconise ce principe est de 357 ms. Il a été observé par [40] que cette règle n'était plus tout à fait directement valable pour des appareils tels que les camera phone et les compacts. En effet, par exemple dans le cas des camera phone, il y a une réduction du temps de pose par un facteur 3. Nos simulations montrent aussi cette tendance. Si nous prenons le cas du Smartphone pour garantir une image de bonne qualité dans 90% des cas, nous préconisons un temps de pose inférieur à 112 ms. Soit une réduction du temps de pose d'un facteur 3,2 par rapport au principe de base utilisé pour les formats 35 mm. Ces résultats sont en accord avec les observations de [40] et s'expliquent dans notre cas par la petite taille des pixels des camera phone.

En conclusion, plus le temps d'exposition est important, plus en moyenne les images sont de mauvaise qualité. Ou d'un point de vue statistique, plus le temps d'exposition est important plus, il est difficile d'avoir une image de bonne qualité.

2.4 Conclusion sur la métrique de Qualité

Nous avons présenté une métrique permettant de prédire la qualité des images dans le cas du flou de bougé. Cette métrique de qualité est directement basée sur les informations de la PSF du bougé de l'appareil photo.

Grâce à l'expérience de vote forcé, nous avons pu déduire une règle de qualité d'image. Cela nous a aussi permis d'étalonner et de valider cette métrique. Notre métrique de qualité de flou de bougé donne une excellente prédiction dans le cas des mesures des flous linéaires.

La deuxième expérience par règle d'ajustement nous a permis de mesurer des flous quelconques. Ceci nous a permis d'obtenir une deuxième vérité de terrain pour le cas de flou de bougé quelconque. Ainsi cette vérité de terrain a été nécessaire pour valider notre métrique dans le cas de flous non linéaires. Les résultats montrent une bonne prédiction de la métrique pour la majorité des flous de bougé ordinaires.

Concernant les conseils sur le temps de pose afin de prendre une photo nette, l'application de la métrique aux flous photographiques a montré de bons résultats, en accord avec la réalité de terrain observée par [40].

Grâce aux deux expériences psycho-visuelles nous avons obtenu une vérité de terrains confirmant ainsi les performances de notre métrique. La métrique permet de prédire la qualité des images perçues soumises aux flous de bougé en calculant leur netteté. Cette connaissance de la netteté des images nous ouvre les portes d'une part pour quantifier les performances des algorithmes de stabilisation et d'autre part, de développer des algorithmes de stabilisation qui tiennent compte de la qualité d'image en cours de stabilisation.

Chapitre 3 *Systemes de stabilisation d'image*

La stabilisation des images pendant une prise de vue est un facteur clé pour la qualité des images, surtout lorsque le temps de pose augmente dans des conditions de scène sombre. Nous avons vu précédemment que du fait de la démocratisation des appareils photo, de la réduction de la taille des pixels, des optiques et de la masse de l'appareil, l'impact du flou s'accroît, et ce malgré une exigence de qualité d'image toujours accrue des utilisateurs.

En s'appuyant, d'une part sur les propriétés du tremblement humain, propriétés ayant donné lieu à la modélisation de son impact sur la PSF du flou de bougé, d'autre part sur une métrique de qualité d'image, j'ai développé des architectures de stabilisation répondant à ces nouveaux défis.

Dans ce chapitre, sont présentés :

- En premier lieu un état de l'art sur la stabilisation des images. Nous allons, entre autres, voir pourquoi les architectures de stabilisation ne permettent pas de garantir une netteté d'image quel que soit le type d'utilisation.
- Puis les algorithmes et architectures bas-coûts pour la stabilisation permettant d'atteindre un niveau de qualité de stabilisation.
- Enfin une architecture qui se base sur un nouveau paradigme : la segmentation du temps de pose pour garantir la qualité de stabilisation de l'image.

3.1 Etat de l'art sur la stabilisation des images

L'utilisateur final s'attend à avoir des images de bonnes qualités quel que soit l'usage qu'il fait de son appareil, des conditions d'exposition et de sa dextérité. Pour cela des

systèmes de stabilisation ont été développés pour réduire le flou induit sur l'image par les tremblements de l'utilisateur.

3.1.1 Position du problème

La chaîne de stabilisation d'image est rappelée plus en détail dans la figure 53.

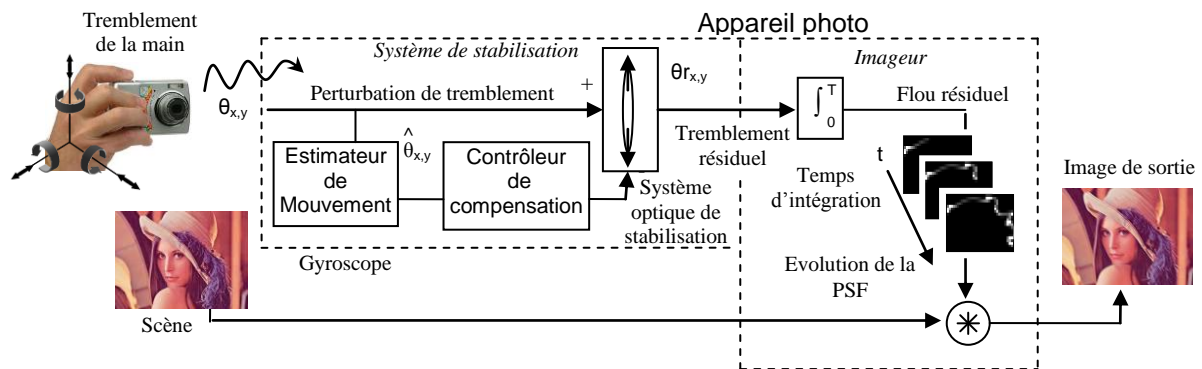


Figure 53. Chaîne de stabilisation d'image classiquement implémentée dans les appareils photos.

Les tremblements de la main sont transmis à l'appareil photo par la tenue de celui-ci comme présenté dans le chapitre 1. Nous avons vu dans le chapitre 1 que même si le tremblement impacte les 6 degrés de liberté de l'appareil photo, seuls les bougés sur les axes de tangage et de lacet auront un impact important sur la PSF de flou de l'image. Ce bougé de l'appareil photo va être estimé par un capteur de mouvement. Il s'agit d'un gyroscope destiné à mesurer de manière précise les rotations autour des axes de tangage et de lacet. Ensuite les informations de mouvement seront interprétées et mises en forme par un contrôleur de compensation, généralement un *digital signal processor* réalise cette fonction, afin de contrôler le système optique de compensation. La compensation optique permet d'aligner l'axe optique avec les rayons incidents comme illustré figure 54.

La stabilisation est réalisée soit par le déplacement de la lentille, soit par le déplacement de l'imageur. Ainsi l'image projetée sur le plan focal est stabilisée en position par rapport à celui-ci, ce qui permet d'obtenir un gain en netteté de l'image.

L'objectif de nos travaux est de proposer une chaîne de l'image simplifiée, c'est-à-dire, comportant moins de composants externes, en supprimant le gyroscope ou le système de stabilisation optique et en le remplaçant par une solution électronique. Cet objectif doit garantir un niveau élevé de qualité d'image.

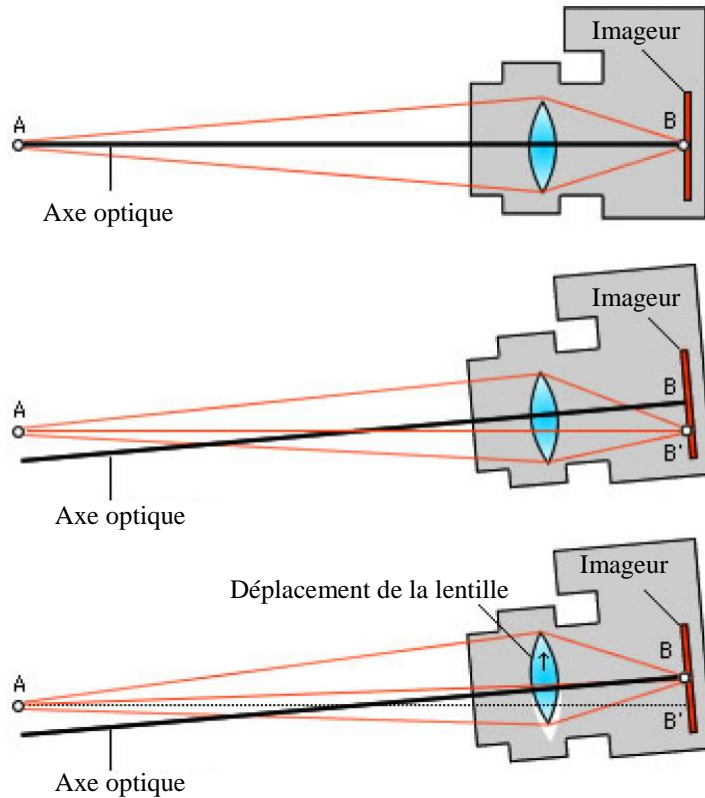


Figure 54. Illustration³ de la stabilisation d'image dans le cas de lentille mobile.

3.1.2 Solution classique de stabilisation d'images

Système de détection

Les stabilisateurs d'image classiques utilisent des gyroscopes comme système de détection du bougé. En effet, comme nous l'avons vu dans le chapitre 1, les bougés importants à détecter sont les rotations de l'appareil photo. Ainsi les capteurs gyroscopiques ont pour fonction de mesurer la rotation, et délivrent un signal de sortie proportionnel à leur rotation. Ces gyroscopes sont des capteurs de vitesse angulaire qui font partie de la famille des capteurs inertiels. Pour mémoire, le principe du gyroscope est présenté figure 55 : un objet de masse m subit des vibrations de vitesse \vec{v} . Lorsque le gyroscope tourne avec une vitesse angulaire $\vec{\omega}$, l'objet de masse m va subir un déplacement additionnel causé par la force de Coriolis $\vec{F}_{Coriolis}$ exprimée par l'équation 14 :

$$\vec{F}_{Coriolis} = -2m(\vec{v} \wedge \vec{\omega}) \quad (14)$$

³ Cette figure illustrant la stabilisation d'image dans le cas de lentille mobile est extrait de [29].

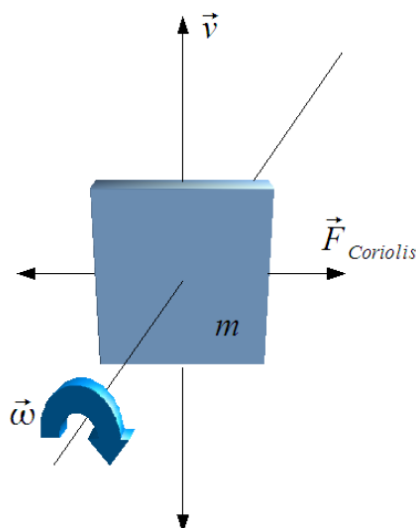


Figure 55. Illustration⁴ présentant la force de Coriolis.

La direction de ce déplacement est perpendiculaire à l'origine du déplacement et de l'axe de rotation [108]. Ainsi lorsque la vitesse et la force de Coriolis sont mesurées, il est possible d'en déduire la vitesse angulaire de l'objet. Cette vitesse angulaire est exprimée en degrés par seconde (dps).

Depuis le développement industriel des *microelectromechanical systems* (MEMS), la miniaturisation des gyroscopes permet leur intégration dans des systèmes embarqués compacts. Ainsi, de nombreuses architectures sont utilisées pour la conception des gyroscopes MEMS [109]. Par exemple pour les résonateurs, il existe des architectures qui utilisent des quartzs, des modules piézoélectriques vibrants ou encore simplement du silicium. Le faible coût et la compacité de ces MEMS, permet aujourd'hui leur présence dans les téléphones portables haut de gamme tels que les smartphones. L'état de l'art des gyroscopes MEMS est présenté dans le tableau 7. Seuls les MEMS utilisés pour la stabilisation d'image sont présentés Ces derniers ont une bonne sensibilité afin de détecter les tremblements de faibles amplitudes.

En conclusion, malgré un espace supplémentaire de quelques mm³ et un coût supérieur de quelques euros, les gyroscopes MEMS sont des composants externes très utilisés pour la stabilisation optique intégrée car ils ont une forte sensibilité, permettant de détecter les faibles tremblements de la main [14]. Cette sensibilité est de quelques millièmes de degré par seconde.

⁴ Cette figure est tirée de [29].

Tableau 7. Etat de l'art des plus petits gyroscopes MEMS disponibles sur le marché adapté à la stabilisation d'image.

| Réf. | Fabricant (composant) | Taille [mm] | Sensibilité [mdps/digit] | Dynamique [dps] |
|-------|-----------------------------|-------------|--------------------------|-----------------|
| [110] | ST (L3GD20H, 2012) | 3×3×1 | 8,75 | 245 |
| [111] | InvenSence (IDG-2020, 2011) | 3×3×0,9 | 30,5 | 250 |
| [112] | Panasonic (EWTS9P, 2011) | 4,6×3,8×0,9 | 20 | 300 |

Dans notre étude, nous allons proposer des architectures de stabilisation qui se passent du gyroscope pour supprimer les contraintes d'espace et de coût supplémentaire de ceux-ci.

Méthodes de contre réaction

Il existe deux technologies pour les systèmes de stabilisation optique : la contre réaction de type déplacement de lentille présenté figure 56.a et la contre réaction par le déplacement de l'imageur illustré figure 56.b.

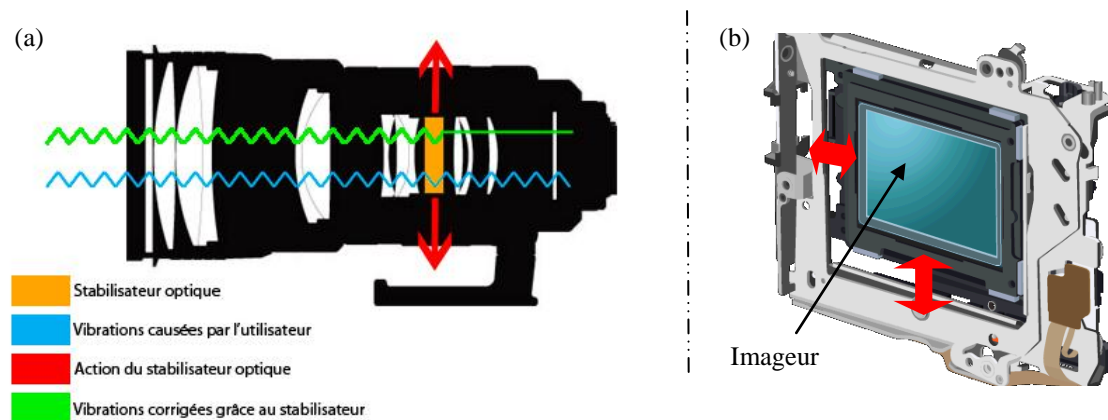


Figure 56. Système de stabilisation optique. (a) Stabilisation par déplacement de l'optique⁵. (b) Stabilisation par déplacement de l'imageur⁶.

Aujourd'hui, la majorité des constructeurs d'appareil photo proposent un système de stabilisation optique avec l'une de ces solutions dans leurs appareils photo reflex ou compact. Nikon, Canon et Sigma sont ceux qui ont le plus développé des systèmes de stabilisation dans les optiques. Alors que Sony, Pentax, Olympus, Fujifilm, Samsung, Casio et Ricoh ont développé des systèmes de stabilisation où c'est l'imageur qui est en mouvement.

Un des avantages de la stabilisation par déplacement de l'imageur est que lorsque l'appareil photo peut utiliser différents optiques, la stabilisation est réalisée indépendamment

⁵ Schéma provenant de [175].

⁶ Schéma tiré de [176].

de l'optique alors que dans l'autre solution la stabilisation doit être intégrée dans chaque optique.

Dans le cas des appareils photos pour téléphone, la contrainte de taille du module optique est une contrainte forte d'intégration. Pour un module ayant une fonction d'auto focus sa taille typique est de 8×8×6 mm [14]. C'est donc un défi pour que la partie stabilisation optique n'augmente pas trop le volume du module. Dans les travaux de [113] un module de stabilisation optique utilisant des bobines mobiles pour faire bouger la partie optique est proposé, il a pour dimensions 12×12×5 mm.

Un autre type de solution utilise des optiques liquides, *liquid lens* en anglais. Cette technologie utilise des optiques composées de liquide afin de faire varier la forme de l'optique ou sa position et ainsi permettre de réaliser une fonction de stabilisation d'image [114]. Ce type de module optique et circulaire a un diamètre de 7,75 mm et une hauteur de 2,2 mm.

Dans la même idée que la stabilisation par déplacement des imageurs mais en utilisant des technique inspiré des MEMS, des solutions miniaturisées sont développées par les chercheurs [115]. Ils arrivent à obtenir des modules de très petite taille : 8,8×8,8×0,2 mm.

Aujourd'hui, même si les modules avec stabilisation approchent les dimensions des modules avec seulement la fonction d'auto focus, leur implémentation engendre un coût et nécessite une technicité plus importante. Dans nos architectures de stabilisation, nous proposons des solutions pouvant se passer du module de stabilisation optique.

Performances des appareils existantes

La majorité des constructeurs d'appareil photo proposent des systèmes de stabilisation optique dans leurs produits. Les performances de stabilisation sont réalisées de manière empirique. Le nombre de « stops » ou *exposure value* (EV) est utilisé pour indiquer le nombre de fois dont le temps de pose peut être multiplié par deux tout en conservant une image nette. C'est-à-dire qu'un système de stabilisation ayant comme performance 3 stops ou 3 EV permettra d'obtenir des images nettes de même qualité avec un temps de pose 8 fois plus long que le même système d'acquisition en l'absence de dispositif de stabilisation. Un exemple de type de graphique est présenté figure 57 par [116].

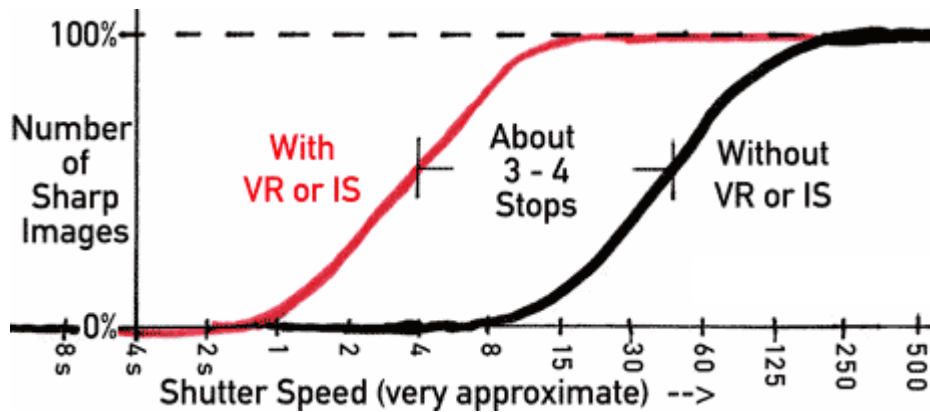


Figure 57. Gain des systèmes de stabilisation optique.

Aujourd'hui, les performances de stabilisation sont entre 2 et 4 EV suivant les appareils photo. Par exemple Sony avec son appareil photo reflex Alpha SLT-A57 [117] estime une réduction du bougé entre 2,5 et 4,5 EV suivant l'utilisation.

La combinaison d'un micro-gyroscopie en technologie MEMS et de modules de stabilisation optique a permis l'apparition des premiers téléphones équipés d'appareil photo avec des systèmes de stabilisation optique. Par exemple le smartphone Lumia 920 [118] avec son concept PureView présente des performances de stabilisation de 3 EV, soit une augmentation possible de 8 fois le temps de pose.

Un premier type de formalisation des performances de systèmes de stabilisation est proposée par [101]. Par l'utilisation d'un système mécanique il simule le bougé avec des données de tremblement humain puis mesure le déplacement résiduel avec ou sans système de stabilisation. Enfin, par un calcul de valeur efficace moyenne du bougé résiduel, il détermine le gain apporté par celui-ci.

Un deuxième type de méthode de mesure de performance est proposée par [119]. Il consiste à photographier un écran calibre composé de LED pendant que l'appareil est soumis à un bougé avec ou sans stabilisation d'image. Puis par le traitement de ces images avec différent temps de pose, le gain du stabilisateur est déterminé en utilisant l'analyse en composante principale des déplacements résiduels.

Malgré la proposition d'harmonisation des méthodes de mesure des performances des systèmes de stabilisation, il existe trois limitations pour comparer les performances de tous les appareils entre eux :

- Les signaux de tremblement injectés dans les différents appareils ne sont pas comparables.

- Les seuils de détermination des images floues ne sont pas les mêmes.
- Les méthodes de mesure du flou sont différentes.

Même si ces systèmes de stabilisation optique se généralisent de plus en plus dans les appareils photo, d'autres solutions de stabilisation existent. Elles sont présentées dans la section suivante.

3.1.3 Autres solutions de stabilisation existantes

Les chercheurs ont développés d'autres solutions dites numériques pour stabiliser les images afin de se passer des composants externes à l'imageur. Elles consistent soit à faire des opérations directement sur l'imageur en cours d'intégration, soit une fois l'image numérisée.

Stoppage de l'intégration localisée

Dans les travaux de [120], la stabilisation de l'image est réalisée de manière localisée sur l'imageur. Pour chaque pixel il y a un circuit électronique de détection du bougé. Cette détection est réalisée de manière ponctuelle suivant une période fixée. Lorsque le circuit détecte un bougé, il stop l'intégration des pixels. Ainsi lorsqu'un objet est en mouvement dans la scène, la zone de l'image contenant la scène est figée. Par extension, lorsque toute la scène sera en mouvement lors d'un flou de bougé, tous les pixels vont stopper leur intégration. Ceci revient finalement à réduire le temps de pose de l'image, ce qui n'est pas toujours souhaitable pour les scènes peu lumineuses.

Couple d'image

Une autre méthode de stabilisation d'image développée par [121] consiste à reconstruire une image nette à partir d'un couple d'image : une avec un temps d'exposition long sans utiliser de flash et une autre avec un temps d'exposition court avec l'utilisation du flash. Ainsi l'image nette acquise avec le flash permet de capturer les détails de la scène et celle sans le flash permet de prendre la luminosité ambiante. L'une des principales limitations de cette technique est que l'utilisation du flash n'est pas toujours possible.

Déconvolution

Un grand nombre d'algorithmes basés sur la déconvolution ont été développés pour réduire le flou sur les images. La déconvolution consiste à résoudre le problème inverse du floutage de l'image. Pour inverser la convolution les algorithmes de déconvolutions ont tous

besoin d'accéder à la PSF du flou : soit en connaissant déjà la PSF par supposition ou données externes, soit en cherchant la PSF du flou directement sur l'image finale. On appelle cette dernière méthode, *blind deconvolution*.

Pour les méthodes de *blind deconvolution*, il existe de nombreuses recherches sur le sujet [122, 123, 124, 125, 126, 127, 128] car l'avantage de celles-ci est de pouvoir utiliser directement les images floutées. Une comparaison d'algorithmes en termes de performance est réalisée dans les travaux de [129], il montre des résultats visuels à partir d'un flou de bougé linéaire ainsi que les valeurs du bruit des images défloutées. Ces méthodes présentent l'inconvénient d'être gourmandes en puissance de calcul et l'erreur sur l'estimation de la PSF se cumule à l'erreur liée au calcul de l'inversion du floutage.

Pour les autres méthodes de déconvolution il existe par exemple celles qui utilisent un imageur hybride [130, 131, 132]. C'est-à-dire que cette méthode utilise deux imageurs (ou deux zones de l'imageur) : un permettant de prendre simplement la scène pour réaliser l'image finale, l'autre pour estimer le mouvement du bougé afin de reconstruire la PSF.

Une autre méthode de déconvolution est celle utilisant un couple d'image comme dans les travaux de [133, 134, 135, 136]. Cette technique consiste à prendre deux images consécutives afin de prendre la même scène : d'une part une image avec un temps d'exposition long (donc floutée) et d'autre part une autre image avec un temps d'exposition rapide. Puis un post traitement est réalisé avec ces deux images pour retrouver efficacement et fidèlement la PSF de bougé de la première image. Ainsi cette seconde image participe à réduire la complexité de la reconstruction de la PSF permettant de faire ces traitements directement sur l'appareil photo numérique.

Globalement, même si certains algorithmes montrent des résultats corrects, ces bons résultats sont obtenus dans les cas de faibles bougés sans forme trop complexe, souvent de quelques dizaines de pixels maximum [129], donc se limite aux temps d'exposition courts correspondant à des bougés linéaires soit d'environ 1/10 de seconde.

Capture multiple

Dans les travaux de [137], un algorithme de capture multiple est réalisé de manière locale au niveau pixel de l'imageur. Ainsi lorsqu'il n'y pas de bougé détecté dans la zone de l'image, il y a une simple accumulation des différentes captures. Lorsqu'un bougé est détecté dans la zone, une opération d'estimation de mouvement local est réalisé afin de recombinaison les séquences successives en tenant compte du mouvement sur la partie de l'image impactée.

Cet algorithme est de type récursif afin de limiter la taille de la mémoire due à la capture multiple. Une autre méthode de capture multiple généralisée à toute l'image est proposée par [138]. Elle consiste à découper le temps de pose long, qui génère des images floutées, en plusieurs images plus nettes grâce à un temps de pose plus court. La décomposition du temps de pose est réalisée par 4 images de temps d'intégration plus courts. Ensuite le décalage entre les 4 images du fait du bougé est détecté afin de recombinaison les images entre elles. Cette méthode permet aussi de réduire le bruit dans l'image finale par rapport à la solution consistant à prendre une seule image avec un temps de pose court compensée par un fort ISO.

Plusieurs fabricants d'appareil photo ont intégré cette fonction de stabilisation particulièrement utilisée dans les scènes nocturnes. Par exemple, Canon dans son PowerShot S100 [139] et Nikon dans Coolpix [140] proposent cette stabilisation pour les prises de vue de la nuit. Le problème de cette approche est qu'elle se limite seulement à quelques images, donc son efficacité s'en trouve limitée par le temps de pose.

Stabilisation sur l'imageur

Grâce à la propriété des imageurs CCD (*charge coupled device*) de pouvoir transférer les charges de proche en proche sur l'imageur, les travaux de [141] réalisent la fonction de stabilisation directement sur l'imageur. Cet imageur particulier est appelé *orthogonal transfert CCD* (OTCCD). Ce capteur permet de répondre aux exigences de l'observation spatiale, permettant de s'adapter aux turbulences atmosphériques qui entraînent un flou de bougé sur l'imageur pour des temps de pose de plusieurs heures. Dans le même type d'idée mais cette fois-ci pour des imageurs CMOS, Advasense [142] propose un *feedback controlled pixel* (FPC) afin de stabiliser l'image pour la photographie. Cette technique prévoit de charger et décharger les pixels en cours d'intégration pour faire suivre les charges sur leur nouveau lieu d'intégration suite au bougé.

Conclusion

En conclusion, toutes ces méthodes de stabilisation présentées dans cet état de l'art s'intéressent à la réduction du flou de bougé mais ne permettent pas de garantir une qualité d'image stabilisée car elles ne prennent pas en compte un critère de qualité, par exemple basé sur la qualité voulue par l'observateur. Par la suite, nous proposerons des architectures qui permettent de remettre au centre la qualité exigée par l'observateur grâce à notre critère de qualité d'image. Ces architectures s'appuieront sur la capacité des imageurs rapides à générer

des séquences d'images à des cadences bien supérieures à la cadence vidéo (25 à 30 i/s). Un état de l'art sur ces imageurs rapides est présenté dans la section suivante.

3.1.4 Imageur rapide

Le premier imageur rapide est apparu dans les années 80', il était de technologie CCD avec une fréquence de 185 images par seconde (ips) pour une résolution de 512×512 pixels [143]. Le développement des imageurs CMOS a permis d'améliorer la fréquence image grâce aux différents types d'architectures pour la lecture des pixels et notamment grâce à la parallélisation de la lecture de ceux-ci [144, 145]. Pour plus de détails sur ces techniques de parallélisations, un résumé est présenté dans les travaux de [146]. Dans ces deux technologies CCD et CMOS, l'obturateur qui permet de prendre les images les unes à la suite des autres est réalisé de manière électronique par l'imageur. Potentiellement comme ils n'utilisent pas de système mécanique, la vitesse d'obturation peut être de quelques nanosecondes soit une fréquence de cadence d'image d'environ 1 GHz [147]. Néanmoins, pour ces imageurs, leur cadence de fonctionnement se trouvera limitée par les circuits d'entrées sorties. Le tableau 8 présente un état de l'art pour les imageurs et les cameras sur le marché ayant des résolutions compatibles avec la photographie.

Tableau 8. Etat de l'art des imageurs et camera les plus rapides ayant une résolution supérieur à 1 méga pixel.

| <i>Réf.</i> | <i>Fabriqueur - Modèle Equipe de recherche</i> | <i>Résolution</i> | <i>Fréquence image [i/s] ; éch./s ; bit/s</i> |
|-------------|--|--------------------|---|
| [148] | Cypress – VITA 25K, 2010 (imageur) | 5120×5120 (10 bit) | 53 i/s ; 1,4 Gé/s ; 14 Gb/s |
| [149] | Vision Research – Phantom v1610, 2011 (camera) | 1280×800 (12 bit) | 16000 i/s ; 16 Gé/s ; 196 Gb/s |
| [150] | Photron – Fastcam SA2, 2009 (camera) | 2048×2048 (12 bit) | 1080 i/s ; 4,5 Gé/s ; 54,4 Gb/s |
| [151] | IDT – Y7-S3, 2012 (camera) | 1920×1080 (10 bit) | 9000 i/s ; 18,7 Gé/s ; 186 Gb/s |
| [152] | Sony, 2011 (imageur) | 8192×2160 (12 bit) | 120 i/s ; 2,1 Gé/s ; 34,8 Gb/s |
| [153] | NHK Science & Tech R. Lab, 2012 (imageur) | 7680×4320 (16 bit) | 120 i/s ; 4 Gé/s ; 51,2 Gb/s |

L'état de l'art montre qu'aujourd'hui il est possible de réaliser des imageurs rapides qui peuvent générer des images à quelques centaines d'image par seconde.

Ces imageurs rapides ont donc une cadence image limitée par la bande passante de leurs sorties. Pour éviter ce goulet d'étranglement, la solution est de garder les images en interne dans l'imageur grâce à des mémoires puis, de les extraire à la vitesse des entrées sorties. Cette nouvelle architecture d'imageur arrive à augmenter de plus d'un facteur 100 la fréquence des images par rapport à ceux présentés ci-dessus. Le problème des systèmes visant à obtenir des

images ultra-rapides, est qu'ils ont aujourd'hui des formats limités à quelques dizaines voire centaines de pixels avec des méthodes difficilement intégrables ni envisageables dans un système embarqué. C'est pourquoi nous ne présenterons pas ce type d'imageur. Néanmoins le lecteur pourra se référer aux travaux de [147] présentant un état de l'art sur le sujet.

Nos algorithmes de stabilisation d'image, présentés dans la section suivante, s'appuieront sur les imageurs rapides.

3.2 Algorithmes et architectures économiques pour la stabilisation

Plusieurs architectures ont été développées afin de démontrer leurs fonctionnalités et de comparer leurs performances. Partant de nos connaissances sur le tremblement humain, nous avons envisagé des architectures sans gyroscope afin d'utiliser l'imageur comme un capteur de bougé.

Dans cette perspective, en premier lieu, j'ai conçu une architecture sans gyroscope mais conservant encore un système de stabilisation optique.

Puis j'en ai décliné une architecture gardant seulement l'imageur. Dans ce cas l'imageur est autonome et réalise aussi bien l'estimation de mouvement que la stabilisation de l'image. De ce fait, cette architecture sans gyroscope ni stabilisation optique est complètement intégrée.

Enfin, j'ai étudié encore une autre solution d'architecture : une architecture sans système de stabilisation optique mais avec gyroscope.

Les systèmes de stabilisation sans gyroscope ont besoins de détecter par eux même les tremblements ; les systèmes de stabilisation sans optique de compensation ont besoin de stabiliser par eux même l'image. Ces deux fonctions sont réalisées d'une part par un estimateur de mouvement et d'autre part par une sommation cohérente et sont présentées dans les paragraphes suivants.

3.2.1 Nouvelle approche pour l'estimation de mouvement sans gyroscope

3.2.1.1 Méthodes d'estimation de mouvement existantes basées sur l'analyse d'image

Le but de notre approche est de simplifier la chaîne du stabilisateur d'image en supprimant le gyroscope qui mesure le bougé de l'appareil photo. Pour réaliser cette fonction, nous avons besoin d'analyser l'image pendant son intégration afin d'en retirer les informations de mouvement global dû au bougé. Un couple d'image doit être utilisé afin d'estimer le mouvement relatif de l'une par rapport à l'autre. Pour réaliser cette estimation, il existe deux principales approches : une approche de bas niveau utilisant l'image brute ; une approche haut niveau utilisant les informations des éléments de l'image. Les approches bas niveau sont résumées ci-dessous.

Méthode fréquentielle

Des méthodes fréquentielles ont été développées pour estimer le mouvement ainsi que les rotations et changement de taille entre deux images. Ces méthodes fréquentielles utilisent les informations de corrélation de la phase [154, 155] des images après transformée de Fourier de celle-ci. Ces méthodes étant basées sur les transformations de Fourier sont encore pour l'instant trop complexes en termes de capacité de calcul pour être embarquées dans des systèmes tel que des appareils photos. Nous n'avons donc pas retenu cette méthode pour réaliser l'estimation de mouvement.

Méthode différentielle

Les méthodes différentielles, développé par [156], s'appuient sur les dérivées spatiales et temporelles pour estimer le mouvement. Elles ont été l'objet de nombreuses améliorations et d'extension [157, 158], leurs performances sont résumées et comparées dans [159]. Elles reposent sur la propriété de conservation de l'illumination des points de l'image, ce qui est le cas lorsque les deux images sont prises l'une à la suite de l'autre avec le même temps de pose. Malgré une complexité importante, due au fait que ces méthodes mettent en jeux des dérivées ainsi que des inversions de matrices, nous avons retenu cette méthode comme référence d'estimateur de mouvement car elles permettent d'atteindre une précision sous-pixélique et sont réputées pour leur robustesse au bruit [159]. Dans les algorithmes proposés,

l'implémentation pyramidale de Lucas Kanade [160, 161] a été utilisée comme estimateur de référence pour ses performances. L'approche pyramidale permet aussi de ne pas être limité par la taille des bougés entre les images [161].

Méthode de mise en correspondance

Il existe une littérature importante sur les méthodes d'estimation de mouvement par mise en correspondance de zone de l'image [162, 163, 164, 165]. Cette méthode consiste à prendre une zone de la première image appelée motif de référence et de retrouver ce motif sur l'image suivante par une méthode apparentée à la corrélation. La distance entre les deux motifs est ainsi la distance de déplacement entre les deux images. Suivant le type de méthodes de mise en correspondance, leurs complexités peuvent être réduites afin d'être embarquées dans des systèmes [165]. Principalement pour cette raison, nous avons choisi d'implémenter cette méthode en vue d'une potentielle intégration sur puce.

L'autre approche utilisée pour faire de l'estimation de mouvement est la méthode d'estimation par zone d'intérêt qui est une approche haut niveau.

Méthode d'estimation par zone d'intérêt

La méthode d'estimation par zone d'intérêt pour l'estimation de mouvement est une combinaison de l'approche bas niveau avec des éléments remarquables de la scène comme par exemple des points sur-brillants, des coins, des contours [166]... Ainsi, une fois que ces éléments sont détectés sur la scène, ils sont suivis afin de connaître le déplacement dans l'image. Du fait de la demande d'un prétraitement des données, donc de l'ajout d'une surcouche, cette approche est plus complexe. De plus, pour mesurer un bougé majoritaire de l'image il faut prendre un nombre important de points pour s'assurer de détecter le mouvement global de la scène [167] ce qui augmente encore la complexité. Pour ces raisons de complexité, nous avons écarté les méthodes d'estimation par zone d'intérêt pour réaliser l'estimation de mouvement dans nos algorithmes de stabilisation.

3.2.1.2 Apport du modèle de tremblement pour l'estimation de mouvement

La modélisation des tremblements de la main appliquée à la photographie nous apporte des connaissances précises en vue de son estimation. L'étude du modèle nous donne des informations en termes d'accélération, de vitesse, de déplacement et de prédictibilité. Ci-dessous, nous allons détailler ces propriétés en termes de déplacement et de prédictibilité.

Longueur des déplacements

Suivant la fréquence de fonctionnement de l'estimateur, les besoins en termes de longueur de mouvement à estimer ne sont pas les mêmes. En effet, plus la fréquence d'estimation est rapide, plus le déplacement à détecter est petit et inversement. Pour démontrer cette affirmation, les tailles des déplacements en 2 dimensions sur le plan imageur suivant la fréquence d'estimation ont été calculées. Les résultats sont présentés figure 58 sous forme de densité de probabilité des déplacements suivant leur taille. Dans le but de rendre ces informations indépendantes de la distance focale ainsi que de la taille des pixels, l'équation 8 a été normalisée par la distance focale ainsi que la taille des pixels.

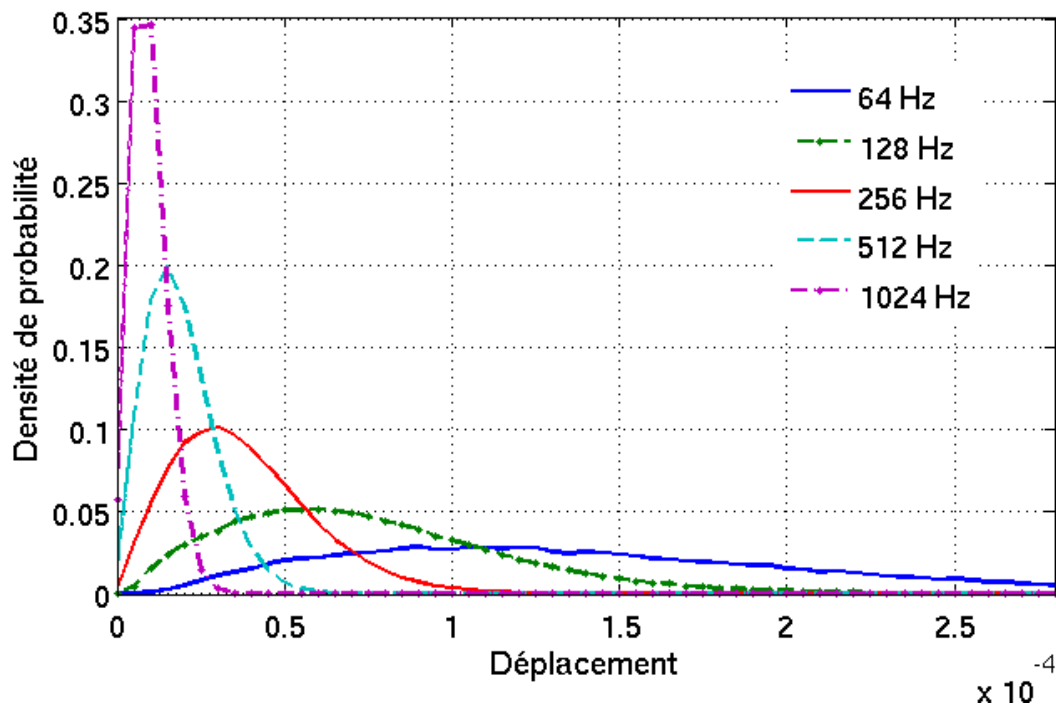


Figure 58. Densité de probabilité de la taille des déplacements normalisés, vis-à-vis de la distance focale et de la taille des pixels, suivant la fréquence d'estimations de mouvement.

Plus la fréquence d'estimation est importante, plus les tailles des déplacements à estimer sont petites et concentrées autour d'une même valeur. Inversement pour des fréquences d'estimation faible, le déplacement à estimer aura plus de chance d'être grand. Ainsi, si nous prenons comme critère de pouvoir estimer 99% des déplacements pour chaque fréquence de fonctionnement, nous pouvons borner le système en terme d'amplitude du bougé sur le plan imageur. Ces informations sur l'amplitude du bougé nous ont permis de dimensionner les

estimateurs de mouvement. Spécialement l'estimateur de mise en correspondance, détaillé par la suite, qui a une amplitude d'estimation fixée par sa zone de recherche.

Prédiction

Plusieurs études ont été menées pour étudier la prédictibilité du bougé bidimensionnel sur le capteur. Une première par l'étude de similarité des vitesses aux instants n et $n+k$. Nous avons observé que plus k est petit, plus il y a une similarité importante sur la vitesse du déplacement. L'autre étude menée a porté sur l'entropie de la taille des déplacements suivant la fréquence d'estimation de celui-ci. Plus l'entropie d'un signal est faible, plus le signal est prédictif.

La formule mathématique de l'entropie est rappelée par l'équation 1.

$$E(X) = \sum_{i=1}^n p(x_i) \log\left(\frac{1}{p(x_i)}\right) \quad (15)$$

où $E(X)$ est l'entropie du signal X , $p(x_i)$ la probabilité d'apparition du signal pour la valeur x_i . L'entropie de la taille de déplacement du bougé sur le plan imageur suivant la fréquence d'estimation du mouvement est présentée figure 59.

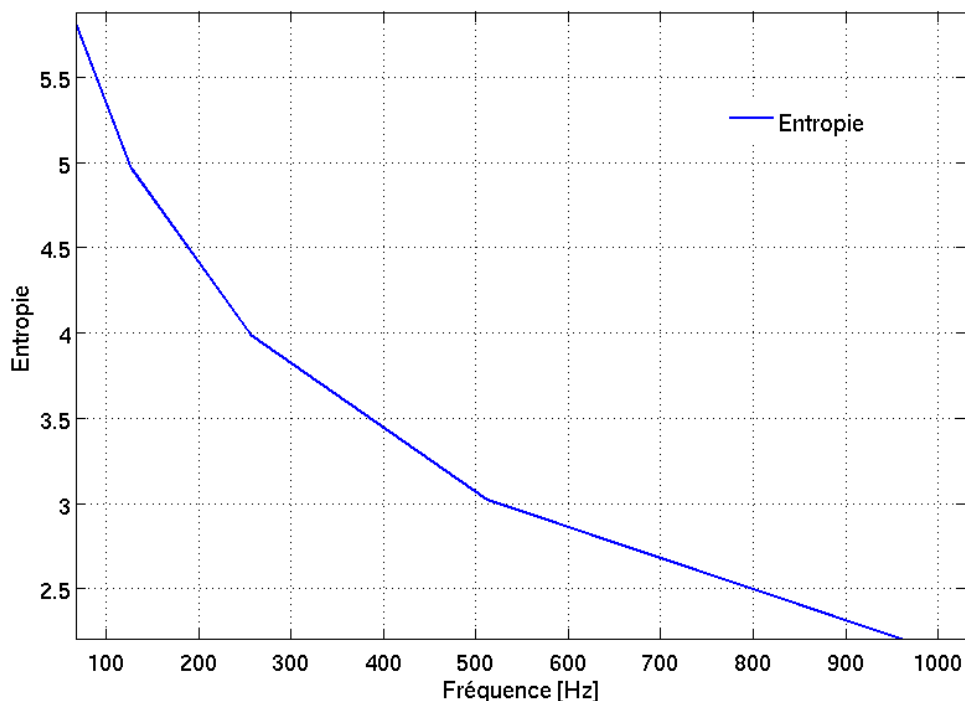


Figure 59. Entropie de la taille de déplacement sur le plan imageur suivant la fréquence d'estimation de celui-ci.

Les résultats montrent que plus la fréquence d'estimation est importante, plus l'entropie du bougé est petite, donc plus le bougé est prédictible concernant la taille de déplacement et inversement. En d'autres termes, plus la fréquence d'estimation sera importante, plus un système prédictif sera efficace. En pratique, pour un estimateur de mouvement associé à un prédicteur, plus la fréquence du fonctionnement sera importante meilleure sera la prédiction du mouvement.

En conclusion, ces études sur le bougé nous permettent pour une utilisation donnée de borner le système afin de dimensionner les architectures au plus proche des besoins réels. Elles nous permettent aussi d'envisager un système de prédiction afin de compenser les retards induits par l'estimation de mouvement ou encore le système de contre réaction. Ces connaissances du bougé nous permettent de développer des architectures de stabilisation d'image performantes et optimisées en vue d'une intégration future.

3.2.1.3 Description du système d'estimation de mouvement sans gyroscope

Principe de base

L'estimation de mouvement sans gyroscope consiste à reconstruire le déplacement initial du bougé contenu dans l'image. Pour réaliser cette fonction, l'image est décomposée dans le temps en trames ayant des temps d'intégration plus courts que le temps de pose global. Nous parlons de trame et non d'image car chaque trame contient une partie de la luminosité de la scène. La somme de ces trames est égale à l'image intégrée à 100%. A partir des trames, une estimation de mouvement est réalisée entre elles comme présenté figure 60.

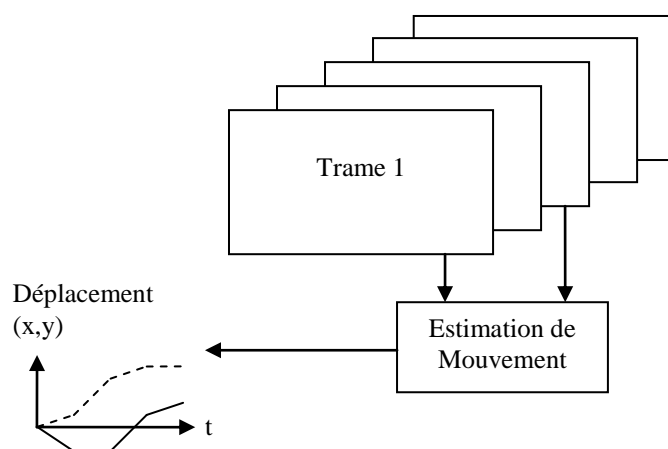


Figure 60. Schéma fonctionnel de l'estimation de mouvement sans gyroscope.

Ainsi par l'estimation de mouvement entre les différentes trames, les déplacements sur le plan image dus aux tremblements de la main sont retrouvés. Ces informations de déplacement du bougé peuvent ainsi remplacer les données du gyroscope pour alimenter le contrôleur de stabilisation. Pour retrouver le déplacement, l'estimateur de mouvement utilise la première trame comme référence car ceci permet d'avoir une estimation absolue. Il est possible de faire une estimation relative en utilisant la trame n et $n+1$ pour estimer le mouvement. Le problème de cette estimation est qu'elle cumule les erreurs d'estimation comme le démontre [167]. Nous avons pu aussi observer cette dérive du système en simulation.

Comme expliqué précédemment, pour réaliser l'estimation de mouvement, nous avons utilisé la méthode différentielle pyramidale comme référence d'estimateur ainsi que l'estimateur de mouvement par mise en correspondance pour une approche plus embarquée. L'estimateur de mouvement par mise en correspondance implémenté est présenté en détail ci-dessous.

Estimation de mouvement par mise en correspondance

L'estimation de mouvement par mise en correspondance, *block-matching* en anglais, consiste à définir une zone de référence dans la première image ou trame dans notre cas et de retrouver cette zone dans l'image suivante. Cette recherche de zone de l'image peut être réalisée par différentes méthodes de pseudo corrélation détaillées dans [165]. Dans ces travaux nous avons choisi la méthode « somme des différences au carré », *sum square difference* (SSD) en anglais pour sa faible complexité et sa robustesse au bruit dans l'image [168]. Comme nous utilisons des trames, nous travaillons avec des faibles luminances et donc des trames plus sensible au bruit d'où un besoin d'un estimateur robuste au bruit. La taille de la zone de mise en correspondance va avoir un impact sur la précision de l'estimation de mouvement. Plus la taille de la fenêtre sera grande, plus les résultats de l'estimation seront précis et robustes au bruit [169]. Par contre utiliser une grande fenêtre aura un coût de calcul plus important. Pour trouver le déplacement, la fenêtre de mise en correspondance va se déplacer dans toute la zone de recherche pour trouver où la correspondance sera la meilleure. Plus la zone de recherche est grande, plus le bougé détectable peut être grand. Avec une seule fenêtre de recherche, il n'est pas possible d'avoir une résolution de recherche du déplacement sous pixel car l'opération de mise en correspondance se fait à l'échelle du pas pixelique. Ainsi pour avoir une résolution sous pixel nous avons défini un estimateur élémentaire que nous répétons suivant une matrice dans l'image comme présenté figure 61. En plus d'avoir une

résolution sous pixel par moyennage des résultats des estimateurs élémentaires, nous réalisons une estimation de mouvement de façon homogène sur l'image.

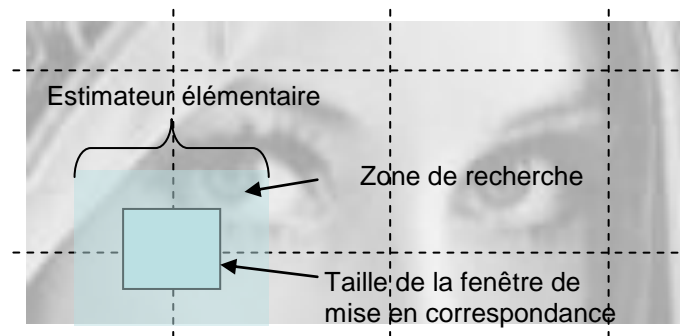


Figure 61. Implémentation de l'estimateur de mouvement par mise en correspondance.

Pour améliorer la précision de l'estimation globale, une fois que la SSD est réalisée sur toute la zone de recherche, nous regardons si la réponse de chaque estimateur élémentaire présente un pic distinct. Si la réponse est trop plate, en d'autres termes qu'elle n'a pas de pic de corrélation, l'estimateur élémentaire est disqualifié car jugé non pertinent. Ce cas arrive lorsque l'estimateur élémentaire se trouve dans une zone uniforme. Ainsi chaque résultat d'estimateur ayant un pic retenu participe au moyennage pour estimer le bougé global dans l'image.

Le choix de la taille des fenêtres, de la zone de recherche et du nombre d'estimateur élémentaire a été déterminé grâce aux entrées du modèle de tremblement physiologique. Nous avons choisi les différents paramètres afin d'avoir un système réaliste permettant de démontrer l'approche de ce système. Pour la taille des fenêtres, nous avons cherché à obtenir des performances similaires en termes de précision à l'estimation de mouvement par méthode différentielle. Ainsi nous avons fixé la taille de la fenêtre à 10×10 pixels, une taille de fenêtre dans la gamme des valeurs couramment utilisées pour faire de l'estimation de mouvement [170] et avec 25 estimateurs élémentaires réparties dans l'image. Comme la taille de la zone de recherche est très dépendante de la fréquence d'estimation, de la distance focale, de la taille des pixels et de la force du bougé, elle a été déterminée grâce au modèle de tremblement. Cette zone de recherche de ± 5 pixels a été déterminée afin de couvrir les différents cas de nos simulations.

Prédicteur

Comme nous avons vu précédemment grâce à l'étude de notre modèle de génération de bougé plus la fréquence de fonctionnement du système est important, plus le bougé est prédictif. Pour tirer parti de cet avantage, nous avons développé des prédicteurs afin d'augmenter les performances de l'estimation du bougé. Ils permettent de rattraper le retard dû à l'intégration des trames avant leurs estimations. Leurs principes est présenté figure 62.

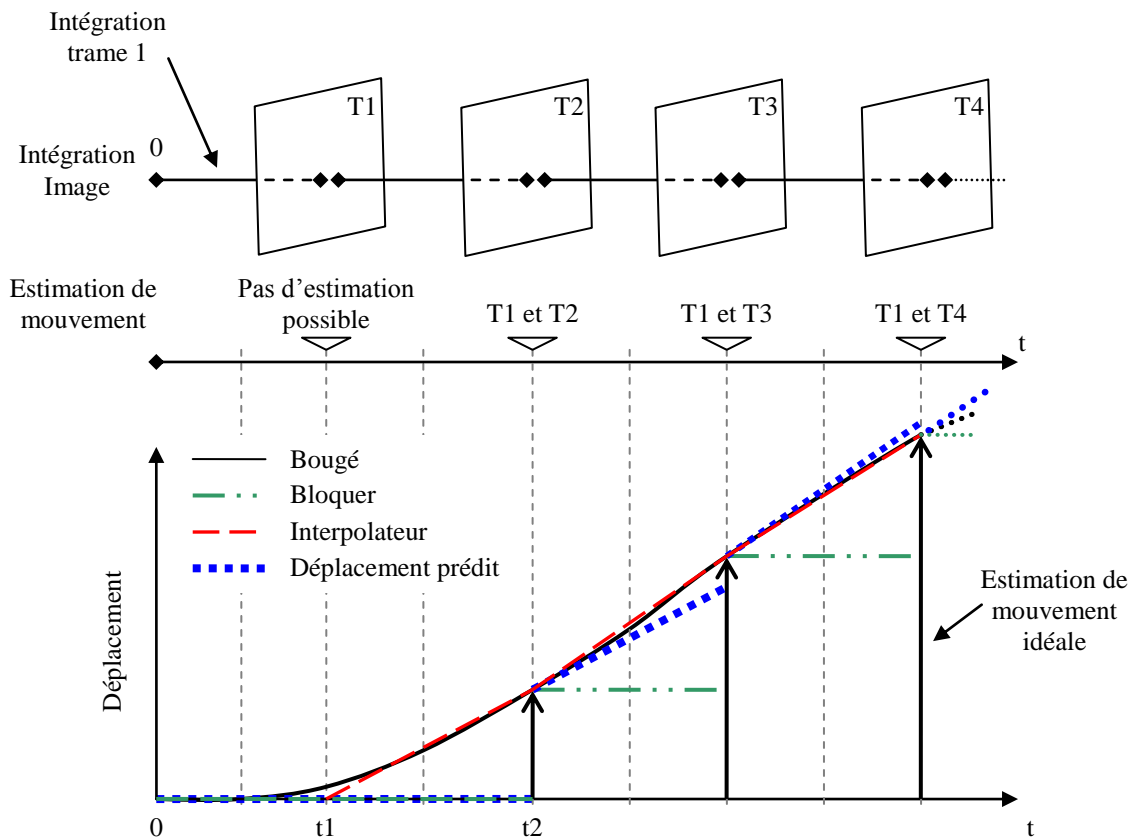


Figure 62. Principe de prédiction pour estimer de manière efficace le déplacement de bougé.

Au démarrage du temps de pose, la première trame commence à s'intégrer. De ce fait à la fin de l'intégration de la trame 1 (T1), il n'est pas encore possible de réaliser une estimation de mouvement car il manque une référence à T1. Ainsi T1 devient la référence pour toutes les futures trames. Lorsque T2 est enfin intégré, si nous avons une estimation idéale, la position du bougé est trouvée. Jusque-là il n'est pas possible de réaliser une prédiction donc le déplacement prédit est toujours à 0. Si le système utilise seulement un bloqueur, la position prédite sera la même position que celle estimée et ainsi la mise à jour du déplacement se fera lors d'une nouvelle estimation. Elle est présentée figure 62 par la courbe du bloqueur. Par

contre si le système utilise un interpolateur entre la position de t1 et t2 pour reconstruire de manière plus fidèle le passé, la prédiction du futur sera meilleure. Ainsi l'estimation du déplacement prédit sera plus proche du bougé initial. Dans cet exemple figure 62, l'interpolateur et le prédicteur sont linéaires.

Pour les algorithmes de stabilisation, nous avons utilisé deux types d'interpolation. L'interpolation linéaire et l'interpolation cubique. Pour les prédicteurs, nous en avons développé un au 1^{er} ordre et l'autre au 2^{ème} ordre. Ils sont présentés ci-dessous.

Si nous développons la prédiction au deuxième ordre avec la série de Taylor nous avons :

$$x_{k+1} = x_k + \left. \frac{\partial x}{\partial t} \right|_{x_k} dt + \left. \frac{\partial^2 x}{\partial t^2} \right|_{x_k} dt^2 \quad (16)$$

où x_{k+1} est la position de déplacement prédit, x_k la position du déplacement à l'instant présent avec :

$$v_k = \left. \frac{\partial x}{\partial t} \right|_{x_k} \text{ et } a_k = \left. \frac{\partial^2 x}{\partial t^2} \right|_{x_k} \quad (17)$$

Où v_k et a_k sont respectivement la vitesse et l'accélération du déplacement à l'instant présent. Ainsi le prédicteur au 1^{er} ordre est basé sur l'équation suivante :

$$x_{k+1} = x_k + v_k dt \quad (18)$$

Le prédicteur au 2^{ème} ordre est basé sur l'équation suivante :

$$x_{k+1} = x_k + v_k dt + a_k dt^2 \quad (19)$$

Les résultats avec ou sans prédicteur seront présentés par la suite et montreront différents gains dans le système de stabilisation d'image suivant le type de prédicteur utilisé.

3.2.2 Nouvelle approche pour la réduction du flou

Description du système de sommation cohérente

La sommation cohérente s'intéresse à estimer la qualité du flou de chaque trame par l'intermédiaire des PSF trames pour reconstituer l'image de telle façon que le flou de l'image finale soit minimum. Pour cela, comme pour le temps d'intégration de l'image qui est découpée dans le temps en trames, l'évolution du bougé est découpé dans le temps pour construire des PSF trames de chaque segment du bougé. Ainsi au final par construction, chaque PSF trame correspond au flou de chaque trame. Pour réaliser la sommation cohérente de manière efficace présentée figure 63, le premier bloc estimateur de mouvement estime le

meilleur déplacement directement sur les PSF trames. Ce déplacement est le meilleur lorsque du point de vue de la perception visuelle, l'addition de ces PSF donne la meilleure netteté. Pour arriver à l'estimation de la meilleure netteté, nous utilisons notre métrique de qualité comme critère de choix analogue au maximum de corrélation. Ces informations de déplacement sont ensuite transmises au translateur pour effectuer le déplacement de la trame avant de l'ajouter aux autres trames stabilisées. La somme finale de trames translatées donne une image stabilisée.

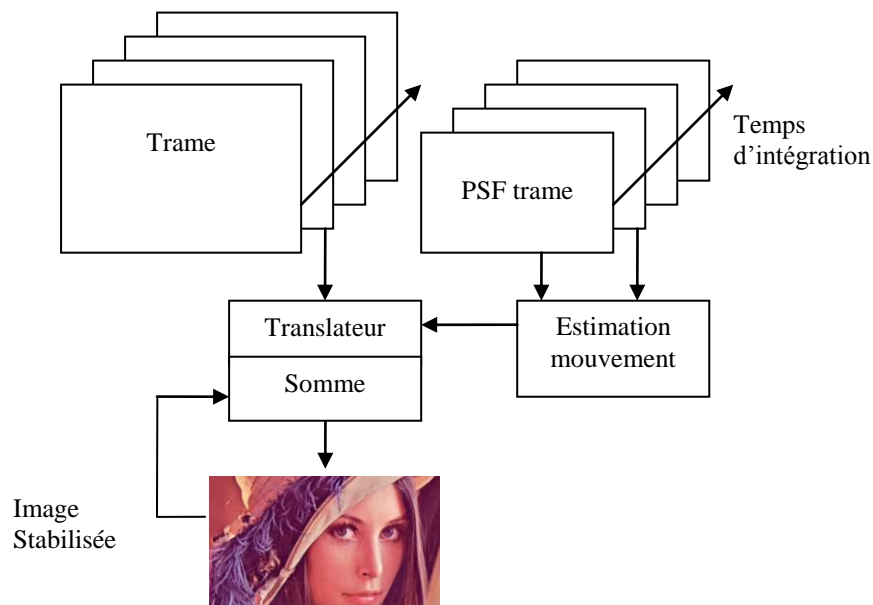


Figure 63. Schéma de principe présentant la sommation cohérente des trames.

Dans la suite nous allons présenter les différentes architectures découlant des différentes combinaisons possibles de la suppression du gyroscope, de l'optique de stabilisation ou des deux. Nous évaluerons aussi les performances de ces différentes combinaisons possibles.

3.2.3 Impact des nouvelles solutions sur l'architecture de stabilisation

Afin de comparer les différentes architectures entre elles, nous avons simulé nos architectures dans 3 cas usuels :

- Un appareil photo de smartphone avec un temps de pose de 1/10 seconde ayant une distance focale de 28 mm (éq. 35 mm) et des pixels de 1,4 μm .
- Toujours le même appareil photo de smartphone mais avec un temps de pose de 1 seconde.

- Un appareil compact avec un temps de pose de 1/10 seconde réglé à une distance focale de 280 mm (éq. 35 mm) avec des pixels de 2,8 μm .

Dans chaque cas, grâce à notre modèle de tremblement physiologique les même 100 bougés ont été générés et fournis comme références d'entrée aux différentes architectures. Les performances de ces architectures sont présentées directement par la qualité moyennée des 100 images stabilisées exprimé en JND grâce à notre métrique de qualité. Donc comme nous avons vu précédemment, pour générer les trames avec une cadence important, les imageurs rapides sont utilisées. L'évaluation de la cadence des imageurs rapides requis sera caractérisée par la suite pour chaque architecture. Le critère de qualité de stabilisation sera d'obtenir une image de bonne qualité quel que soit le bougé initial, c'est-à-dire une image ayant au minimum une qualité de -3 JND.

3.2.3.1 Architecture sans gyroscope avec système de stabilisation optique

Description de l'architecture

L'architecture de stabilisation d'image sans gyroscope mais avec un système de stabilisation optique s'appuie sur un imageur rapide pour générer les trames à une cadence importante. Cette vitesse de génération de trame est étudiée dans les résultats car suivant la fréquence de fonctionnement, un niveau de qualité de stabilisation peut être atteint. Le principe de base de cette architecture est le suivant : l'imageur va être utilisé comme un capteur de mouvement pour remplacer le gyroscope. Les informations de déplacement générées par l'estimateur de mouvement seront envoyées au système optique de compensation. Le système optique de compensation se chargera d'ajuster le chemin optique afin que les rayons lumineux arrivent de manière stable par rapport à l'axe optique de l'imageur. Ainsi au cours du temps d'intégration, les bougés dû aux tremblements seront compensés ainsi nous obtiendrons une image stabilisée. Le schéma détaillé de cette architecture est présenté figure 64.

Afin de simuler l'architecture dans son ensemble, une perturbation est ajoutée en entrée. Dans notre cas, nous avons utilisé notre modèle de tremblement physiologique pour générer les bougés. Avec un système réel, ces perturbations proviendront des tremblements de la main. L'opérateur de convolution simule l'intégration de la scène sur le capteur prenant compte des bougés résiduels du système optique de compensation. Ainsi les trames sont créées pendant le temps de pose.

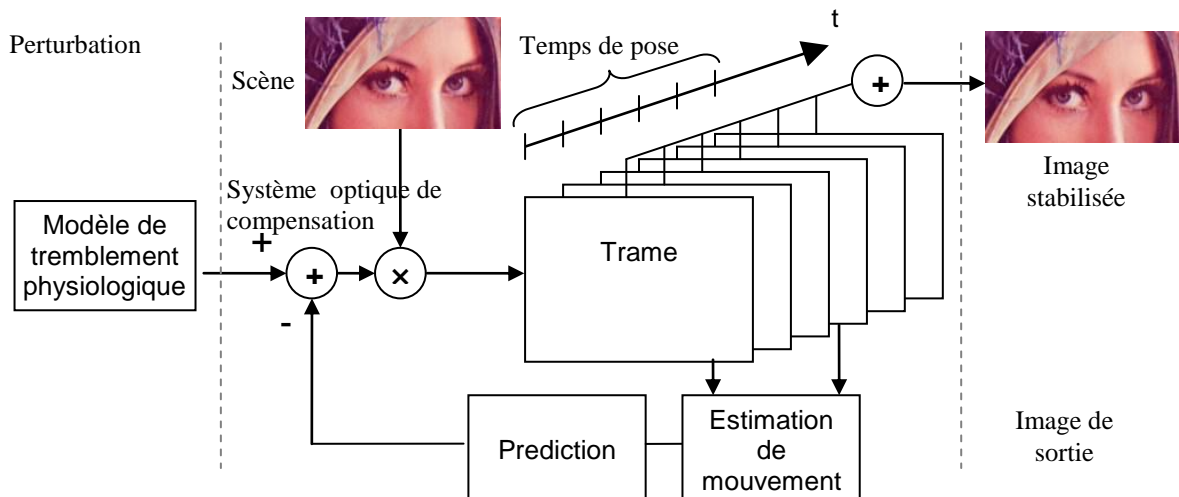


Figure 64. Détail de l'architecture de stabilisation d'image sans gyroscope avec système de stabilisation optique.

La somme de toutes les trames donne l'image finale de sortie stabilisée. Par ce découpage temporel du temps de pose en multi trame, chaque trame est composée d'une faible luminosité mais aussi d'un faible bougé. Plus le temps d'intégration des trames sera court, plus les trames seront nettes. Et plus les trames seront nettes, plus il sera facile pour l'estimateur de mouvement de mesurer le bougé entre les trames. L'estimation de mouvement peut être réalisée par tout type d'estimateur. Comme vu précédemment, l'estimation doit être réalisée de manière absolue en utilisant la première trame comme référence. Si l'estimation est réalisée de manière relative, en estimant toujours le mouvement entre 2 trames adjacentes, l'erreur d'estimation de mouvement est cumulée entre chaque trame. Alors que dans le cas de l'estimation absolue, l'erreur d'estimation n'est pas cumulée, elle est seulement propre à chaque estimation de déplacement. Cette estimation de mouvement absolue demande d'estimer des déplacements importants par rapport à une estimation relative qui va chercher des déplacements de proche en proche. Pour réduire la taille de zone de recherche, au lieu de lancer la recherche à partir de la position de la trame de référence, il est possible de partir du lieu de déplacement de la trame précédente.

Dans notre étude, nous avons utilisé deux types d'estimateurs présentés en 3.2.1.3 que sont le *block-matching* (BM) en vue d'une intégration potentielle et *Motion 2D* (M2D) utilisé comme estimateur de référence.

Après l'estimation de mouvement, les déplacements estimés sont transmis à un prédicteur permettant de rattraper le retard dû à l'attente de l'intégration de la nouvelle trame à analyser. Les différents types de prédicteurs implémentés présentés dans la section 3.2.1.3 sont : le

prédicteur 1^{er} ordre avec interpolateur linéaire, le prédicteur 1^{er} ordre avec interpolateur cubique et le prédicteur 2^{ème} ordre avec interpolateur cubique. Ensuite les informations de déplacements prédites sont transmises au système de stabilisation optique de compensation afin de réaliser la stabilisation de l'image. En simulation ceci est réalisé par la soustraction du mouvement estimé prédit au bougé d'entrée. En réalité ce système de stabilisation optique est classique, soit composé d'une lentille amovible ou soit d'un imageur amovible comme présenté dans l'état de l'art.

Résultats

Afin d'étudier les performances de cette architecture de stabilisation suivant les différents types d'estimateur et types de prédicteur, nous présentons dans la figure 65 les résultats de stabilisation dans le cas du smartphone (distance focale de 28 mm et taille pixel de 1,4 μm) avec un temps de pose d'une seconde. Cette étude a été réalisée sur la base d'images Kodak composées de 24 images [171] présentant des scènes naturelles différentes.

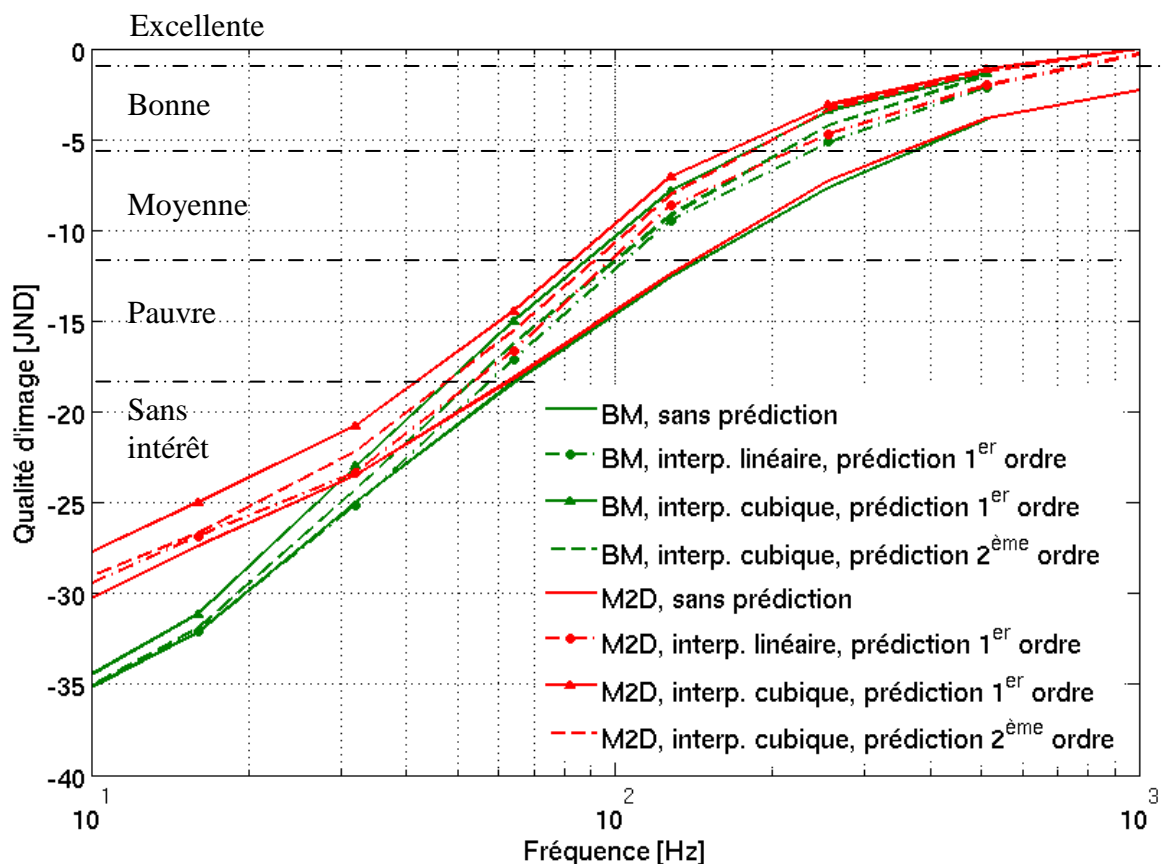


Figure 65. Résultat de simulation de l'architecture sans gyroscope suivant la fréquence de fonctionnement du système. Deux types d'estimation de mouvement sont présentés celui de mise en correspondance (BM) et *Motion 2D* (M2D).

Nous présentons figure 66 les mêmes résultats avec leur l'écart type car suivant la composition de la scène, les performances de stabilisation peuvent être différentes. En effet, c'est la fonction d'estimation de mouvement sur l'image qui va avoir ses performances dépendantes de la composition de la scène. Aux notes de qualité d'image exprimées en JND sont notés les intervalles qualificatifs de qualité d'image facilitant la lecture des résultats.

Ces résultats montrent très clairement que plus la fréquence de génération des trames est rapide, plus cette architecture de stabilisation est performante. Les deux principales raisons :

- plus les trames sont courtes, moins elles ont un flou interne donc plus l'estimation de mouvement est précise ;
- plus l'estimation est fréquente plus les trames à combiner sont nettes donc la somme de celles-ci sera nette.

Dans tous les cas, l'estimateur le plus performant est M2D. Pour des fréquences de trame faible, inférieures à 60 Hz, l'estimateur par M2D est bien meilleur que le BM. Ceci s'explique par le fait que l'estimateur par BM est limité dans sa zone de recherche de déplacement (à +/- 5 pixels). Plus la fréquence trame sera faible, plus le déplacement à estimer entre les trames sera important. Le décrochage se produit lorsque les déplacements à estimer se retrouvent en dehors de la zone de recherche. Pour éviter ce problème il est possible d'augmenter la taille de la zone de recherche au prix d'une augmentation de la complexité. Néanmoins cette action est limitée du fait qu'à ces faibles fréquences de fonctionnement, la qualité d'image sera si mauvaise que cette dernière ne sera pas bonne à garder.

Nous pouvons voir que l'ajout d'un prédicteur dans ce type d'architecture est fortement bénéfique car il permet de gagner 5,3 JND de qualité d'image dans le meilleur des cas, donc de faire souvent passer les images dans une catégorie supérieure. Par exemple à la fréquence de 256 Hz, l'image passe de la catégorie moyenne à bonne.

L'utilisation de l'interpolation cubique par rapport à l'interpolateur linéaire permet au prédicteur de gagner dans le meilleur des cas 3 JND.

Dans la figure suivante, nous reprenons les résultats de l'architecture sans gyroscope dans les cas avec et sans prédiction en y ajoutant l'écart type de qualité d'image obtenu par la stabilisation des différentes images.

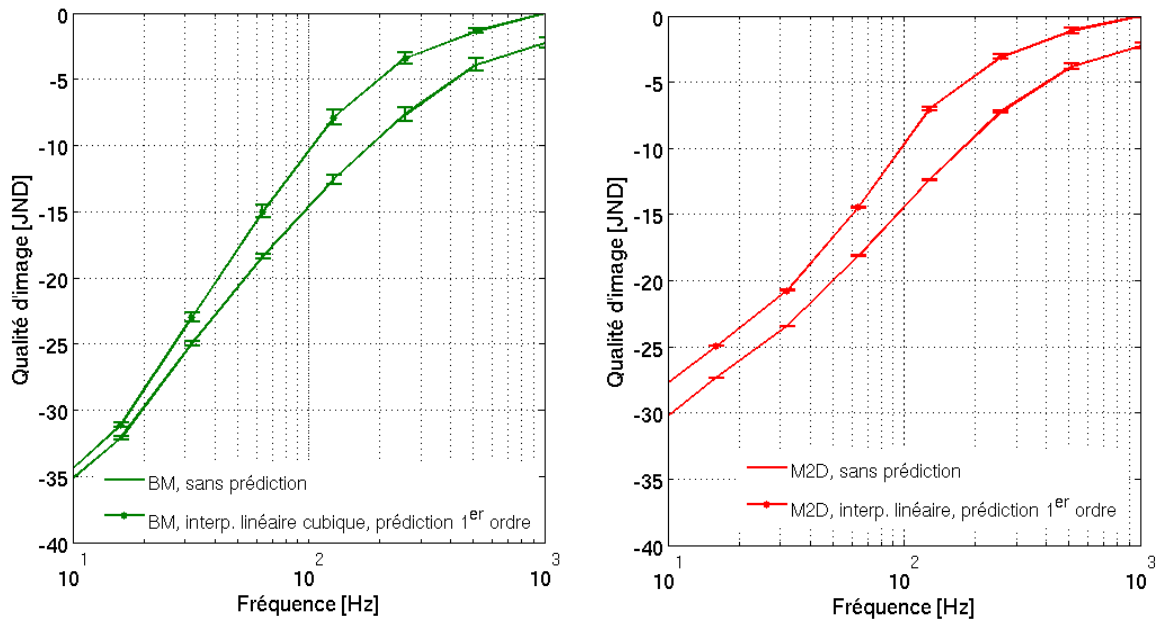


Figure 66. Qualité d'image lors de la stabilisation suivant la fréquence de cadence des imageurs avec leurs écarts types respectifs.

Nous pouvons remarquer que l'estimateur de mouvement M2D est moins sensible au type de scène que l'estimateur de mouvement BM. L'écart type maximum dans le cas de M2D est de 0,21 JND alors que pour le cas de BM il est de 0,6 JND. Ces résultats montrent qu'en termes de qualité finale de l'image les résultats de stabilisation sont similaires pour des scènes naturelles.

Conclusion

Grâce aux résultats de cette architecture de stabilisation, nous avons démontré qu'il est possible de se passer du gyroscope en utilisant l'imageur comme un capteur d'environnement pour estimer le bougé. Plus la qualité de service recherchée sera importante, plus l'imageur devra avoir une fréquence de fonctionnement importante. Afin d'obtenir une bonne qualité d'image définie à -3 JND en utilisant un prédicteur 1^{er} ordre avec interpolateur cubique, l'imageur devra avoir une fréquence de fonctionnement de plus de 270 Hz dans le cas de l'estimateur M2D et plus de 300 Hz dans le cas de l'estimateur BM. Ces fréquences rapides de fonctionnement sont atteignables vis-à-vis de l'état de l'art sur certains imageurs rapidement. Par contre comme nous allons le voir par la suite, cette approche sera limitée par le bruit interne aux trames.

3.2.3.2 Architecture sans gyroscope ni système de stabilisation optique

Détail de l'architecture

L'architecture sans gyroscope ni système de stabilisation optique utilise seulement un imageur rapide. Le principe général de cette architecture est de réaliser une somme cohérente de toutes les trames afin d'obtenir une image stabilisée. Nous pouvons parler d'une architecture totalement intégrée car elle ne comporte aucun composant externe. De même que l'architecture précédente, pour les simulations nous avons utilisé notre modèle de tremblement physiologique comme source de perturbation. A une certaine fréquence, l'imageur rapide génère des trames tout au long du temps de pose. Puis une estimation de mouvement absolue est réalisée entre les trames pour estimer le déplacement afin de sommer les trames entre elles en décalant les trames suivant leur déplacement. Cette sommation avec décalage, nommée sommation cohérente est présentée dans la section 3.2.2. L'estimation de mouvement peut être réalisée par tout type d'estimateur de mouvement. Pour démontrer la fonctionnalité dans notre cas, l'estimateur de référence M2D a été utilisé.

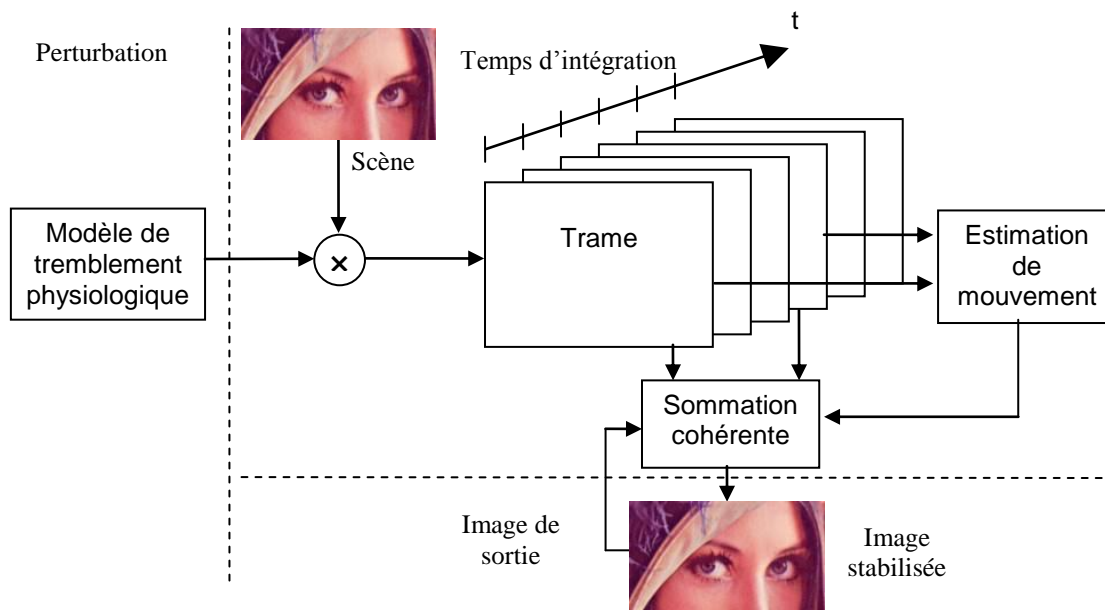


Figure 67. Détail de l'architecture de stabilisation d'image sans gyroscope ni système de stabilisation optique.

Résultats

Les performances de l'architecture complètement intégrée sont présentées figure 68. Les trois cas ont été simulés afin de voir l'évolution du JND suivant le temps de pose et la

distance focale : smartphone (distance focale de 28 mm ; taille de pixels de 1,4 μm) avec des temps de pose de 1/10 et 1 seconde ; appareil photo compact (distance focale de 280 mm ; taille de pixels de 2,8 μm) avec un temps de pose de 1/10 de seconde.

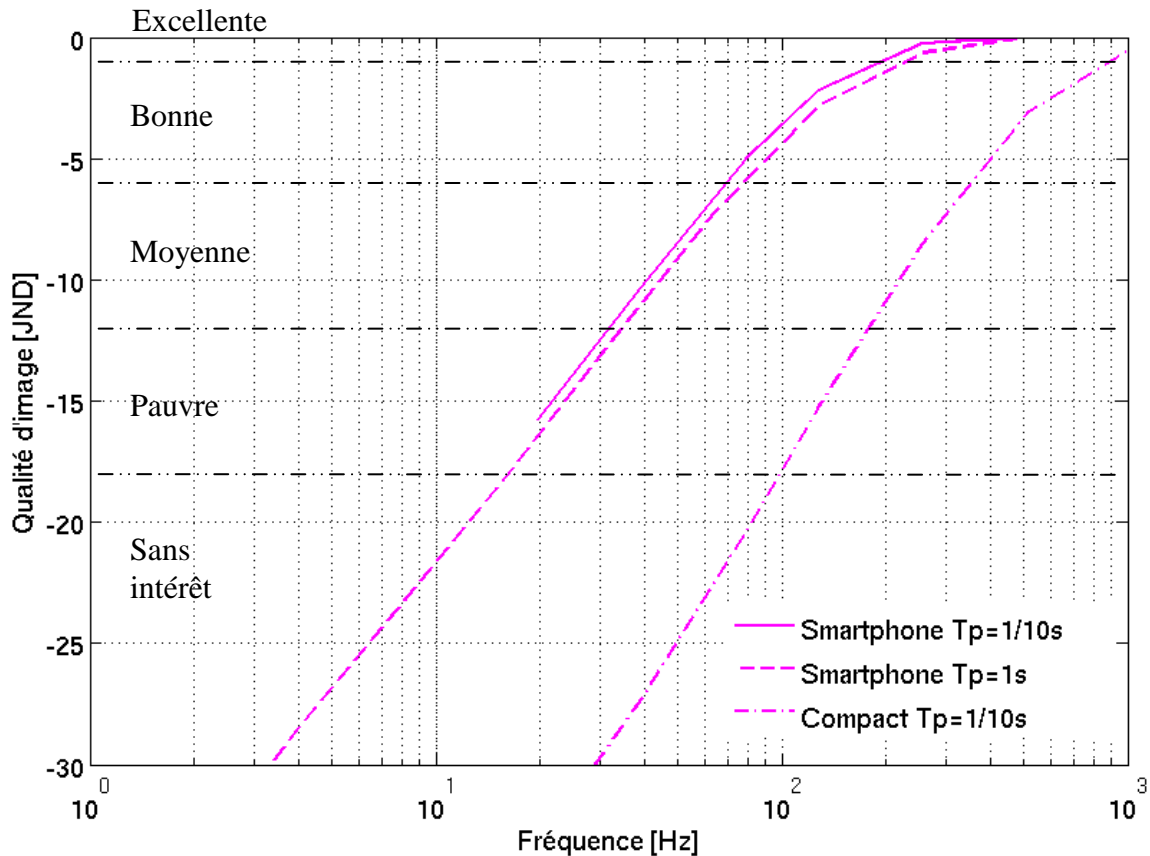


Figure 68. Résultat de simulation de l'architecture de stabilisation sans gyroscope ni système de stabilisation optique en qualité d'image suivant la fréquence de fonctionnement du système.

Les résultats montrent que plus la fréquence du système est élevée, meilleur est la qualité des images donc meilleur est la stabilisation. Les résultats du smartphone montrent que plus le temps de pose est important plus la fréquence de fonctionnement doit être légèrement plus élevée pour atteindre une même qualité d'image. Cette tendance observée sera approfondie par la suite.

Logiquement, l'augmentation de la distance focale amplifie le bougé sur l'imageur, ce qui a pour conséquence de nécessiter une forte augmentation de la fréquence de fonctionnement pour maintenir la bonne qualité de l'image.

Conclusion

En conclusion cette architecture démontre qu'il est possible de se passer de tous les éléments extérieurs à l'imageur pour la stabilisation que sont le gyroscope et le système de stabilisation optique.

Pour atteindre une bonne qualité de service définie à -3 JND, le Smartphone devra atteindre une cadence de 111 Hz et 125 Hz pour un temps de pose de 1/10 et 1 seconde respectivement. Dans le cas de l'appareil photo compact à une distance focale de 280 mm, pour atteindre cette qualité de service, le système devra atteindre une cadence de trame de 532 Hz pour le cas d'un temps de pose de 1/10 seconde.

3.2.3.3 Architecture sans système de stabilisation optique avec gyroscope

Description de l'architecture

L'architecture sans système de stabilisation optique mais avec gyroscope est présentée en détaille figure 69.

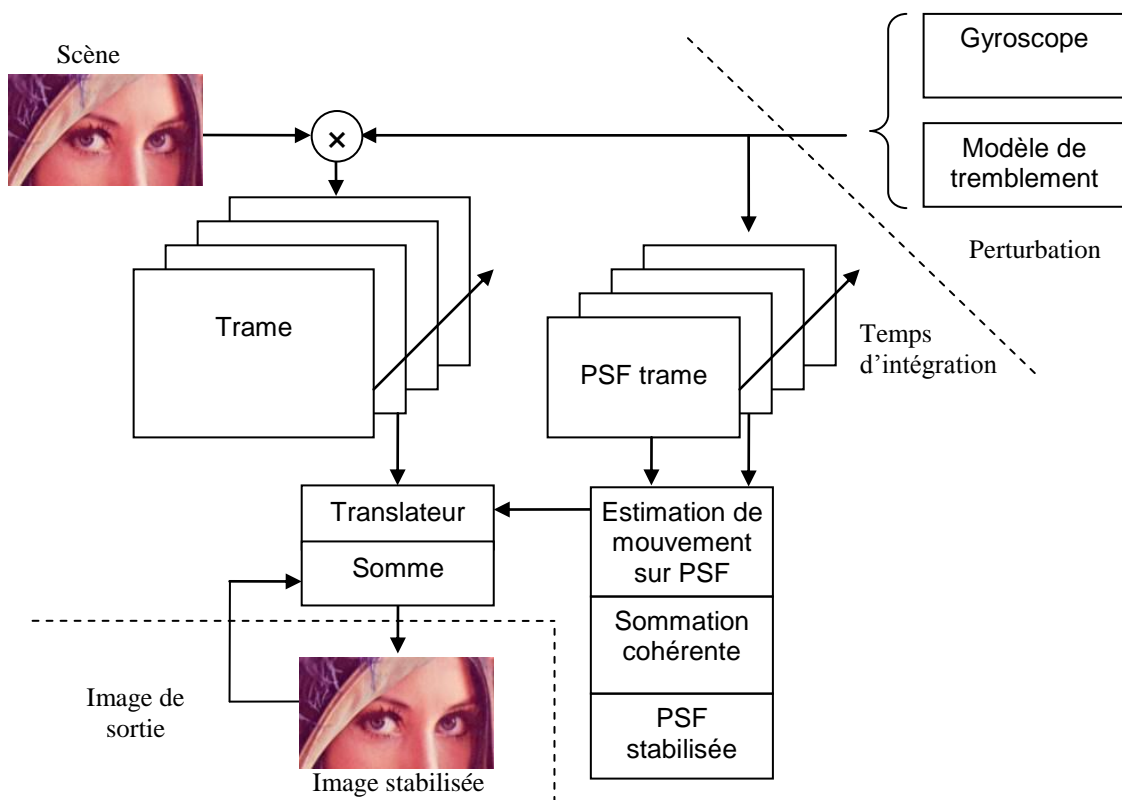


Figure 69. Détail de l'architecture de stabilisation sans système optique avec gyroscope.

Le principe général de cette architecture est de travailler directement sur la PSF du bougé pour stabiliser l'image finale. Les perturbations d'entrée ont été simulées par notre modèle de tremblement physiologique, mais dans un système réel, les informations de tremblement sont fournies par les gyroscopes. A la même cadence trames, des PSF trames sont construites à partir des données du bougé, ainsi pour chaque trame est associée sa PSF trame. Entre les PSF trames, une estimation de mouvement sur PSF est réalisée pour retrouver le déplacement entre les trames. Cette estimation de mouvement sur PSF est réalisée par la métrique de qualité d'image. En utilisant la métrique de qualité, l'estimateur de mouvement sur PSF va chercher à combiner de la meilleure manière les PSF trames afin d'avoir la meilleure netteté possible. Ainsi une sommation cohérente des PSF trames ainsi que des trames seront réalisées comme présenté précédemment dans la section 3.2.2.

Résultats

Les performances de l'architecture avec gyroscope sont présentés figure 70. Les trois cas ont été simulés afin de voir les variations suivant le temps de pose et la distance focale : smartphone (distance focale de 28 mm ; taille de pixels de 1,4 μm) avec des temps de pose de 1/10 et 1 seconde ; appareil photo compact (distance focale de 280 mm ; taille de pixels de 2,8 μm) avec un temps de pose de 1/10 de seconde.

De même que les autres architectures, les résultats montrent que plus la fréquence du système est élevée, meilleure est la qualité des images donc meilleure est la stabilisation. De même que l'architecture précédente, les résultats du smartphone montrent que plus le temps de pose est important plus la fréquence de fonctionnement doit être légèrement plus élevée pour atteindre une même qualité d'image. L'augmentation de la distance focale entraîne une forte augmentation de la fréquence de fonctionnement pour atteindre une bonne qualité de service.

Conclusion

En conclusion cette architecture démontre qu'il est possible de se passer du système de stabilisation optique et de travailler directement sur la PSF pour stabiliser l'image de sortie.

Pour atteindre une bonne qualité de service définit à -3 JND, le smartphone devra atteindre une cadence de 108 Hz et 118 Hz pour un temps de pose de 1/10 et 1 seconde respectivement. Dans le cas de l'appareil photo compact avec une distance focale de 280 mm, pour atteindre cette qualité de service, le système devra atteindre une cadence de trame de 500 Hz pour le cas d'un temps de pose de 1/10 seconde.

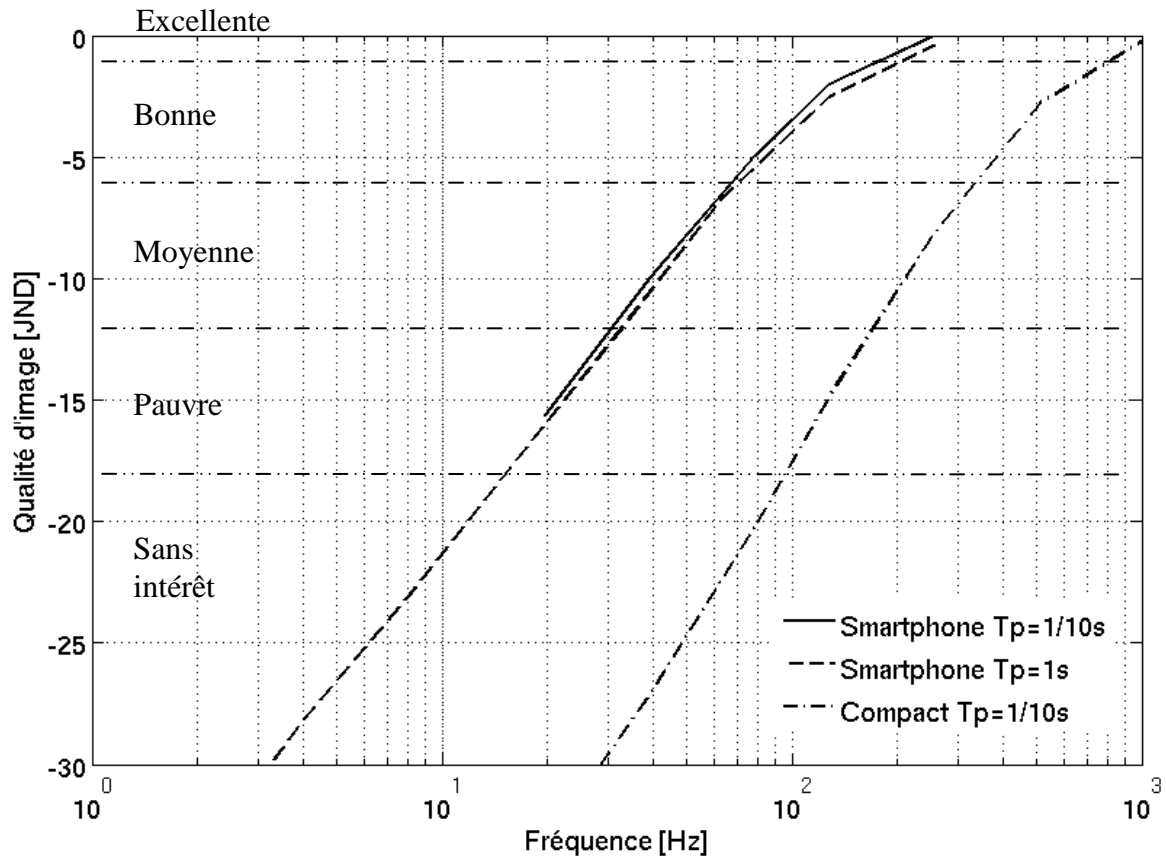


Figure 70. Résultat de simulation de l'architecture de stabilisation sans système de stabilisation optique mais avec gyroscope en qualité d'image suivant la fréquence de fonctionnement du système.

3.2.4 Comparaison des performances de stabilisation des architectures

Nous allons maintenant comparer les performances de stabilisation des architectures. Les résultats de simulation présentés figure 71 utilisent le smartphone avec un temps de pose de 1 seconde comme référence de comparaison. Pour chaque fréquence de fonctionnement, la qualité d'image moyenne des trames a été calculée afin d'obtenir une limite de performance de nos architectures de stabilisation. Nous parlons de performance limite car ceci correspond au flou intrinsèque moyen pour une trame. De ce fait, comme ces algorithmes ne peuvent pas détecter un flou plus petit qu'une trame, ce type d'algorithme ne pourra jamais dépasser cette limite. Pour l'architecture avec système de stabilisation optique sans gyroscope, les résultats avec prédicteur idéal ont été ajoutés afin de connaître le gain maximal de la prédiction. Le prédicteur idéal a été construit de manière à ne plus avoir de retard dans la compensation. Ce prédicteur idéal prend comme entrée l'estimation idéale de bougé interpolé de façon cubique.

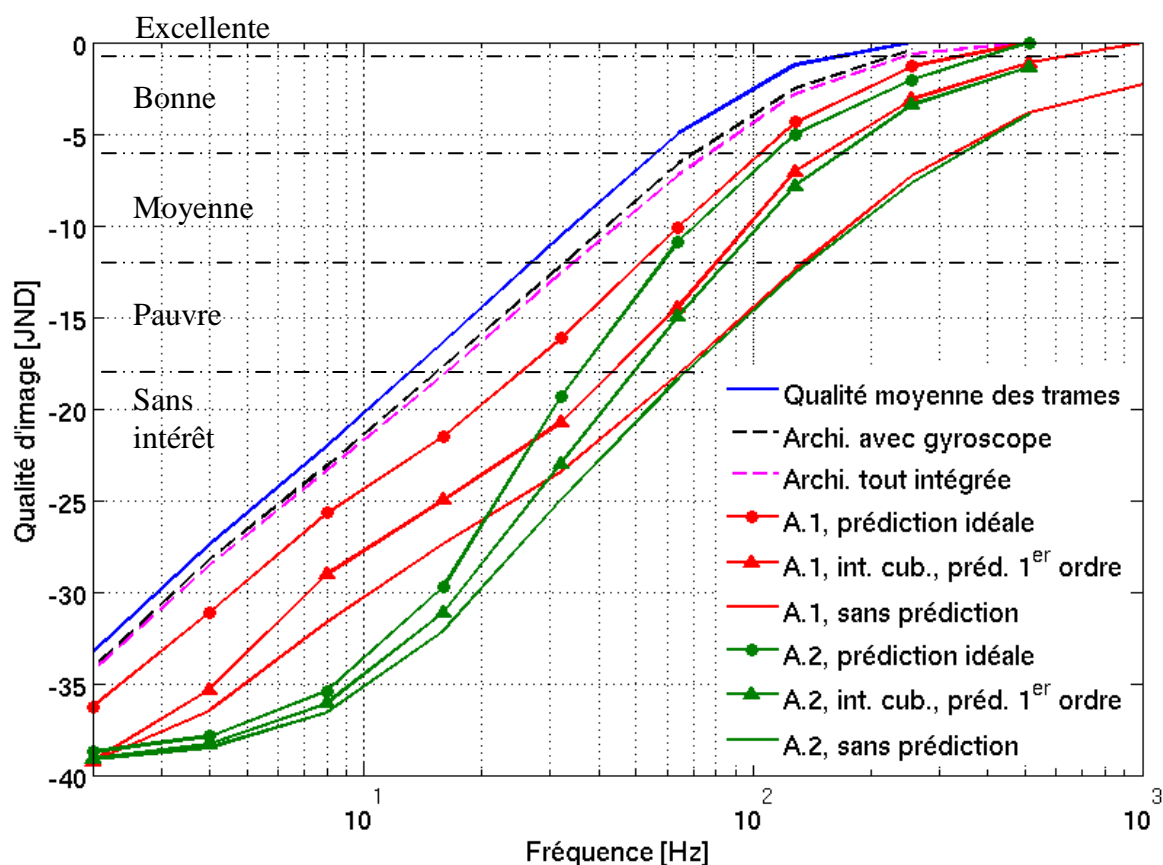


Figure 71. Comparaison des différentes architectures suivant leurs fréquences de fonctionnement pour le cas du smartphone avec un temps de pose de 1 seconde. A.1 et A.2 : Architecture de stabilisation avec système de stabilisation optique sans gyroscope utilisant l'estimateur de mouvement M2D (A.1) et BM (A.2).

Les architectures avec seulement le gyroscope et celle complètement intégrée utilisant la technique de sommation cohérente sont les plus efficaces en termes de fréquence de fonctionnement et de qualité de stabilisation. Elles approchent quasiment les performances limites de stabilisation telles que présentées ci-dessus. Si nous reprenons notre critère à atteindre de -3 JND, la fréquence du système de stabilisation sera de 118 Hz, 125 Hz, 270 Hz et 295 Hz pour les architectures respectives : avec gyroscope, tout intégrée, avec prédiction 1^{er} ordre dans le cas M2D et BM. Les architectures avec gyroscope et celle intégrée sont plus efficaces que l'architecture avec système de stabilisation optique sans gyroscope car du fait que pour bien reconstruire le tremblement il faut à la base des trames nettes, contrairement aux autres architectures qui vont travailler directement sur le flou des trames.

Nous pouvons remarquer qu'en dessus de -18 JND, les courbes des deux architectures avec système de stabilisation optique mais sans gyroscope (A.1 et A.2) ne se croisent plus et se rejoignent. L'estimation de mouvement BM arrive à atteindre les performances de

l'estimation de mouvement M2D, donc pour les systèmes embarqués, nous pouvons privilégier la solution avec le BM car cette solution est plus faible en complexité par rapport à la solution M2D.

Pour l'architecture avec système de stabilisation optique sans gyroscope, le calcul de l'entropie permet de voir que plus le mouvement est estimé à une fréquence élevée, plus il sera précis. Si nous reprenons l'apport du modèle de tremblement dans le calcul de l'entropie de la section 3.2.1.2 mais cette fois ci en regardant l'entropie du résidu dans le cas de l'estimateur M2D, nous remarquons qu'il faut atteindre une fréquence de fonctionnement importante pour que la probabilité du signal soit faible. Le signal de résidu est simplement le signal de bougé d'entrée moins l'estimation du bougé prédit. La figure 72 présente les résultats de l'entropie de la taille des déplacements du résidu dans le cas du smartphone pour un temps de pose 1/10 seconde.

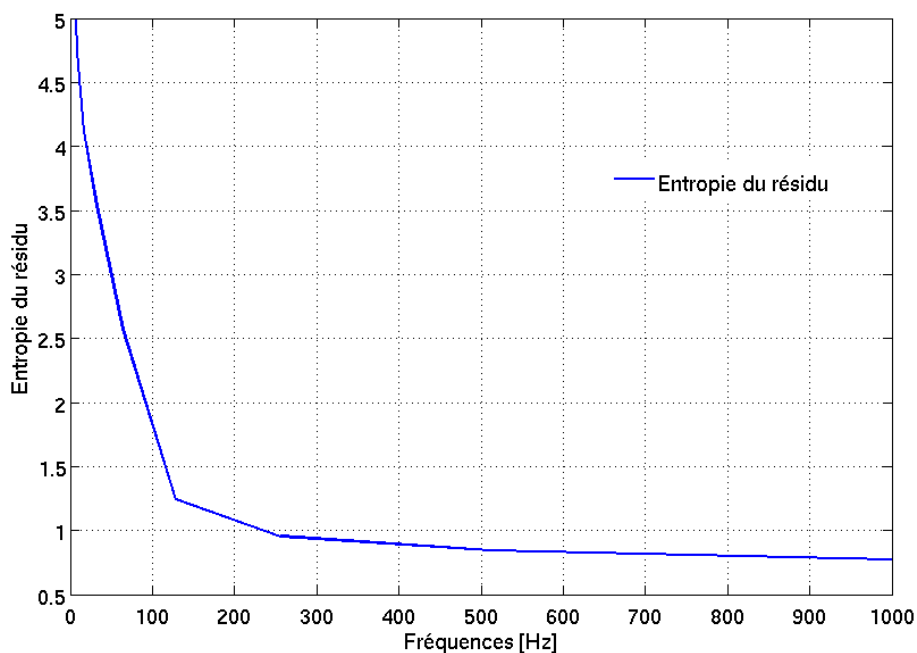


Figure 72. Résultat de l'entropie de la taille des déplacements du signal résiduelle de bougé suivant la fréquence trame pour l'architecture avec système de stabilisation optique sans gyroscope.

Les résultats de cette étude montrent qu'il faut atteindre une certaine fréquence pour avoir une taille de résidu prédictif. Dans le cas présent, la fréquence à atteindre est de plus de 100 Hz afin d'envisager un gain de prédiction important, ce que nous observons dans la figure 65. Au-delà de cette fréquence de fonctionnement, le gain dû à la prédiction est important alors

que pour des fréquences de fonctionnement plus faible ce gain est minime ou parfois inexistant.

Pour simplifier la suite de la présentation des résultats, le nom des 2 architectures, que sont d'une part l'architecture complètement intégrée et d'autre part l'architecture sans système de stabilisation optique mais avec gyroscope, seront regroupées en : « architecture avec sommation cohérente ». En effet, les résultats de ces deux architectures sont très proches du fait de l'introduction de la sommation cohérente dans leur fonctionnement.

Concernant l'architecture avec système de stabilisation optique mais sans gyroscope nous retiendrons la solution avec l'estimateur de mouvement BM ainsi qu'avec le prédicteur le plus efficace (1^{er} ordre avec interpolation cubique) car cette solution nous semble plus intégrable que celle avec l'estimation de mouvement M2D. Pour aussi simplifier la présentation nous appellerons cette architecture, l'architecture sans gyroscope.

Nous allons étudier la fréquence trame requise pour ces deux types d'architecture afin d'atteindre une qualité de service fixée à -3 JND suivant l'évolution du temps de pose ou de la distance focal pour le cas de taille de pixel fixé à $1,4 \mu\text{m}$. Pour réaliser ces courbes, nous avons réalisés des simulations avec différents paramètres (distance focale, temps de pose) afin d'obtenir des courbes similaires à la figure 71. Puis nous en avons tiré les caractéristiques (fréquence trame, nombre de trame) du système pour atteindre une qualité d'image de -3 JND.

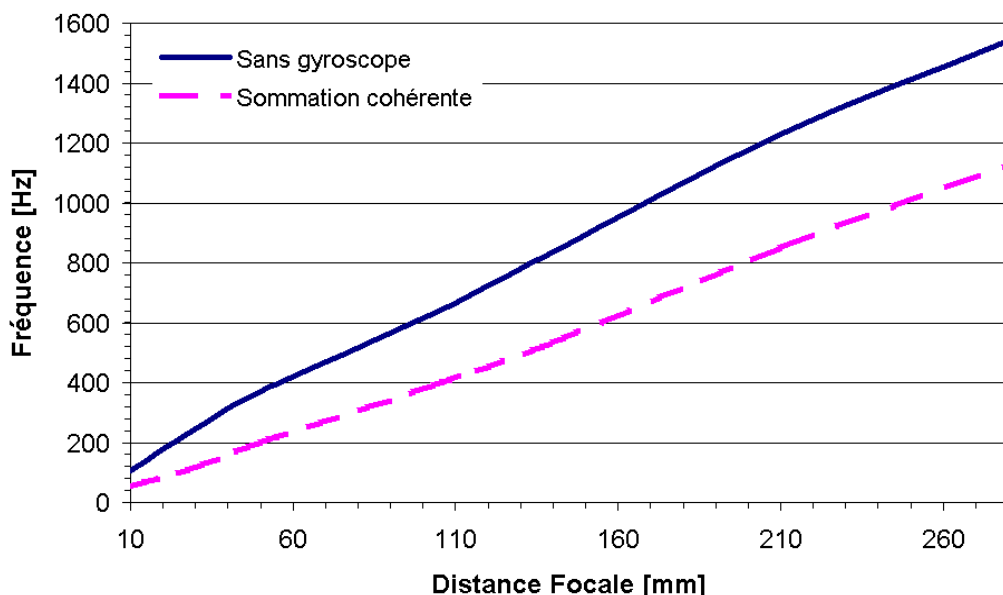


Figure 73. Evolution de la fréquence de fonctionnement du système pour garantir une stabilisation optimale (-3 JND) suivant la distance focale de l'appareil photo pour un temps de pose $1/10$ de seconde.

L'évolution de la fréquence des trames suivant la distance focale est présentée figure 73. Cette étude a été réalisée pour un temps de pose de 1/10 seconde. En conséquence dans cette étude, le temps de pose est toujours le même (1/10 s), seulement la distance focale de l'appareil change.

Les résultats montrent que plus la distance focale est importante, plus la fréquence trame devra être importante. Cette augmentation de fréquence provient de l'augmentation du bougé sur l'imager lorsque la distance focale augmente.

Maintenant si l'on regarde non plus l'évolution de la fréquence des trames mais l'évolution du nombre de trames suivant la distance focale, toujours avec un temps de pose de 1/10 seconde, nous avons les résultats suivants présentés figure 74.

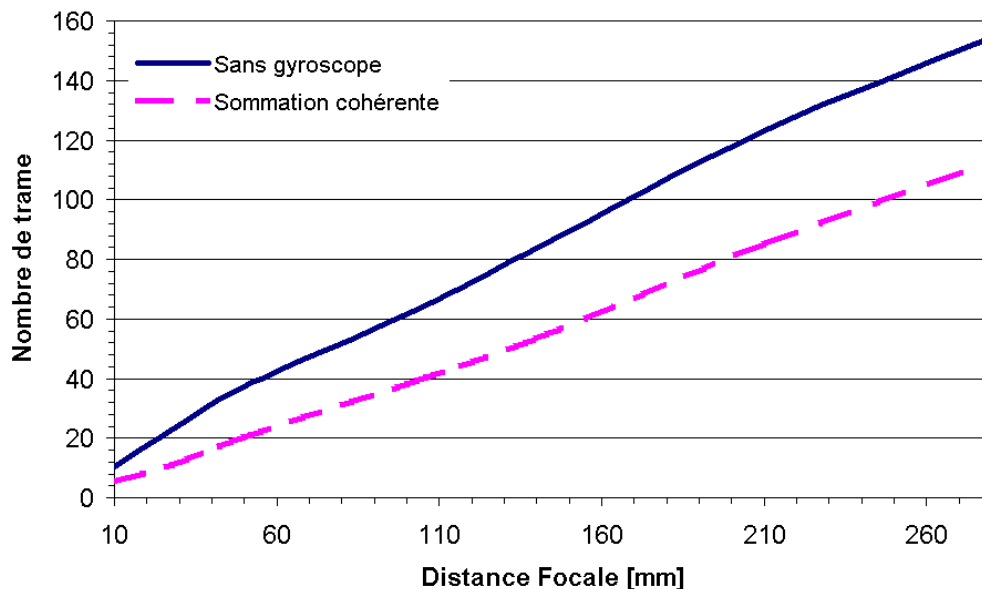


Figure 74. Evolution du nombre de trame requis suivant la distance focale pour garantir une bonne qualité de service.

Pour un même temps de pose, plus la distance focale sera importante, plus le nombre de trames sera important pour garantir une bonne qualité de service. Nous avons une augmentation quasi linéaire du nombre de trames suivant la distance focale. Les régressions linéaires pour les courbes des architectures sans gyroscope et de sommmation cohérente sont respectivement :

$$y = 0,53x + 8,27 \quad (20)$$

$$y = 0,4x + 0,01 \quad (21)$$

avec un coefficient de corrélation de 0,99 pour les deux droites.

Dans les tests suivants, le cas du smartphone sera utilisé. On rappelle qu'il a une distance focale fixe de 28 mm et une taille pixel de 1,4 μm . Dans les résultats de la figure 75, l'évolution de la fréquence de fonctionnement suivant le temps de pose est étudiée. Nous pouvons remarquer que le temps de pose minimal est de 1/50 seconde car à cette vitesse-là, il y a plus de 50 % de chance d'avoir une image de bonne qualité d'après les résultats présentés figure 52 section 2.3.3.2. Donc pour des temps de pose inférieurs à 1/50 seconde, il est préférable d'avoir une seule trame car le temps d'obturation sera suffisamment rapide pour obtenir une image nette.

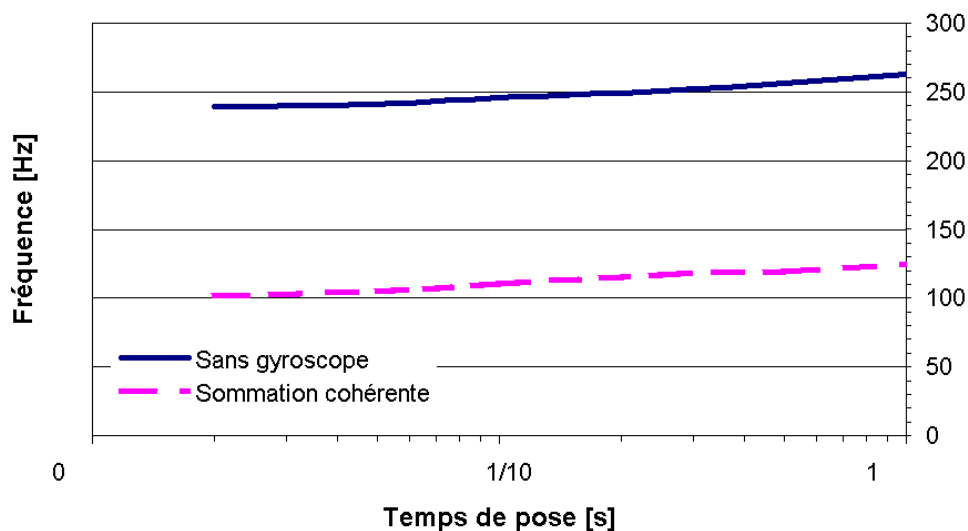


Figure 75. Evolution de la fréquence de fonctionnement du système suivant le temps de pose afin d'obtenir une qualité de service définie à -3 JND.

Nous pouvons remarquer que suivant le temps de pose, l'augmentation de la fréquence de fonctionnement requise par le système est faible pour garantir la qualité de service, que ce soit pour l'architecture dite sans gyroscope que pour celles avec sommaton cohérente. Cette légère augmentation provient du fait que plus le temps de pose est long, plus l'erreur résiduelle cumulée au cours du temps sera importante. Ainsi il faudra augmenter légèrement la fréquence de fonctionnement pour compenser cette perte.

De façon complémentaire avec les mêmes conditions que précédemment, l'évolution du nombre de trames suivant le temps de pose est présentée figure 76.

L'évolution du nombre de trame est quasiment proportionnelle suivant le temps de pose car comme vu précédemment la fréquence augmente peu suivant l'augmentation du temps de pose.

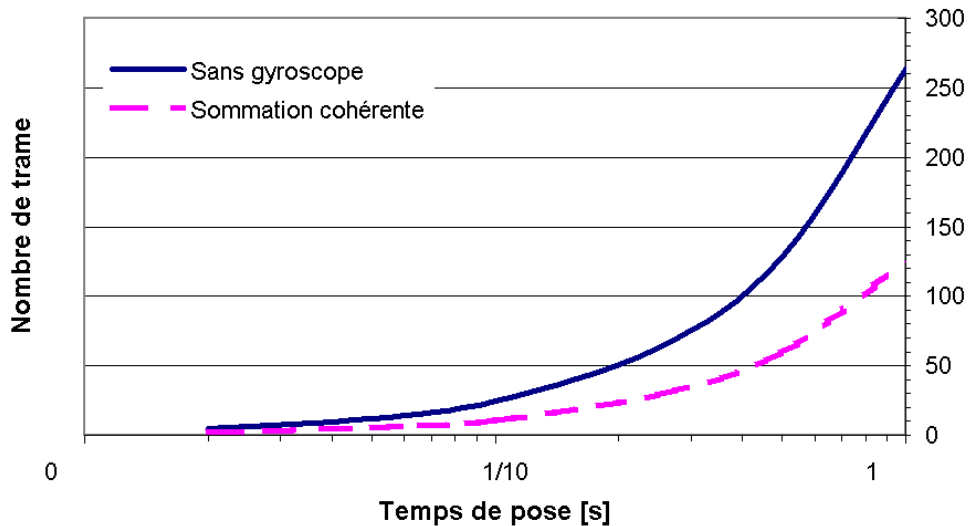


Figure 76. Evolution du nombre de trame des architectures de stabilisation suivant le temps de pose pour garantir une qualité de -3 JND.

3.2.5 Conclusion et limitations

En conclusion, pour garantir une qualité d'image stabilisée, il faudra une fréquence de trame d'autant plus importante que :

- La distance focale augmente ou que la taille des pixels de l'imageur diminue.
- Le temps de pose augmente.

L'augmentation du temps de pose ou de la distance focale aura un impact fort sur le nombre de trames, ce qui va poser des problèmes aux algorithmes d'estimation de mouvement car plus le nombre de trame sera important, plus le rapport signal sur bruit ou *signal noise ratio* (SNR) de chacune sera faible. Plus précisément, alors que pour un signal on parle de SNR, pour les images, ou trame dans notre cas, nous allons parler de *peak signal to noise ratio* (PSNR).

Limitations dues au PSNR

Différentes sources de bruit vont faire diminuer le SNR du signal d'entrée des pixels comme présenté figure 77. Le signal de la photodiode représentant les pixels de l'imageur va être perturbé par un bruit lié à la nature physique de la lumière et de sa capture appelé le bruit photonique. Le bruit photonique provient du fait que le flux de photons reçu n'est pas parfaitement continu, il est soumis à un processus de Poisson. Ce bruit est proportionnel au carré du signal, donc proportionnel au carré du nombre de photons captés par la photodiode.

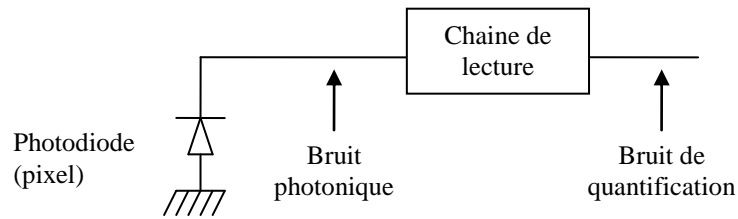


Figure 77. Chaîne simplifiée de l'acquisition d'un pixel de l'imager avec les différentes sources de bruit.

En plus de cela va s'ajouter un bruit de quantification. Ce bruit provient de la conversion du signal analogique en signal numérique.

Pour simuler les limites de nos architectures, nous avons mis en place une chaîne de bruit présenté figure 78.

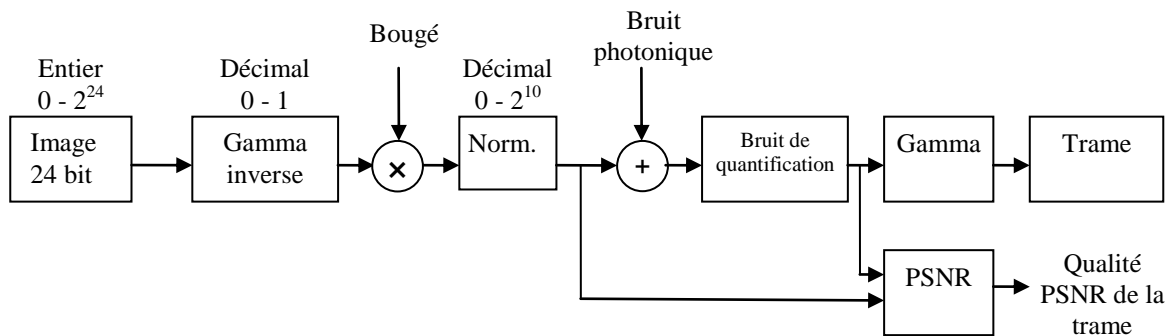


Figure 78. Schéma de la chaîne du bruit utilisée pour les simulations Matlab permettant de mesurer le PSNR de l'image.

Cette chaîne de bruit est construite de la façon suivante :

- Elle prend en entrée une image couleur échantillonnée sur 24 bits.
- Puis elle est convertie en niveau de gris en décimal compris entre 0 et 1 afin d'être convertie par la fonction gamma inverse. Cette transformation permet de retrouver les valeurs réelles des pixels sur le capteur. Nous avons utilisé un gamma usuel de 2,2.
- De la même manière que dans le système globale de simulation, le bougé est appliqué sur la trame en cours d'intégration.
- Puis la trame est normalisée sur 10 bits en vue de lui appliquer un bruit photonique sur 10 bits de dynamique.

- Pour chaque valeur de pixel de la trame, suivant une répartition gaussienne, le bruit photonique est ajouté avec une valeur de sigma en charge par l'équation suivante :

$$\sigma_{\text{Bruit}(e^-)} = \sqrt{Ne^-} \quad (22)$$

Avec Ne^- le nombre total de charges dans la photodiode pour une image finale. Maintenant si l'on converti ce bruit en bits et suivant la trame en cours d'intégration, le nombre de charges compris dans la photodiode sera le nombre total de charges (Ne^-) divisé par le nombre de trames composant l'image finale. Par conséquence l'équation calculant la valeur du sigma en bit sera :

$$\sigma_{\text{Bruit}(bit)} = \frac{\text{code}}{\sqrt{\frac{Ne^-}{NbTrame}}} \quad (23)$$

Avec le code égale à 1024 car nous considérons un échantillonneur 10 bits, un $NbTrame$ variant suivant le nombre de trames qui sera utilisé pour la stabilisation d'image (cf. figure 74 et 76).

- Ensuite la quantification est réalisée par l'arrondissement des valeurs entières suite à l'ajout du bruit photonique pour les transformer en valeurs décimales simulant ainsi l'échantillonnage des valeurs des pixels. Ceci correspond au bruit de quantification de la trame estimée en général à une valeur de 1 bit de poids faible, de 1 *least significant bit* (LSB).
- Enfin le PSNR est calculé sur la trame. Si la trame doit être affichée, la correction gamma est réalisée.

Pour cette étude de PSNR, nous nous sommes placés dans le cas du smartphone ayant un pas pixel de 1,4 μm donc pouvant contenir 6000 électrons en pleine capacité [172]. L'évolution du PSNR trame suivant le nombre de trames est présenté figure 79.

Ces résultats montrent que plus le nombre de trames augmente, plus la valeur du PSNR est faible. Lorsque le bruit de chaque trame est trop élevé (PSNR < 20 dB), les estimateurs de mouvement ne permettent plus d'estimer le mouvement. Dans nos simulations nous avons une image en entrée avec bon PSNR de 38,3 dB correspondant aussi au cas où il n'y a qu'une seule trame. Dans le cas où il y a 100 trames, le PSNR va chuter à 19,1 dB. Afin de calculer la robustesse des estimateurs de mouvement, nous avons repris les simulations des différentes architectures en insérant la chaine de bruit.

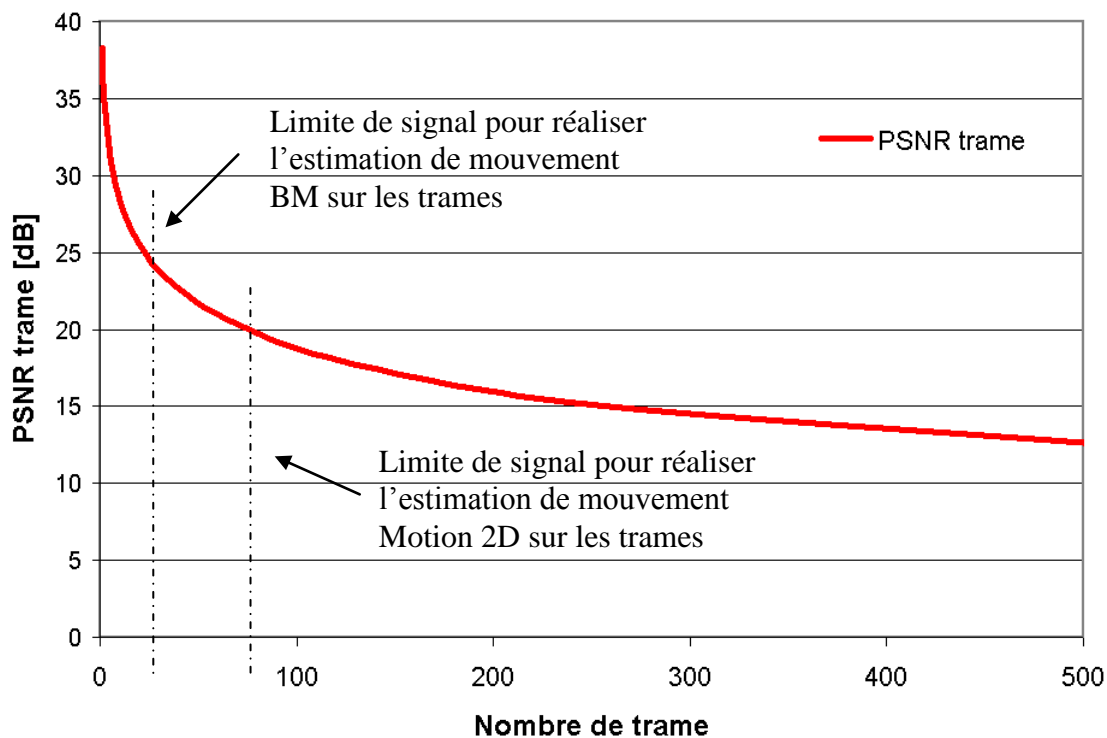


Figure 79. Evolution du PSNR d'une trame en fonction du nombre de trames requis pour les architectures de stabilisation des images. Leur seuil de limite de fonctionnement est présenté pour les estimateurs de mouvement *Motion 2D* (M2D) et *block-matching* (BM).

Nous en avons déduit les limites de fonctionnement lorsque l'estimateur décroche de 1 JND par rapport au cas idéal, c'est-à-dire sans bruit. Ainsi pour l'algorithme d'estimation de mouvement *Motion 2D*, nous avons trouvé une limite de fonctionnement pour les trames ayant un PSNR de 20 dB. Cette bonne résistance au bruit vient de l'approche pyramidale de cet estimateur, pour lequel, plus l'image sera grande, plus la résistance au bruit sera importante. En dessous de 20 dB dans chaque trame, l'estimateur de mouvement n'arrive plus à estimer le mouvement de manière correcte, les trames ont un bruit trop fort face au signal. Cette limite imposée par le PSNR des trames correspond à 74 trames dans l'image.

Dans le cas de l'estimation de mouvement par mise en correspondance (BM), notre estimateur avec des fenêtres ayant pour taille 10×10 pixels va trouver sa limite de fonctionnement pour 23,7 dB, ce qui correspond à 30 trames dans l'image.

Nous voyons donc qu'il y a une forte contrainte due au PSNR des trames pour les architectures qui utilisent des estimateurs de mouvements basées sur les trames. Si l'on regarde les figures 74 et 76, nos architectures seront limitées en temps d'exposition et distance focale pour garantir une bonne qualité d'image.

Conclusion

En conclusion, pour garantir la netteté des images grâce à ces architectures, il faudra arriver à une fréquence de trame d'autant plus grande que la distance focale augmente ou que la taille des pixels de l'imageur diminue. De même, plus le temps de pose sera important, plus la fréquence de génération des trames devra être légèrement plus importante. Ainsi suivant la fréquence maximale de l'imageur rapide, une première limitation fréquentielle sera présente pour garantir la stabilisation de l'image.

Nous avons vu aussi que plus le nombre des trames utiles pour la stabilisation est important, plus le PSNR de chaque trame sera faible. Ainsi à partir d'un certain nombre de trame dépendant du type d'estimateur de mouvement utilisé, l'estimation de mouvement ne sera plus réalisée correctement. Nous verrons en conséquence les performances des architectures de stabilisation chuter jusqu'à ne plus garantir la qualité de service souhaitée. Néanmoins, cette limitation s'applique seulement aux deux architectures sans gyroscope que sont l'architecture avec système de stabilisation optique sans gyroscope et l'architecture sans système de stabilisation optique ni gyroscope.

L'architecture sans système de stabilisation optique mais avec gyroscope sera donc seulement limitée par la fréquence de fonctionnement de l'imageur rapide.

Pour contourner les deux limitations que sont la vitesse de fonctionnement ainsi que la valeur du PSNR des trames, nous avons imaginé une nouvelle architecture qui permet d'optimiser le temps d'intégration de chaque trame afin de relâcher la contrainte fréquentielle ainsi que de maximiser le PSNR des trames tout en s'assurant de la netteté finale de l'image.

3.3 Nouveau paradigme : segmentation du temps de pose pour garantir la qualité de l'image

Cette section présente une nouvelle architecture s'appuyant sur un nouveau paradigme : la segmentation du temps de pose pour garantir la qualité de l'image. Cette architecture se base d'une part sur les observations faites sur le modèle de tremblement humain présenté chapitre 1 et d'autre part sur la métrique de qualité d'image présenté chapitre 2.

3.3.1 Présentation de l'architecture de stabilisation avec consigne de qualité

Le principe de base de cette architecture de stabilisation est de surveiller la qualité de l'image en cours d'intégration. Cette surveillance est rendue possible grâce à la métrique de qualité. Une veille continue est réalisée pendant l'intégration de chaque trame afin de garantir leur netteté comme présenté figure 80.

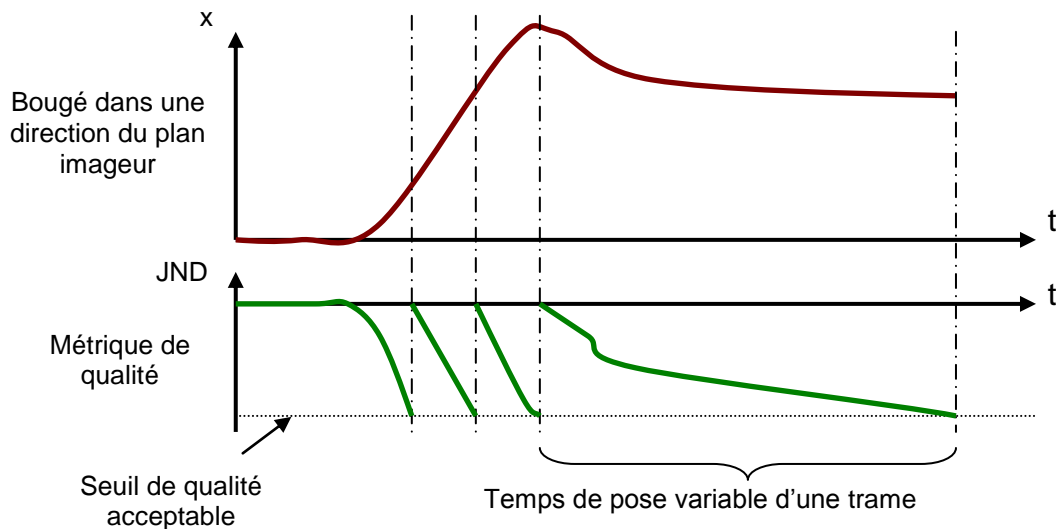


Figure 80. Schéma de principe de l'évaluation du bougé par la métrique de qualité au cours du temps de pose.

Ainsi lorsque le seuil de qualité acceptable de la trame est dépassé, une nouvelle trame commence son intégration. Cette étape est répétée jusqu'à ce que le temps d'intégration de l'image finale soit terminé. En observant plus précisément le bougé dû au tremblement de la main, nous observons qu'il y a des grandes zones de temps qui présentent peu ou pas de bougé. Ainsi nous pouvons gagner du temps d'intégration pour certaines trames qui ne présentent pas un fort bougé visuel. Par la suite nous verrons différentes possibilités pour exploiter ce concept de segmentation du temps de pose afin de l'intégrer dans un système global de stabilisation. Nous appellerons cette surcouche, architecture avec consigne de qualité.

Cette nouvelle architecture basée sur les trames à intégration variable est présentée figure 81.

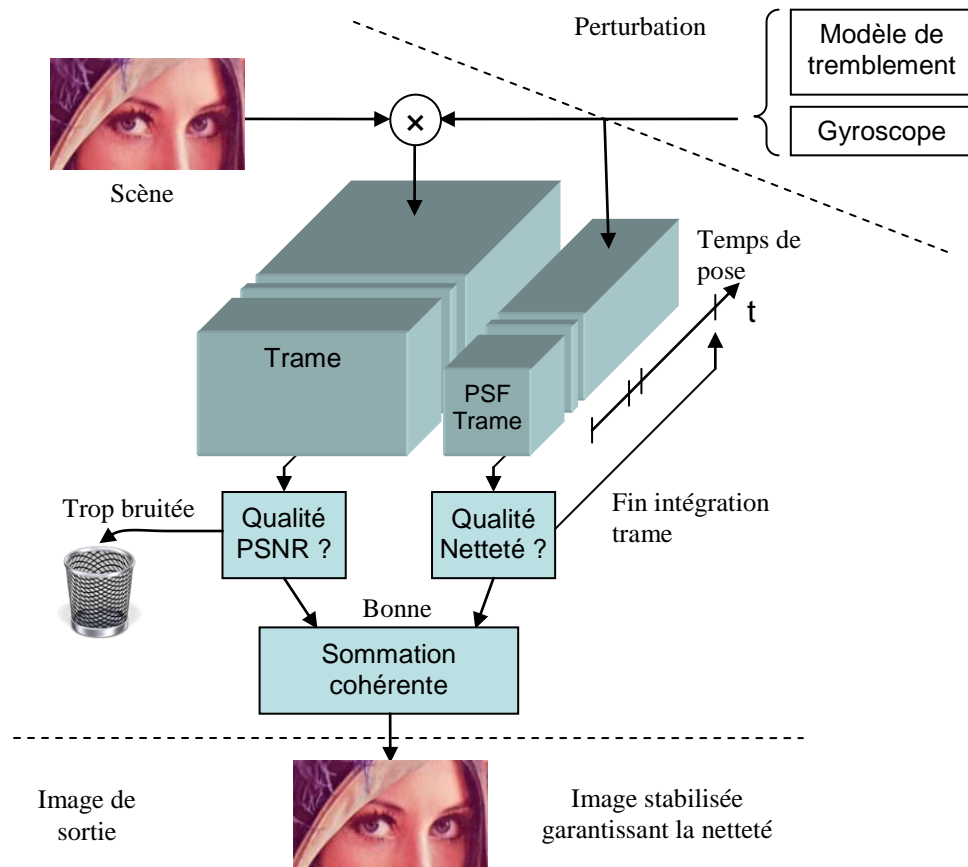


Figure 81. Détaille de l'architecture basée sur la qualité de service.

Cette architecture utilise en entrée les informations de perturbation mesurées par gyroscope dans le cas classique, dans notre cas en simulation nous avons utilisé les sorties de notre modèle de tremblement physiologique. Le modèle nous permet d'une part de construire la PSF des trames et d'autre part, par la convolution de la PSF du bougé par l'image, de simuler l'intégration de la scène sur le capteur en cours du temps de pose. A la fréquence d'échantillonnage des informations du bougé, la métrique de qualité va surveiller la qualité de chaque trame en cours d'intégration en calculant la valeur instantanée en JND de la PSF trame. Si la qualité de la trame devient inférieure à la consigne de qualité, l'intégration de cette trame sera interrompue. Et ainsi de suite, n trames seront intégrées avec des temps différents suivant les perturbations de bougé d'entrée avec la particularité d'avoir toute la même qualité de netteté. Avant l'étape de reconstruction de l'image finale, la qualité du PSNR de la trame peut être calculée afin de juger si la trame est trop bruitée ou non. Par exemple s'il y a eu un bougé important, la trame va avoir un temps d'intégration très faible donc si la trame contient beaucoup de bruit par rapport au signal utile (par exemple à cause du bruit de quantification), cette trame pourra être détruite car jugée non pertinente. Puis en utilisant les informations de la PSF trame, l'étape de sommation cohérente présentée section

3.2.2 est réalisée avec les trames retenues ayant une bonne qualité de PSNR. Un avantage de réaliser l'estimation de mouvement sur les PSF frame et non directement sur les trames est de réduire la complexité de l'estimateur car les dimensions des PSF frames sont beaucoup plus petites que les trames elles-mêmes.

Cette architecture peut être combinée à différents types d'architecture de stabilisation d'image. Par exemple, elle peut être utilisée dans une chaîne classique de stabilisation utilisant un système de stabilisation optique avec un gyroscope, la perturbation d'entrée sera le bougé résiduel. Dans le cas d'une architecture utilisant seulement le gyroscope, les informations de la perturbation d'entrée seront les données gyroscopiques.

3.3.2 Résultats et conclusion

Les tests ont été réalisés pour le cas du smartphone ayant une distance focale de 28 mm et une taille pixel de 1,4 μm . Cette fois un temps de pose long de $\frac{1}{2}$ seconde a été choisi. Lorsque des résultats moyennés sont présentés, ceux-ci sont calculés à partir des mêmes 100 bougés afin de rendre les résultats comparables. La consigne de qualité a été fixée à -3 JND pour garantir une bonne qualité d'image. La figure 82 montre les résultats de l'intégration variable des trames pour 10 prises de vue.

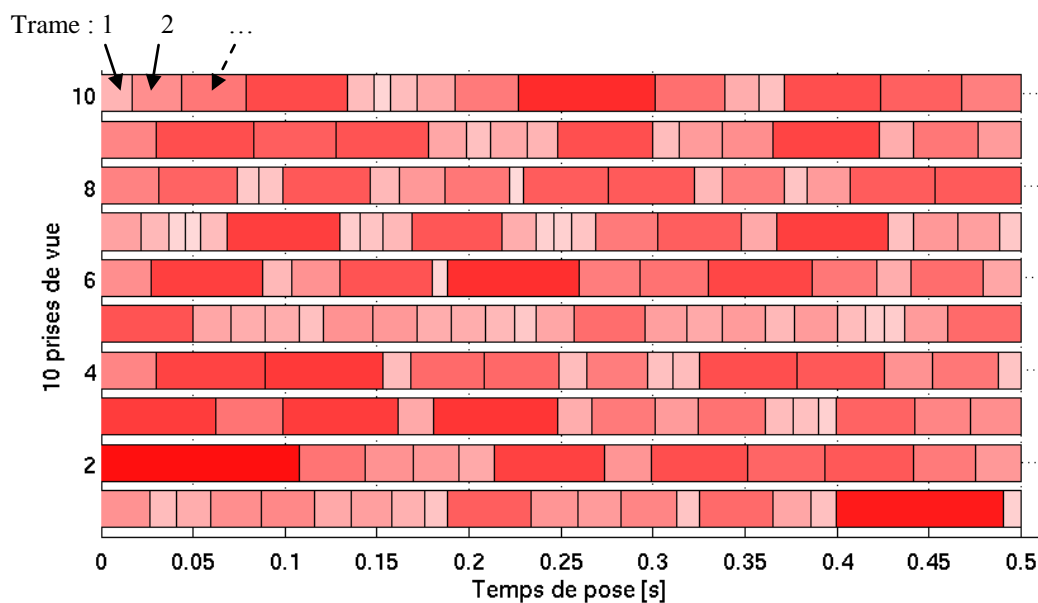


Figure 82. 10 prises de vue de $\frac{1}{2}$ seconde avec leurs trames ayant des temps d'intégration différents.

Sur cette figure nous pouvons remarquer que 2 prises de vue différentes, ne vont pas avoir le même nombre de trames ni les mêmes temps d'intégration trame. Ceci montre que le bougé a parfois un fort impact sur la qualité de l'image, caractérisé par les trames courtes, et parfois un faible impact sur la qualité d'image, caractérisé par les trames longues.

La figure 83 montre les résultats de qualité d'image en JND pour ces mêmes 10 prises de vue.

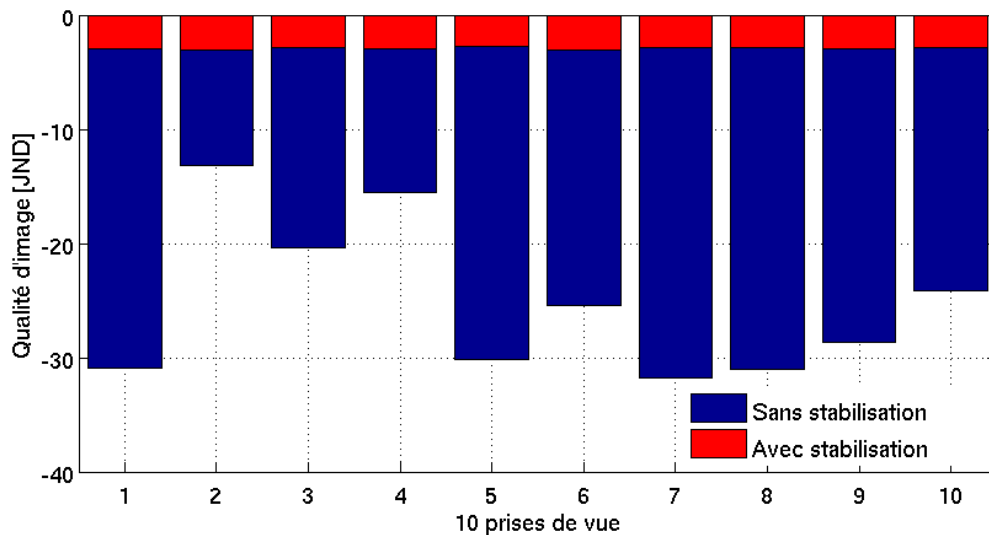


Figure 83. Qualité des images perçues en JND avec ou sans stabilisation pour 10 prises de vue.

Comme nous pouvons le remarquer sur la figure, la qualité de stabilisation est bien garantie malgré un temps de pose important de $\frac{1}{2}$ seconde. La moyenne de qualité obtenue des 100 images stabilisées est de -2, 9 JND avec un écart type de 0,1 JND. Quel que soit l'impact du bougé initial sur la perception, notre architecture de stabilisation arrive à garantir la qualité des images.

Pour juger visuellement des résultats de la stabilisation par notre architecture, la première prise de vue est présentée figure 84. Pour une bonne visualisation des images, de la même manière que dans la section 2.3.3.1, les images présentées par la suite ont une définition de 180 dpi. Ainsi une fois imprimée en couleur avec une définition en 180 dpi ou plus, elles devront être visualisées à 48,5 cm afin de retrouver les résultats inscrits en JND.

Sans stabilisateur d'image, il est quasiment impossible de prendre une prise de vue de $\frac{1}{2}$ seconde avec un smartphone comme le montre l'exemple présenté figure 84.a, l'image sera trop floue. Cette image est jugée sans intérêt par notre métrique qui lui donne une note de -31 JND.



Figure 84. Rendu visuel de l'architecture de stabilisation d'une prise de vue. (a) Image non stabilisée. (b) Image stabilisée (sans les sources de bruit parasites). (c) Image stabilisée (avec simulation du bruit). (d) Image de référence (scène). Pour visualiser les résultats correctement, attention à la résolution d'impression couleur (>180 dpi) ainsi qu'à la distance de visualisation (48,5 cm).

Grâce à notre architecture de stabilisation d'image nous arrivons à obtenir une image nette ayant comme qualité -2,95 JND. Si nous ignorons les différentes sources de bruit tels que le bruit photonique et le bruit de quantification, nous avons une image de bonne qualité (figure 84.b) légèrement moins bonne que l'image de référence (figure 84.d). Si nous prenons en compte les différentes sources de bruit avec les mêmes paramètres que dans la section précédente (3.2.5), nous obtenons encore une image de bonne qualité (figure 84.c) avec un bruit visible mais non gênant.

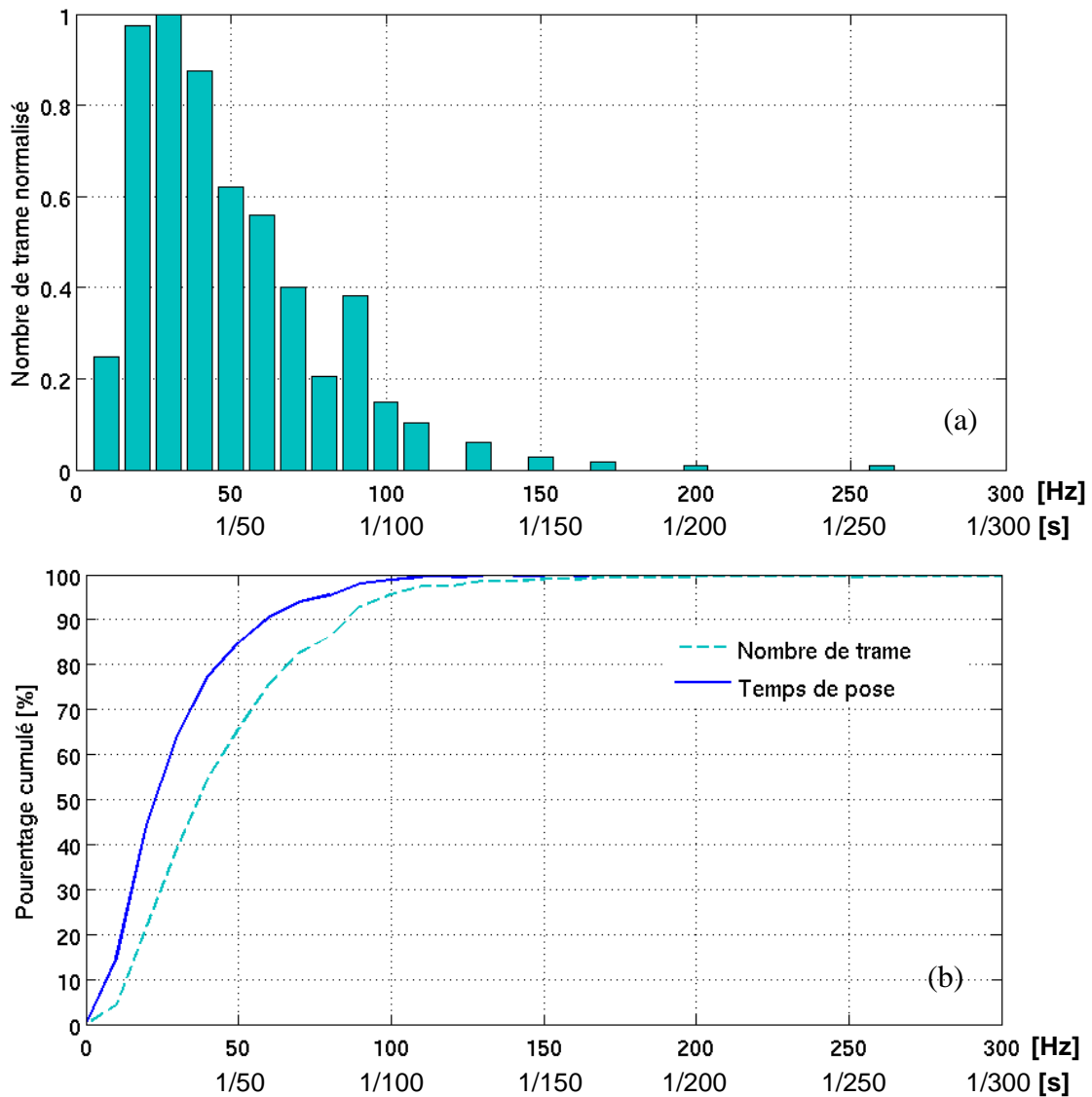


Figure 85. (a) Histogramme normalisé du nombre de trames suivant la cadence de l'imageur et leur temps d'intégration. (b) Pourcentage cumulé du nombre de trames et du temps de pose en fonction de la fréquence maximum de fonctionnement de l'imageur et du temps d'intégration des trames.

Maintenant, si nous regardons les fréquences requises des imageurs afin de générer les trames pour stabiliser les 100 bougés, nous avons la répartition fréquentielle des trames présentées figure 85. La figure 85.a présente l'histogramme normalisé des trames suivant leurs temps d'intégration correspondant à des fréquences de génération de celles-ci par l'imageur réparties tous les 10 Hz. Les fréquences utilisées pour générer les trames s'étendent de 5 à environs 300 Hz et sont plutôt nombreuses dans les basses fréquences avec un pic d'apparition autour de 30 Hz, soit 33 ms de temps d'intégration des trames.

Maintenant, si l'on calcule le pourcentage cumulé du nombre de trames en fonction du temps d'intégration (figure 85.b), plus de 90 % des trames sont supérieures à 11 ms soit une fréquence maximale imageur de 90 Hz et 50 % pour les trames supérieures à 25 ms soit une fréquence maximale imageur de 40 Hz.

Si l'on regarde les résultats de façon différente : au cours du temps de pose, une trame générée à une faible fréquence aura une grande période, donc elle contribuera fortement au temps de pose final de l'image. Inversement une trame générée à une grande fréquence aura une petite période, donc elle contribuera faiblement au temps de pose final de l'image. Dans cette idée, si on ramène les temps d'intégration des trames en pourcentage du temps de pose présenté figure 85.b, avec seulement une fréquence imageur fonctionnant à 23 Hz en moyenne ceci correspond à 50 % du temps de pose total. Si l'on cherche à avoir 90% de l'intégration finale de l'image en moyenne, la fréquence de fonctionnement pourra être inférieure à 60 Hz.

Dans cette optique, un moyen de relâcher la contrainte fréquentielle de fonctionnement des imageurs est de supprimer les trames ayant un faible temps d'intégration au profit des trames avec un temps d'intégration plus long comme présenté figure 86.

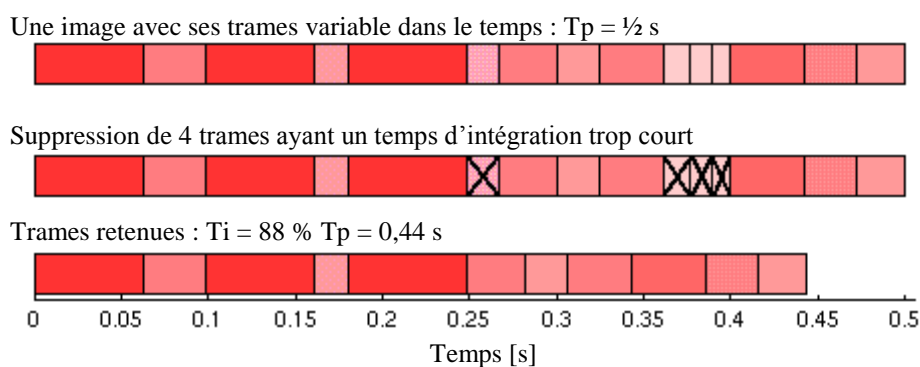


Figure 86. Principe de suppression des trames afin de relâcher la contrainte fréquentielle de l'architecture de stabilisation. Dans cet exemple, le temps d'intégration (T_i) final de l'image est de 88 % du temps de pose (T_p) requis de $\frac{1}{2}$ seconde.

Pendant l'intégration de chaque trame, la qualité intrinsèque des trames est calculée par la métrique de qualité. En fonction de la fréquence maximale du système, si une trame est trop peu intégrée c'est-à-dire qu'à ce moment-là il y a eu un bougé qui a rapidement dégradé la trame, cette trame sera simplement supprimée comme dans l'exemple présenté figure 86. Du coup à la fin du temps de pose, l'image ne sera pas complètement intégrée, une partie de l'exposition de l'image sera manquante.

Deux solutions se présentent au système pour compléter cette partie manquante :

- Augmenter le temps de pose : en continuant la pose jusqu'à atteindre 100% de l'exposition requise de l'image.
- Augmenter l'ISO : en arrêtant la pose et en la complétant par un gain global sur l'image.

Ces deux approches différentes sont illustrées figure 87.

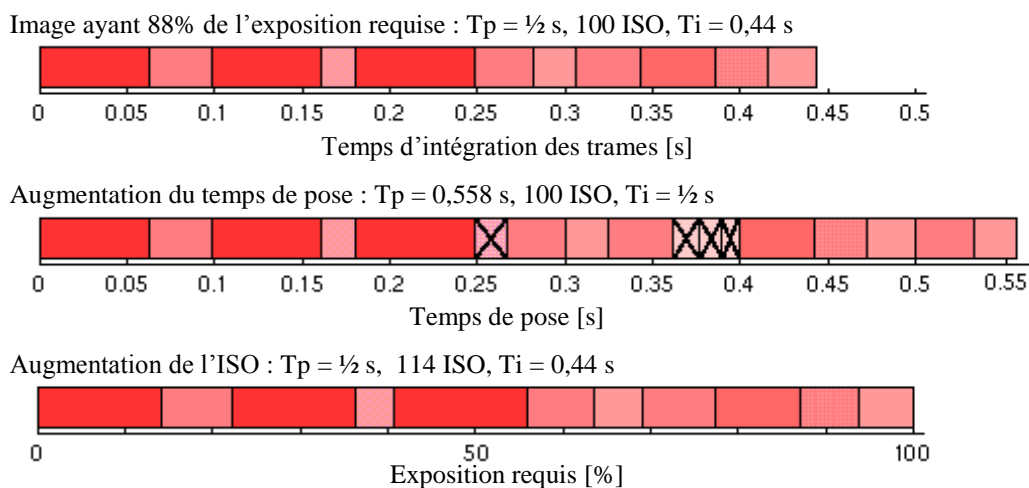


Figure 87. Solutions afin d'obtenir l'exposition requise pour avoir une image suffisamment lumineuse : par augmentation du temps de pose ou par augmentation de l'ISO.

L'augmentation du temps de pose va permettre d'obtenir ainsi l'exposition requise pour que l'image soit suffisamment lumineuse. Dans certains cas, l'augmentation du temps de pose peut être gênante, par exemple lorsqu'il y a du mouvement dans la scène. En conséquence la solution d'augmentation de l'ISO peut être une bonne alternative dans ce cas-là.

En reprenant l'exemple du premier bougé produisant un flou sur l'image estimé à -31 JND, les résultats visuels sont présentés dans le cas de l'augmentation du temps de pose (figure 88) et dans le cas de l'augmentation de l'ISO (figure 89).



Figure 88. Rendu de l'architecture de stabilisation d'image avec consigne de qualité fixée à -3 JND utilisant la technique de l'augmentation du temps de pose. (a) Image initiale comprenant toutes les trames. (b) Image avec seulement les trames générées à une fréquence inférieure à 60 Hz. (c) Image avec seulement les trames générées à une fréquence inférieure à 23 Hz. Pour visualiser les résultats correctement, attention à la résolution d'impression couleur (>180 dpi) ainsi qu'à la distance de visualisation (48,5 cm).

Au niveau de la netteté de l'image, dans tous les cas différents présentés figures 88 et 89, la qualité de netteté est bien respectée grâce à la consigne de qualité de l'architecture fixé à -3 JND. Dans le cas de l'augmentation du temps de pose (figure 88) nous pouvons observer une

légère réduction du bruit dans l'image liée à la sélection de trame de plus longues durées. Cette réduction de bruit est mesurée par l'augmentation du PSNR de l'image, elle contribue à l'amélioration de la qualité globale de l'image. La réduction de bruit vient du fait que lorsque l'on élimine les trames ayant des temps d'intégration très courts, on élimine aussi les trames ayant un plus fort bruit intrinsèque.



Figure 89. Rendu de l'architecture de stabilisation d'image avec consigne de qualité fixée à -3 JND utilisant la technique de l'augmentation de l'ISO. (a) Image initiale comprenant toutes les trames. (b) Image avec seulement les trames générées à une fréquence inférieure à 60 Hz. (c) Image avec seulement les trames générées à une fréquence inférieure à 23 Hz.

Au final l'image finale par l'addition de toutes les trames s'en trouve moins bruitée. Dans le cas de l'augmentation de l'ISO (figure 89) plus le système utilise des trames générées à basse fréquence, plus il est important d'augmenter l'ISO pour compenser le manque d'exposition. Le fait d'augmenter le gain dans l'image va augmenter le bruit dans l'image finale. Néanmoins comme le montre la figure 89.c comparée à la figure 89.a l'augmentation du bruit est seulement de 1,1 dB pour un gain ISO de 40 % permettant au système de fonctionner à une fréquence bien inférieure de 23 Hz.

La segmentation du temps d'intégration en trame variable grâce à la métrique de qualité nous permet d'envisager l'intégration future de cette nouvelle architecture de stabilisation au vu des performances actuelles des imageurs rapides présentés dans l'état de l'art section 3.1.4.

Ainsi la qualité de service concernant la netteté de l'image pourra satisfaire l'observateur quel que soit son type d'utilisation.

3.4 Conclusion sur les algorithmes de stabilisation priorisant la netteté

Les architectures de stabilisation présentées dans ce manuscrit permettent de répondre aux inconvénients de la stabilisation classique qui sont :

- Intégration plus difficile car utilisation de composants externes dans la stabilisation.
- Ne permet pas de garantir une qualité d'image donnée.

De même, les systèmes de stabilisation d'aujourd'hui présentés dans l'état de l'art ne permettent pas de garantir une netteté de l'image donnée quel que soit le type d'utilisation.

Dans ces travaux, nous avons présenté 3 architectures qui permettent de réduire le nombre de composant externe, elles seront dites économiques en composants externes :

- L'architecture de stabilisation sans gyroscope avec système de stabilisation optique.
- L'architecture complètement intégrée.
- L'architecture sans système de stabilisation optique mais avec gyroscope.

Nous présentons une synthèse des principales caractéristiques des différentes solutions en figure 90.

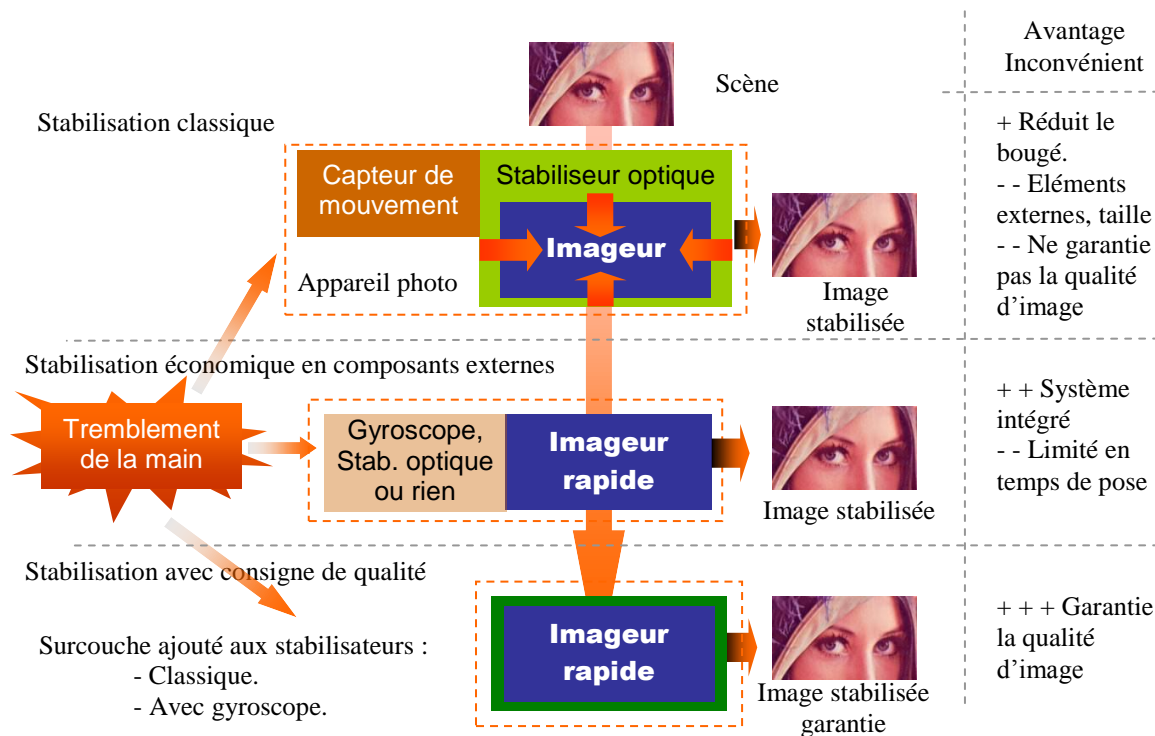


Figure 90. Synthèse des différentes architectures de stabilisation.

Ces architectures économiques permettent de garantir une netteté d'image dans les cas où la fréquence de génération des trames est atteinte, suivant les caractéristiques de l'appareil photo telles que la distance focale, le temps de pose et la taille des pixels. Néanmoins, selon la robustesse des algorithmes de stabilisation implémentés, les 2 architectures sans gyroscope pourront être limitées en temps de pose. En effet ce sont les performances d'estimation de mouvement qui limitent ces architectures par rapport au temps de pose. Les performances d'estimation de mouvement sont limitées par le niveau de bruit élevé interne dans les trames. Plus le temps de pose augmente, plus le nombre de trames utilisées pour la stabilisation est important, plus le bruit de chaque trame est important et donc plus l'estimation de mouvement sera difficile à réaliser.

Afin de relâcher les contraintes fréquentielles tout en garantissant la netteté de l'image, l'architecture avec consigne de qualité a été présentée. Cette architecture est proposée comme surcouche pour certaines architectures de stabilisation. Cette architecture permet aussi de réduire le bruit dans l'image en cherchant à maximiser le PSNR dans les trames qui contribuent à l'image finale.

Un comparatif des différentes architectures de stabilisation d'image est présenté tableau 9 suivant ces 5 critères :

- La faisabilité : dépend du type d'imageur, de sa cadence de fonctionnement et de la difficulté d'intégrer un tel système.
- Les performances de stabilisation : dépend de la qualité de stabilisation que peut garantir le système.
- La possibilité de faire de longs temps de pose ($> 1/2$ s) en gardant une bonne qualité de netteté.
- Le niveau d'intégration matériel : dépend du nombre de composants externes ainsi que de leur espace requis.
- La faible complexité : dépend de la complexité calculatoire de l'architecture de stabilisation.

Dans le tableau proposé ci-dessous, pour chaque architecture de stabilisation, les 3 types d'appareils sont présentés avec leurs différentes utilisations possibles. Les différentes utilisations notées dans le tableau sont grand angle, portrait et zoom ayant respectivement des distances focales autour de 28 mm, 100 mm et au-delà de 180 mm.

Pour le cas de la stabilisation classique, la complexité est faible car il n'y a pas besoin d'analyser l'image, un *digital signal processor* se charge de traiter les données gyroscopiques pour contrôler le système de stabilisation optique. Par contre au niveau de la faisabilité, intégrer un système de stabilisation optique n'est pas quelque chose de facile dans les smartphones comme en témoigne leurs récentes apparitions. Pour les appareils photos compacts et reflex, cette technique est largement répandue en conséquence jugée avec une forte faisabilité. Pour leur intégration, plus l'appareil est gros, plus leur intégration est simple. Concernant les performances de stabilisation garantissant la netteté des images, plus la distance focale augmente, plus le bougé sur l'image augmente donc plus les performances s'en trouvent réduites. Ceci est nuancé par la masse de l'appareil permettant de réduire l'impact du tremblement de la main sur l'appareil ainsi les appareils de masse plus lourde s'en sortent mieux. Pour la capacité à faire des longs temps de pose, seul avec un reflex en grand angle cela semble envisageable. Dans tous les cas, ce système de stabilisation n'arrive pas à garantir la qualité de netteté des images. Par l'ajout de la surcouche de l'architecture avec consigne de qualité sur la stabilisation classique, les performances de stabilisation sont augmentées car grâce à la consigne de qualité la qualité de la stabilisation est assurée. De plus

grâce à la sélection des trames ayant un faible bougé, l'utilisation de long temps de pose sera aussi envisageable.

Tableau 9. Comparatif des architectures de stabilisation d'image proposées ainsi que la stabilisation classique. Plus il y a d'étoiles, meilleure est la note. Les étoiles entourées s'ajoutent lorsque l'architecture avec consigne de qualité est ajoutée en surcouche.

| | Appareil | Focale | Faisabilité | Performance stabilisation | Long temps de pose | Intégration matériel | Faible complexité | Total |
|------------------------------------|-------------|------------------|----------------|---------------------------|--------------------|----------------------|-------------------|--------------------|
| Stabilisation classique | Smartphone | Grand angle | ☆☆ | ☆☆☆ | ☆☆☆ | ☆ | ☆☆☆☆ | 9 (12) |
| | Compact | Grand angle | ☆☆☆☆ | ☆☆☆ | ☆☆☆ | ☆☆ | ☆☆☆☆ | 11 (14) |
| | | Portrait Zoom | ☆☆☆☆ | ☆☆☆ | ☆☆☆ | ☆☆ | ☆☆☆☆ | 10 (14) 10 (14) |
| Reflex | Grand angle | ☆☆☆☆ | ☆☆☆ | ☆☆☆ | ☆☆☆ | ☆☆☆☆ | ☆☆☆☆ | 13 (15) |
| | Portrait | ☆☆☆☆ | ☆☆☆ | ☆☆☆ | ☆☆☆ | ☆☆☆☆ | ☆☆☆☆ | 12 (15) |
| | Zoom | ☆☆☆☆ | ☆☆☆ | ☆☆☆ | ☆☆☆ | ☆☆☆☆ | ☆☆☆☆ | 11 (15) |
| Architecture sans gyroscope | Smartphone | Grand angle | ☆☆ | ☆☆☆☆ | ☆ | ☆☆ | ☆ | 9 |
| | Compact | Grand angle | ☆☆ | ☆☆☆☆ | | ☆☆☆☆ | ☆ | 10 |
| | | Portrait Zoom | ☆☆ ☆☆ ☆ | ☆☆☆☆ | ☆ | ☆☆☆☆ | ☆ | 10 9 |
| Reflex | Grand angle | ☆☆☆☆ | ☆☆☆☆ | | ☆☆☆☆ | ☆ | 11 | |
| | Portrait | ☆☆☆☆ | ☆☆☆☆ | ☆ | ☆☆☆☆ | ☆ | 10 | |
| | Zoom | ☆☆ ☆☆ | ☆☆☆☆ | ☆ | ☆☆☆☆ | ☆ | 9 10 | |
| Architecture tout intégrée | Smartphone | Grand angle | ☆☆ | ☆☆☆☆ | ☆ | ☆☆☆☆ | ☆ | 10 |
| | Compact | Grand angle | ☆☆☆☆ | ☆☆☆☆ | ☆ | ☆☆☆☆ | ☆ | 11 |
| | | Portrait Zoom | ☆☆ ☆☆ ☆ | ☆☆☆☆ | ☆ | ☆☆☆☆ | ☆ | 10 9 |
| Reflex | Grand angle | ☆☆☆☆ | ☆☆☆☆ | | ☆☆☆☆ | ☆ | 11 | |
| | Portrait | ☆☆☆☆ | ☆☆☆☆ | ☆ | ☆☆☆☆ | ☆ | 11 | |
| | Zoom | ☆☆ ☆☆ | ☆☆☆☆ | ☆ | ☆☆☆☆ | ☆ | 10 | |
| Architecture avec gyroscope | Smartphone | Grand angle | ☆☆☆ | ☆☆☆☆ | ☆☆☆ | ☆☆☆☆ | ☆☆☆☆ | 12 (15) |
| | Compact | Grand angle | ☆☆☆☆ | ☆☆☆☆ | ☆☆☆ | ☆☆☆☆ | ☆☆☆☆ | 13 (15) |
| | | Portrait Zoom | ☆☆ ☆☆ ☆☆ | ☆☆☆☆ | ☆☆☆ | ☆☆☆☆ | ☆☆☆☆ | 12 (15) 11 (15) |
| Reflex | Grand angle | ☆☆☆☆ | ☆☆☆☆ | | ☆☆☆☆ | ☆☆☆☆ | ☆☆☆☆ | 13 (15) |
| | Portrait | ☆☆☆☆ | ☆☆☆☆ | ☆☆☆ | ☆☆☆☆ | ☆☆☆☆ | ☆☆☆☆ | 13 (15) |
| | Zoom | ☆☆ ☆☆ | ☆☆☆☆ | ☆☆☆ | ☆☆☆ | ☆☆☆☆ | ☆☆☆☆ | 12 (15) |

☆☆☆☆ = Passage de 1 à 3 étoiles grâce à la surcouche de l'architecture avec consigne de qualité.

Pour le cas de l'architecture sans gyroscope mais avec système de stabilisation optique, la faisabilité sera difficile car il faudra utiliser des imageurs rapides. Plus la distance focale sera importante, plus l'imageur devra être rapide. Au niveau de l'intégration matériel, seulement

dans le cas du smartphone il sera techniquement plus difficile d'intégrer le stabilisateur optique. Dans le cas d'un imageur assez rapide (>250 trames/s pour $f = 28$ mm), les performances de stabilisation seront élevées (-3 JND) car le niveau de qualité sera garanti. Mais pour les temps de pose long, comme le nombre de trames requis sera important, le système n'arrivera plus à réaliser l'estimation sur les trames car elles auront un PSNR trop faible pour réaliser l'estimation de mouvement comme démontré dans la section 3.2.5. De plus comme l'opération d'estimation de mouvement est réalisée sur les trames, elle est gourmande en calculs et donc complexe. Si l'on souhaite ajouter la surcouche du stabilisateur avec consigne de qualité, il faudra reconstruire la PSF trame en temps réel et donc avoir une première étape d'estimation de mouvement à la cadence élevée. Du coup nous ne pourrons pas relâcher les contraintes de cadence de l'imageur.

Pour le cas de l'architecture tout intégré, nous avons quasiment les mêmes contraintes que l'architecture sans gyroscope mais avec système de stabilisation optique. L'avantage principal est l'intégration maximale du système car seul l'imageur est requis pour réaliser la stabilisation. Comme nous l'avons vu figure 75, les architectures avec sommation cohérente ont des contraintes de cadence plus faibles par rapport à l'architecture sans gyroscope mais avec stabilisation optique, de ce fait la faisabilité d'imageur rapide s'en trouve légèrement augmentée. De même que pour l'architecture sans gyroscope si l'on souhaite ajouter une surcouche de stabilisation avec consigne de qualité, nous allons retrouver le même problème pour estimer la PSF trame en cours d'intégration, donc les contraintes de cadence ne seront pas relâchées.

Pour le cas de l'architecture avec seulement le gyroscope, nous allons avoir les mêmes faisabilités, performances, problèmes pour les longs temps de pose que l'architecture complètement intégrée. Pour l'intégration matérielle, seul un gyroscope est requis donc même pour le cas des smartphones ceci n'est plus une forte contrainte car aujourd'hui, ils sont toujours présents dans ce type d'appareil. Grâce à la réalisation de l'estimation de mouvement réalisé directement sur les PSF trames, la complexité d'estimation de mouvement s'en trouve fortement réduite. Si l'on ajoute la surcouche de stabilisation avec consigne de qualité, grâce à l'estimation de la netteté de chaque trame pendant leur intégration, nous pouvons relâcher fortement la cadence de l'imageur rapide et ainsi rendre sa faisabilité plus aisée. De plus, grâce à la sélection des trames ayant un faible bougé, les longs temps de pose seront réalisables.

Pour conclure sur les architectures proposées ces architectures répondent à différents besoins. Si le critère d'intégration est maximal et que pour des raisons de place ou de réduction de coût, il n'est pas envisageable d'avoir un gyroscope MEMS dans le système, l'architecture de stabilisation tout intégrée semble être un très bon compromis.

Si le critère d'intégration est maximal et qu'un gyroscope MEMS est disponible ou ne présente pas un surcoût, l'architecture avec gyroscope utilisant la surcouche de l'architecture avec consigne de qualité sera préférée.

Si la performance de stabilisation recherchée est maximale, l'ajout de la surcouche de l'architecture avec consigne de qualité sur les architectures de stabilisation classique permettra de garantir la qualité de stabilisation des images même pour des long temps d'exposition.

Conclusions et Contributions

Dans cette thèse, nous avons proposé des architectures de stabilisation d'image permettant de garantir la qualité de celle-ci quel que soit l'utilisation de l'utilisateur. En introduction nous avons vu qu'aujourd'hui nous arrivons à deux situations conflictuelles.

D'un côté, nous observons une tendance à l'accentuation du flou de bougé causé par :

- La réduction des dimensions des modules photographique par la réduction de l'optique, de l'imageur ainsi que la taille des pixels de l'imageur.
- La réduction de la masse des appareils qui les rendent plus sensibles aux perturbations externes tel que le bougé, en particulier celles hautes fréquences.
- La démocratisation des appareils photo qui entraine l'utilisation de ceux-ci dans toutes les situations sans prendre de précaution particulière pour limiter le bougé, en particulier dans des conditions de faible luminosité.

De l'autre côté, l'utilisateur final est de plus en plus exigeant en termes de qualité d'images, notamment avec une exigence plus élevée en termes de netteté.

Les différentes solutions de stabilisation de l'état de l'art qui visent à réduire le flou de bougé, ont été présentées dans la section 3.1.2. Nous en avons montré les limites suivant le type d'architecture :

- Elles peuvent être constitués des modules de dimensions importantes car requérant des éléments externes. Cela limite leur intégration dans les téléphones portables et entraine un coût supérieur.
- Elles ne tiennent pas en compte de la netteté finale souhaitée par l'utilisateur, car elles ne permettent pas de garantir la qualité de netteté de sortie de l'image.

Dans cette thèse nous avons proposé différents types d'architectures qui permettent de répondre aux différentes demandes d'intégration ainsi qu'une architecture utilisé en surcouche

des autres architectures permettant de garantir une qualité de netteté d'image quelle que soit la situation.

Pour développer ces différentes architectures, nous nous sommes basés sur deux briques de base que nous allons rappeler et discuter avant d'arriver à la discussion finale sur les architectures de stabilisation proposées.

Modèle de tremblement physiologique

Une première contribution de cette thèse est l'apport d'un modèle de génération de flous de bougé basé sur la physiologie du tremblement de la main présentée dans le chapitre 1. A notre connaissance, c'est le premier modèle de bougé qui réunit les observations faites dans les domaines statistique, temporel et fréquentiel. Nous avons démontré que notre modèle de tremblement est bien plus fidèle que les modèles généralement utilisés que sont le modèle de bougé linéaire et le modèle de marche aléatoire. Notre modèle présente néanmoins une limitation temporelle : il est valide pour des temps de pose inférieurs à 1 seconde. Au-delà de ce temps, notre modèle ne prend pas en compte les contre-réactions volontaires du photographe face à sa propre dérive de bougé. Pour répondre à ce problème, il serait intéressant de mener une étude comportementale sur les contre-réactions volontaires du photographe afin de modéliser l'impact des réactions du photographe pour des temps de poses de plusieurs secondes voir de plusieurs minutes. Il est probable que cette contre réaction se traduise par des composantes très basse fréquences par rapport aux tremblements de la main et pourrait être ajoutée au modèle de tremblement que nous proposons. L'avantage d'avoir un tel modèle serait d'aider à développer un système de stabilisation pouvant se passer complètement de trépied même pour des temps de pose supérieurs à plusieurs secondes.

Comme notre modèle de tremblement est basé sur la densité spectrale de puissance des tremblements de la main, pour des tremblements non physiologiques mais pathologiques il sera possible de simuler ces types de tremblement en injectant la densité spectrale de puissance des différents type de pathologie. Ceci permettrait de dimensionner ou de rendre adaptatif des architectures de stabilisation d'image même pour des personnes ayant un tremblement pathologique, plus important que la moyenne.

Métrique de qualité d'image

Une seconde contribution de cette thèse est l'établissement d'une relation quantitative entre la métrique de qualité d'image (JND) et « l'amplitude » du flou définie à partir de sa PSF. Nous avons démontré que cette relation permet de prédire avec précision l'impact sur la

qualité perçue des flous résultants de mouvements linéaires et d'une façon correcte des flous quelconques. Pour cela nous avons mis en place deux expériences psycho-visuelles, une première par expérience de vote forcé entre deux images de qualité de flou proche afin d'en déduire une règle de qualité d'image. Une deuxième expérience par règle d'ajustement nous a permis de mesurer des flous quelconques. La relation que nous avons ainsi définie permet une bonne prédiction pour la majorité des flous ordinaires mais présente une limite pour les flous de type dédoublé comme discuté dans la section 2.3.2.2. De ce fait cette métrique peut être en partie améliorée afin de prendre en compte de ces cas particuliers. Il faut cependant noter que la relation que nous avons établie sous-estime la qualité des images pour ces cas et donc garantit une limite inférieure de qualité, soit autrement dit la métrique garantit le pire cas.

L'estimation de la qualité de netteté pourrait être aussi utilisée pour estimer la qualité des images finales. Ceci serait possible par l'utilisation d'un algorithme mesurant la PSF de l'image de manière fidèle. Puis grâce à la métrique de qualité, la netteté de l'image pourrait être retrouvée. Et ainsi par exemple pour des systèmes tout intégrés ayant seulement un imageur à la fin de l'acquisition de l'image une mesure de netteté pourrait être réalisée. Ceci permettrait d'indiquer au photographe la qualité de netteté de la photo afin de lui suggérer de reprendre une photo ou non pour obtenir une image plus nette. Cette solution peut être aussi envisagée lorsque la PSF de l'image finale est disponible, par exemple pour un système de stabilisation classique en calculant la PSF résiduel provenant des bougés résiduels.

Dans nos travaux, cette métrique nous a permis de prédire la qualité des images perçues soumises aux flous de bougé en calculant leur netteté. Par cette connaissance de la netteté des images nous avons pu d'une part quantifier les performances des algorithmes de stabilisation et d'autre part, développer des algorithmes de stabilisation qui tiennent compte de la qualité d'image en cours de stabilisation.

Architecture de stabilisation d'image

Enfin cette thèse contribue à la démonstration d'architecture de stabilisation d'image permettant de garantir à l'utilisateur un niveau de qualité de netteté et suivant les cas de réduire le nombre de composants externes.

Nous avons présenté dans la section 3.2.3.1 une première architecture de stabilisation avec système de stabilisation optique qui permet de se passer du gyroscope. Cette architecture s'appuie sur une estimation de mouvement réalisée sur les trames générées au court du temps de pose. L'estimation de mouvement permet de recréer la consigne de contre réaction pour le

système de stabilisation d'optique. Afin d'améliorer la stabilisation nous avons mis en place un prédicteur. Nous pensons que par le développement d'un meilleur prédicteur il est possible d'encore améliorer les performances de stabilisation de cette architecture.

Une deuxième architecture de stabilisation sans gyroscope ni système de stabilisation optique a été présentée section 3.2.3.2. Cette architecture de stabilisation présente l'avantage d'être la plus intégrée de toute.

Ensuite une troisième architecture de stabilisation avec gyroscope mais sans système de stabilisation a été présentée section 3.2.3.3. Par rapport au deux précédentes, le fait d'avoir directement les données gyroscopiques permet de réduire la complexité due à l'estimation de mouvement entre les trames afin de réaliser la stabilisation de l'image. Donc si le critère d'intégration est fort et que les données gyroscopiques sont disponibles dans le système, cette architecture sera préférée.

Enfin une architecture utilisée en surcouche a été présentée section 3.3. Cette architecture s'appuie sur un nouveau paradigme, la segmentation du temps de pose, pour garantir la qualité de l'image. Cette surcouche permet de relâcher les contraintes de cadence des imageurs rapides car les trames ne sont plus générées à une fréquence fixée mais sont générées suivant l'impact du bougé dans chaque trame. De plus, par la sélection des trames, les longs temps de pose sont rendu possibles grâce à cette nouvelle architecture.

Vers un stabilisateur d'image idéal ?

Grâce aux différentes architectures proposées, quelle serait le stabilisateur d'image idéal de demain ? En premier lieu, nous pensons que ce serait une solution complètement intégrée où le module photographique réalise intégralement la fonction de stabilisation. L'avantage d'avoir un seul module est la simplicité d'intégration pour le fabricant car il n'a plus besoin d'intégrer différents composants externes qui communiquent entre eux. Ceci permet aussi d'atteindre un niveau d'intégration maximal suivant le type de réalisation. Ensuite, c'est un système qui permettrait de garantir la stabilisation des images quel que soit la prise de vue même pour des temps de pose supérieurs à la seconde.

Pour atteindre le plus haut niveau de performance, comme nous l'avons vu, le système de détection le plus efficace est un gyroscope dédié à l'estimation du bougé du capteur. En effet réaliser l'estimation sur l'image demande un coût important de calcul et surtout ne permet pas d'avoir un grand nombre de trames à cause de leurs faibles PSNR, limitant le temps de pose final. Cependant le gyroscope pourrait être intégré sur la même puce que l'imageur si les

techniques hybrides MEMS et circuit rendent avantageuses la possibilité d'avoir un seul et même circuit. Pour obtenir la plus grande surface dédiée aux photodiodes et embarquer la stabilisation sur l'imageur, la technologie 3D des imageurs pourrait être utilisée. Une première couche serait dédiée aux photodiodes permettant d'obtenir un *fill factor* de 100 % afin de capter le plus de luminosité possible. Pour atteindre un premier niveau de stabilisation, un système de stabilisation optique type *liquid lens*, serait intégré utilisant directement les informations de bougé du gyroscope MEMS permettant ainsi de réduire le bougé en gardant un fort niveau d'intégration. Puis pour garantir la qualité de stabilisation, en dessous des photodiodes serait intégrée une partie mémorisation afin de stocker une ou plusieurs trames avant leurs additions. Le temps d'intégration de chaque trame serait déterminé par la consigne de qualité, réalisée par la métrique de qualité sur les PSF trame du bougé résiduelles. Permettant ainsi de garantir la qualité finale de stabilisation comme vue dans la section 3.3. Ensuite sous la mémoire, un étage de circuit de contrôle permettrait de réaliser la sommation cohérente avec les trames arrivant les unes à la suite des autres. Ainsi l'image accumulée serait une image stabilisée garantissant la netteté. Cette réalisation pourrait faire l'objet de futurs travaux de recherche.

Bibliographie

- [1] Celebrating 30 Years of Mobile Phones : Thank You NTT of Japan, <http://communities-dominate.blogs.com/brands/2009/11/celebrating-30-years-of-mobile-phones-thank-you-ntt-of-japan.html>.
- [2] How many photos have ever been taken ?, <http://blog.1000memories.com/94-number-of-photos-ever-taken-digital-and-analog-in-shoebox>.
- [3] R. Oudghiri, L. Brunet, and S. de Barthes, “Le cycle de vie d’une photo à l’ère numérique,” tech. rep., Ipsos, 2011.
- [4] J. Hémon, “Les chiffres officiels 2011,” tech. rep., Observatoire des Professions de l’image, 2012.
- [5] “Photography – electronic still-picture cameras – determination of iso speed,” 1998. ISO Standard 12232.
- [6] L. Gimena, “Exposure value in photography. a graphics concept map proposal,” pp. 256–269, 2004.
- [7] F. Xiao, J. Farrell, P. Catrysse, and B. Wandell, “Mobile imaging: The big challenge of the small pixel,” vol. 7250, 2009.
- [8] M. Okincha, “Omnibsi technology backgrounder,” June 2009.
- [9] B. W. Keelan, *Handbook of image quality: characterization and prediction*. Optical engineering, Marcel Dekker, 2002.
- [10] A. Silverstein, F. Xiao, and J. Farrell, “Camera-motion and effective spatial resolution,” pp. 33–36, May 2006.
- [11] J. Hémon, “Les chiffres officiels 2006,” tech. rep., Observatoire des Professions de l’image, 2007.
- [12] H. Furukawa and A. Tajima, “Image stabilizing optical system having a variable prism,” 1974. US Patent 3942862.
- [13] Vibration Reduction (VR) Technology, <http://imaging.nikon.com/history/scenes/16/index.htm>.
- [14] M. Ayat, S. Sh, and S. Nasiri, “Advances in mems gyroscopes enable ois in camera phones,” 2010.
- [15] G. Deuschl, P. Bain, and M. Brin, “Consensus statement of the movement disorder society on tremor,” *Movement Disorders*, vol. 13, no. S3, pp. 2–23, 1998.
- [16] T. A. Zesiewicz and R. A. Hauser, “Phenomenology and treatment of tremor disorders,” *Neurologic Clinics*, vol. 19, no. 3, pp. 651–680, 2001.

- [17] S. Smaga, "Tremor," Oct. 2003.
- [18] J. Raethjen, M. Lindemann, M. Dümpelmann, R. Wenzelburger, H. Stolze, G. Pfister, C. E. Elger, J. Timmer, and G. Deuschl, "Corticomuscular coherence in the 6-15 hz band: Is the cortex involved in the generation of physiologic tremor?," *Experimental Brain Research*, vol. 142, no. 1, pp. 32–40, 2002.
- [19] J. Timmer, M. Lauk, W. Pflieger, and G. Deuschl, "Cross-spectral analysis of physiological tremor and muscle activity: 2 application to synchronized electromyogram," *Biological Cybernetics*, vol. 78, no. 5, pp. 359–368, 1998.
- [20] J. Raethjen, F. Pawlas, M. Lindemann, R. Wenzelburger, and G. Deuschl, "Determinants of physiologic tremor in a large normal population," *Clinical Neurophysiology*, vol. 111, no. 10, pp. 1825–1837, 2000.
- [21] M. M. Sturman, D. E. Vaillancourt, and D. M. Corcos, "Effects of aging on the regularity of physiological tremor," *Journal of Neurophysiology*, vol. 93, no. 6, pp. 3064–3074, 2005.
- [22] B. Carignan, J. F. Daneault, and C. Duval, "Quantifying the importance of high frequency components on the amplitude of physiological tremor," *Experimental Brain Research*, vol. 202, pp. 299–306, 2010.
- [23] B. J. Davis and J. O'Connell, "Shoulder, elbow and wrist components of physiologic tremor amplitude as measured using a laser penlight," *European Neurology*, vol. 43, no. 3, pp. 152–154, 2000.
- [24] B. Carignan, J. F. Daneault, and C. Duval, "The amplitude of physiological tremor can be voluntarily modulated," *Experimental Brain Research*, vol. 194, no. 2, pp. 309–316, 2009.
- [25] S. Spieker, P. Loschmann, C. Jentgens, A. Boose, T. Klockgether, and J. Dichgans, "Tremorolytic activity of budipine: A quantitative study with long-term tremor recordings," *Clinical Neuropharmacology*, vol. 18, no. 3, pp. 266–272, 1995.
- [26] A. Chwaleba, J. Jakubowski, and K. Kwiatos, "The measuring set and signal processing method for the characterization of human hand tremor," in *CAD Systems in Microelectronics, 2003. CADSM 2003. Proceedings of the 7th International Conference. The Experience of Designing and Application*, pp. 149–154, feb. 2003.
- [27] G. Deuschl, J. Raethjen, M. Lindemann, and P. Krack, "The pathophysiology of tremor," *Muscle and Nerve*, vol. 24, no. 6, pp. 716–735, 2001.
- [28] R. J. Elble, R. Sinha, and C. Higgins, "Quantification of tremor with a digitizing tablet," *Journal of Neuroscience Methods*, vol. 32, no. 3, pp. 193–198, 1990.
- [29] B. Golik, *Development of a Test Method for Image Stabilizing Systems*. PhD thesis, Cologne University of Applied Sciences, 2006.
- [30] J. Keogh, S. Morrison, and R. Barrett, "Augmented visual feedback increases finger tremor during postural pointing," *Experimental Brain Research*, vol. 159, pp. 467–477, 2004.
- [31] R. J. Elble, "Characteristics of physiologic tremor in young and elderly adults," *Clinical Neurophysiology*, vol. 114, no. 4, pp. 624–635, 2003.
- [32] J. A. Gallego, E. Rocon, J. O. Roa, J. C. Moreno, and J. L. Pons, "Real-time estimation of pathological tremor parameters from gyroscope data," *Sensors*, vol. 10, no. 3, pp. 2129–2149, 2010.

- [33] K. A. Mann, F. W. Wernere, and A. K. Palmer, "Frequency spectrum analysis of wrist motion for activities of daily living," *Journal of Orthopaedic Research*, vol. 7, no. 2, pp. 304–306, 1989.
- [34] C. N. Riviere, R. S. Rader, and N. V. Thakor, "Adaptive cancelling of physiological tremor for improved precision in microsurgery," *Biomedical Engineering, IEEE Transactions*, vol. 45, pp. 839–846, July 1998.
- [35] C. N. Riviere, W. T. Ang, and P. K. Khosla, "Toward active tremor canceling in handheld microsurgical instruments," *IEEE Transactions on Robotics and Automation*, vol. 19, no. 5, pp. 793–800, 2003.
- [36] J. Zhang and F. Chu, "Real-time modeling and prediction of physiological hand tremor," vol. 5, pp. 645–648, 2005.
- [37] R. N. Stiles and J. E. Randall, "Mechanical factors in human tremor frequency.," *Journal of applied physiology*, vol. 23, no. 3, pp. 324–330, 1967.
- [38] S. Morrison and K. M. Newell, "Postural and resting tremor in the upper limb," *Clinical Neurophysiology*, vol. 111, no. 4, pp. 651–663, 2000.
- [39] B. Koster, M. Lauk, J. Timmer, T. Winter, B. Guschlbauer, F. X. Glocker, A. Danek, G. Deuschl, and C. H. Lucking, "Central mechanisms in human enhanced physiological tremor," *Neuroscience Letters*, vol. 241, no. 2-3, pp. 135–138, 1998.
- [40] F. Xiao, J. Pincetti, G. John, and K. Johnson, "Camera-motion and mobile imaging," vol. 6502, 2007.
- [41] "Understanding image sharpness," white paper, LumoLabs, 2011.
- [42] E. M. Or and D. Pundik, "Hand motion and image stabilization in hand-held devices," *Consumer Electronics, IEEE Transactions on*, vol. 53, pp. 1508–1512, Nov. 2007.
- [43] K. Forbes and E. Fiume, "An efficient search algorithm for motion data using weighted pca," pp. 67–76, 2005.
- [44] S. Kazi, A. As'array, M. Z. M. Zain, M. Mailah, M. Hussein, and I. Z. M. Darius, "Biodynamic response of the human hand-arm models under vibration," in *Proceedings of the 10th WSEAS International Conference on Automatic Control, Modelling & Simulation*, pp. 237–242, 2008.
- [45] H. Bruneel., S. Wittevrongel, T. Cherian, S. Rakheja, and R. B. Bhat, "An analytical investigation of an energy flow divider to attenuate hand-transmitted vibration," *International Journal of Industrial Ergonomics*, vol. 17, no. 6, pp. 455–467, 1996.
- [46] A. F. Pop and M. Arghir, "Hand-arm simulation, under the vibration action, with vibration flow divider," in *Automation Quality and Testing Robotics (AQTR), 2010 IEEE International Conference*, vol. 2, pp. 1–4, May 2010.
- [47] J. Z. Zhang, B. T. Price, R. D. Adams, K. Burbank, and T. J. Knaga, "Detection of involuntary human hand motions using empirical mode decomposition and hilbert-huang transform," pp. 157–160, 2008.
- [48] B. Pillman and D. Jasinski, "Camera exposure determination based on a psychometric quality model," *Journal of Signal Processing Systems*, pp. 1–12, 2011.
- [49] F. Spitzer, *Principles of Random Walk*. Springer, 1976.
- [50] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*. 1991.

- [51] E. Friedmann, “Etude d’un système de capture du mouvement multimodal combinant accéléromètres, magnétomètres et capteurs images,” 2004. Mémoire DRT.
- [52] “Imaging apparatus, and method and device for shake correction in imaging apparatus,” 2005. US patent 6940542.
- [53] J. B. Reppas, A. M. Dale, M. I. Sereno, and R. B. Tootell, “La vision, une perception subjective : A la (re)découverte de la panoplie d’aires visuelles du cortex,” 1996.
- [54] S. F. Ray, *Applied photographic optics: Lenses and optical systems for photography, film, video, electronic and digital imaging*. 2002.
- [55] G. Smith and D. A. Atchison, *The eye and visual optical instrument*. New York: Cambridge University Press., 1997.
- [56] G. B. Arden, “Le standard de mesure de l’acuité visuelle,” vol. 11, no. 11, pp. 779–792, 1988.
- [57] F. W. Campbell and D. G. Green, “Optical and retinal factors affecting visual resolution,” vol. 181, pp. 576–593, December 1965.
- [58] J. Mannos and D. Sakrison, “The effects of a visual fidelity criterion of the encoding of images,” *Information Theory, IEEE Transactions on*, vol. 20, pp. 525–536, jul. 1974.
- [59] P. Le Callet and D. Barba, *Critères objectifs avec référence de qualité visuelle des images couleur*. PhD thesis, 2001.
- [60] T. Ho Phuoc, N. Guyader, F. Landragin, and A. Guérin Dugué, “When viewing natural scenes, do abnormal colors impact on spatial or temporal parameters of eye movements?,” *Journal of Vision*, vol. 12, no. 2, 2012.
- [61] “Subjective audiovisual quality assessment methods for multimedia applications,” 1998. ITU-T Recommendation P.911.
- [62] “Methodology for the subjective assessment of the quality of television pictures,” 2002. ITU-T Recommendation BT.500-11.
- [63] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, “Live image quality assessment database release 2,” 2005. <http://live.ece.utexas.edu/research/quality>.
- [64] N. Ponomarenko, V. Lukin, K. Egiazarian, J. Astola, M. Carli, and F. Battisti, “Color image database for evaluation of image quality metrics,” pp. 403–408, 2008.
- [65] E. C. Larson and D. M. Chandler, “Most apparent distortion: Full-reference image quality assessment and the role of strategy,” *Journal of Electronic Imaging*, vol. 19, no. 1, 2010.
- [66] P. Le Callet and F. Atrousseau, “Subjective quality assessment irccyn/ivc database,” 2005. <http://www2.irccyn.ec-nantes.fr/ivcdb/>.
- [67] D. M. Chandler and S. S. Hemami, “A57 dataset,” 2007. <http://foulard.ece.cornell.edu/dmc27/vsnr/vsnr.html>.
- [68] Y. Horita, Y. Kawayoke, and Z. Parvez Sazzad, “Image quality evaluation database,” <http://mict.eng.u-toyama.ac.jp/mictdb.html>.
- [69] U. Engelke, H. J. Zepernick, and M. Kusuma, “Wireless imaging quality database,” <http://www.bth.se/tek/rcg.nsf/pages/wiq-db>.

- [70] P. G. Engeldrum, *Psychometric Scaling: A Toolkit for Imaging Systems Development*. Imcotek Press, 2000.
- [71] A. Parducci and L. F. Perrett, "Category rating scales: Effects of relative spacing and frequency of stimulus values," *Journal of Experimental Psychology*, vol. 89, no. 2, pp. 427–452, 1971.
- [72] "Photography - psychophysical experimental methods for estimating image quality," 2005. Tech. Rep. ISO 20462.
- [73] E. W. Jin, B. W. Keelan, J. Chen, J. B. Phillips, and Y. Chen, "Softcopy quality ruler method: Implementation and validation," vol. 7242, 2009.
- [74] S. J. Erasmus and K. C. A. Smith, "An automatic focusing and astigmatism correction system for the sem and ctem," *Journal of Microscopy*, vol. 127, no. 2, pp. 185–199, 1982.
- [75] L. Firestone, K. Cook, K. Culp, N. Talsania, and K. Preston, "Comparison of autofocus methods for automated microscopy," *Cytometry*, vol. 12, no. 3, pp. 195–206, 1991.
- [76] C. F. Batten, *Autofocusing and Astigmatism Correction in the Scanning Electron Microscope*. PhD thesis, University of Cambridge, 2000.
- [77] N. Ng Kuang Chern, P. A. Neow, and M. H. Ang, "Practical issues in pixel-based autofocusing for machine vision," in *Robotics and Automation, 2001. Proceedings 2001 ICRA. IEEE International Conference*, vol. 3, pp. 2791–2796, 2001.
- [78] X. Marichal, W. Y. Ma, and H. J. Zhang, "Blur determination in the compressed domain using dct information," in *Image Processing, 1999. ICIP 99. Proceedings. 1999 International Conference*, vol. 2, pp. 386–390, 1999.
- [79] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "A no-reference perceptual blur metric," vol. 3, pp. 57–60, 2002.
- [80] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "Perceptual blur and ringing metrics: Application to jpeg2000," *Signal Processing: Image Communication*, vol. 19, no. 2, pp. 163–172, 2004.
- [81] J. Caviedes and F. Oberti, "A new sharpness metric based on local kurtosis, edge and energy information," *Signal Processing: Image Communication*, vol. 19, no. 2, pp. 147–161, 2004.
- [82] S. Winkler, "Visual fidelity and perceived quality: Towards comprehensive metrics," in *in Proc. SPIE*, pp. 114–125, 2001.
- [83] E. Ong, W. Lin, Z. Lu, X. Yang, S. Yao, F. Pan, L. Jiang, and F. Moschetti, "A no-reference quality metric for measuring image blur," in *Signal Processing and Its Applications, 2003. Proceedings. Seventh International Symposium on*, vol. 1, pp. 469–472, July 2003.
- [84] E. Ong, W. Lin, Z. Lu, S. Yao, X. Yang, and L. Jiang, "No-reference jpeg-2000 image quality metric," in *Multimedia and Expo, 2003. ICME '03. Proceedings. 2003 International Conference*, vol. 1, p. 545, July 2003.
- [85] W. Lin and C. C. Jay Kuo, "Perceptual visual quality metrics: A survey," *Journal of Visual Communication and Image Representation*, vol. 22, no. 4, pp. 297–312, 2011.

- [86] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *Image Processing, IEEE Transactions on*, vol. 13, pp. 600–612, april 2004.
- [87] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *Image Processing, IEEE Transactions on*, vol. 15, pp. 430–444, feb. 2006.
- [88] D. M. Chandler and S. S. Hemami, "Vsnr: A wavelet-based visual signal-to-noise ratio for natural images," *Image Processing, IEEE Transactions on*, vol. 16, pp. 2284–2298, sept. 2007.
- [89] A. Shnayderman, A. Gusev, and A. M. Eskicioglu, "An svd-based grayscale image quality measure for local and global assessment," *Image Processing, IEEE Transactions on*, vol. 15, pp. 422–429, feb. 2006.
- [90] E. Dunic, S. Grgic, and M. Grgic, "New image-quality measure based on wavelets," *Journal of Electronic Imaging*, vol. 19, no. 1, p. 011018, 2010.
- [91] R. Ferzli and L. J. Karam, "No-reference objective wavelet based noise immune image sharpness metric," in *Image Processing, 2005. ICIP 2005. IEEE International Conference*, vol. 1, sept. 2005.
- [92] D. Shaked and I. Tastl, "Sharpness measure: towards automatic image enhancement," in *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, vol. 1, p. 937, sept. 2005.
- [93] N. B. Nill and B. H. Bouzas, "Objective image quality measure derived from digital image power spectra," *Optical Engineering*, vol. 31, pp. 813–825, apr. 1992.
- [94] R. Ferzli and L. J. Karam, "A no-reference objective image sharpness metric based on the notion of just noticeable blur (jnb)," *Image Processing, IEEE Transactions on*, vol. 18, pp. 717–728, april 2009.
- [95] S. Wu, W. Lin, S. Xie, Z. Lu, E. P. Ong, and S. Yao, "Blind blur assessment for vision-based applications," *Journal of Visual Communication and Image Representation*, vol. 20, no. 4, pp. 231–241, 2009.
- [96] J. Dijk, M. van Ginkel, R. van Asselt, L. van Vliet, and P. Verbeek, "A new sharpness measure based on gaussian lines and edges," in *Computer Analysis of Images and Patterns* (N. Petkov and M. Westenberg, eds.), vol. 2756 of *Lecture Notes in Computer Science*, pp. 149–156, Springer Berlin / Heidelberg, 2003.
- [97] L. N. Thibos, X. Hong, A. Bradley, and R. A. Applegate, "Accuracy and precision of objective refraction from wavefront aberrations," *Journal of Vision*, vol. 4, no. 4, 2004.
- [98] X. Cheng, A. Bradley, and L. N. Thibos, "Predicting subjective judgment of best focus with objective image quality metrics," *Journal of Vision*, vol. 4, no. 4, 2004.
- [99] S. F. Ray, *Applied photographic optics: Lenses and optical systems for photography, film, video, electronic and digital imaging*. Focal Press, 2002.
- [100] J. Buzzi and F. Guichard, "Uniqueness of blur measure," in *Image Processing, 2004. ICIP '04. 2004 International Conference*, vol. 5, pp. 2985–2988, oct. 2004.
- [101] B. Brown, "A performance measure of stills optical image stabilization," 2011. Cambridge Mechatronics Ltd.

- [102] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *Journal of the Optical Society of America. A, Optics and image science*, vol. 4, no. 12, pp. 2379–2394, 1987.
- [103] H. Knutsson, C. F. Westin, and G. Granlund, "Local multiscale frequency and bandwidth estimation," in *Image Processing, 1994. Proceedings. ICIP-94., IEEE International Conference*, vol. 1, pp. 36–40, nov. 1994.
- [104] C. Massot and J. Héroult, "Model of frequency analysis in the visual cortex and the shape from texture problem," *International Journal of Computer Vision*, vol. 76, no. 2, pp. 165–182, 2008.
- [105] F. Gavant, L. Alacoque, A. Dupret, , and D. David, "A physiological camera shake model for image stabilization systems," in *Sensors, 2011 IEEE*, pp. 1461–1464, oct. 2011.
- [106] E. W. Jin and B. W. Keelan, "Slider-adjusted softcopy ruler for calibrated image quality assessment," *Journal of Electronic Imaging*, vol. 19, no. 1, p. 12, 2010.
- [107] R. Hoffner, "Hdtv and the resolving power of the eye," 2005.
- [108] D. Roetenberg, *Inertial and magnetic sensing of human motion*. PhD thesis, Enschede, May 2006.
- [109] J. Bernstein, "An overview of mems inertial sensing technology," vol. 20, no. 2, pp. 14–21, 2003.
- [110] STMicroelectronics, "L3gd20h : Memes motion sensor : three-axis digital output gyroscope," tech. rep., 2012.
- [111] InvenSense, "Idg-2020 & ixz-2020 product specification," tech. rep., 2011.
- [112] Panasonic, "Angular rate sensors : Ewts9p," tech. rep., 2011.
- [113] M. G. Song, Y. J. Hur, N. C. Park, Y. P. Park, K. S. Park, S. C. Lim, and J. H. Park, "Development of small sized actuator for optical image stabilization," in *Optomechatronic Technologies, 2009. ISOT 2009. International Symposium*, pp. 152–157, sept. 2009.
- [114] E. Simon, B. Berge, F. Fillit, H. Gatou, M. Guillet, O. Jacques-Sermet, F. Laune, J. Legrand, M. Maillard, and N. Tallaron, "Optical design rules of a camera module with a liquid lens and principle of command for af and ois functions," 2010.
- [115] C. Y. Lin and J. C. Chiou, "Design and fabrication of mems-based thermally-actuated image stabilizer for cell phone camera," *Solid-State Electronics*, vol. 77, pp. 64–71, 2012.
- [116] "Why IS and VR Matter", <http://www.kenrockwell.com/tech/image-stabilization.htm>.
- [117] Sony, "Sony a57 dslr camera," tech. rep., 2012.
- [118] Nokia, "Pureview image technology white paper 2 - the next breakthrough in purity of imaging," tech. rep., 2012.
- [119] K. Nishi and T. Onda, "Evaluation system for camera shake and image stabilizers," in *Multimedia and Expo (ICME), 2010 IEEE International Conference*, pp. 926–931, july 2010.

- [120] T. Hamamoto and K. Aizawa, "A computational image sensor with adaptive pixel-based integration time," *Solid-State Circuits, IEEE Journal*, vol. 36, pp. 580–585, apr. 2001.
- [121] G. Petschnigg, R. Szeliski, M. Agrawala, M. Cohen, H. Hoppe, and K. Toyama, "Digital photography with flash and no-flash image pairs," *ACM Trans. Graph.*, vol. 23, pp. 664–672, Aug. 2004.
- [122] A. Foi, S. Alenius, M. Trimeche, V. Katkownik, and K. Egiazarian, "A spatially adaptive poissonian image deblurring," vol. 1, pp. 925–928, 2005.
- [123] R. Fergus, B. Singh, A. Hertzmann, S. Roweis, and W. T. Freeman, "Removing camera shake from a single photograph," pp. 787–794, 2006.
- [124] A. Levin, P. Sand, T. Sang, C. F. Durand, and W. T. Freeman, "Motion-invariant photography," 2008.
- [125] P. K. Sa, R. Dash, and B. Majhi, "Restoration of motion blurred images," *International Journal of Signal and Imaging Systems Engineering*, vol. 2, no. 1-2, pp. 51–56, 2009.
- [126] R. Dash, P. K. Sa, and B. Majhi, "Rbfn based motion blur parameter estimation," in *Advanced Computer Control, 2009. ICACC '09. International Conference*, pp. 327–331, jan. 2009.
- [127] J. Xie, W. Lin, H. Li, K. Guo, B. Jin, Y. Zhang, and D. Liu, "A new algorithm for improving deblurring effects and addressing spatially-variant blur problems for image motion deblurring," vol. 2, pp. 651–655, 2011.
- [128] S. Fiori, "Blind deconvolution by a newton method on the non-unitary hypersphere," *International Journal of Adaptive Control and Signal Processing*, 2012.
- [129] S. Schuon and K. Diepold, "Comparison of motion de-blur algorithms and real world deployment," *Acta Astronautica*, vol. 64, pp. 1050–1065, 2009.
- [130] M. Ben-Ezra and S. K. Nayar, "Motion deblurring using hybrid imaging," in *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference*, vol. 1, pp. 657–664, june 2003.
- [131] S. K. Nayar and M. Ben Ezra, "Motion-based motion deblurring," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, pp. 689–698, june 2004.
- [132] S. k. Xu, S. k. Liang, D. Tu, and G. h. Li, "A deblurring technique for large scale motion blur images using a hybrid camera," in *Image and Signal Processing (CISP), 2010 3rd International Congress*, vol. 2, pp. 806–810, oct. 2010.
- [133] A. Rav Acha and S. Peleg, "Two motion-blurred images are better than one," *Pattern Recognition Letters*, vol. 26, no. 3, pp. 311–317, 2005.
- [134] L. Yuan, J. Sun, L. Quan, and H.-Y. Shum, "Image deblurring with blurred/noisy image pairs," *ACM Trans. Graph.*, vol. 26, jul. 2007.
- [135] H. Kano, H. Hatanaka, S. Fukumoto, and H. Murata, "Motion blur estimation of handheld camera using regular and short exposure image pair," in *Image Processing (ICIP), 2009 16th IEEE International Conference*, pp. 1317–1320, nov. 2009.
- [136] H. Hatanaka, S. Fukumoto, H. Murata, H. Kano, and K. Chihara, "An image stabilization technology for digital still camera based on blind deconvolution," *IEICE Transactions on Information and Systems*, vol. E94-D, no. 5, pp. 1082–1089, 2011.

- [137] X. Liu and A. El Gamal, "Simultaneous image formation and motion blur restoration via multiple capture," in *Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP '01). 2001 IEEE International Conference*, vol. 3, pp. 1841–1844, 2001.
- [138] M. Tico, "Multi-frame image denoising and stabilization," in *EUSIPCO*, 2008.
- [139] Canon PowerShot S100, <http://www.canon.com/camera-museum/tech/report/2011/12/>.
- [140] Digital Compact Camera Nikon COOLPIX S6400/S800c/S01, http://www.nikon.com/news/2012/0822_cpx-s_01.htm.
- [141] D. Jewitt, "Project pan-starrs and the outer solar system," *Earth, Moon, and Planets*, vol. 92, pp. 465–476, 2003.
- [142] A. Gavriely, "Handsets: Mpixels, small pixel and image quality challenge," Advasense, 2008.
- [143] G. E. Smith, "The invention and early history of the ccd," *Journal of Applied Physics*, vol. 109, no. 10, 2011.
- [144] A. Krymski, D. Van Blerkom, A. Andersson, N. Bock, B. Mansoorian, and E. R. Fossum, "High speed, 500 frames/s, 1024*1024 cmos active pixel sensor," pp. 137–138, 1999.
- [145] S. Kleinfelder, S. Lim, X. Liu, and A. El Gamal, "A 10000 frames/s cmos digital pixel sensor," *IEEE Journal of Solid-State Circuits*, vol. 36, no. 12, pp. 2049–2059, 2001.
- [146] M. El Desouki, M. Jamal Deen, Q. Fang, L. Liu, F. Tse, and D. Armstrong, "Cmos image sensors for high speed applications," *Sensors*, vol. 9, no. 1, pp. 430–444, 2009.
- [147] W. Uhring and M. Zlatanski, "Ultrafast imaging in standard (bi)cmos technology," 2012.
- [148] Cypress, "Vita 25k 53fps," tech. rep., 2010.
- [149] V. Research, "Phantom v1610," tech. rep., 2011.
- [150] Photron, "Fastcam sa2 high-definition high-speed video system," tech. rep., 2009.
- [151] IDT, "Model: Y7-s3," tech. rep., 2012.
- [152] T. Toyama, K. Mishina, H. Tsuchiya, T. Ichikawa, H. Iwaki, Y. Gendai, H. Murakami, K. Takamiya, H. Shiroshita, Y. Muramatsu, and T. Furusawa, "A 17.7mpixel 120fps cmos image sensor with 34.8gb/s readout," in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2011 IEEE International*, pp. 420–422, feb. 2011.
- [153] T. Watabe, K. Kitamura, T. Sawamoto, T. Kosugi, T. Akahori, T. Iida, K. Isobe, T. Watanabe, H. Shimamoto, H. Ohtake, S. Aoyama, S. Kawahito, and N. Egami, "A 33mpixel 120fps cmos image sensor using 12b column-parallel pipelined cyclic adcs," in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2012 IEEE International*, pp. 388–390, feb. 2012.
- [154] B. S. Reddy and B. N. Chatterji, "An fft-based technique for translation, rotation, and scale-invariant image registration," *Image Processing, IEEE Transactions on*, vol. 5, pp. 1266–1271, aug. 1996.
- [155] E. Kumar, M. Biswas, and T. Q. Nguyen, "Global motion estimation in frequency and spatial domain," in *Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference*, vol. 3, may 2004.

- [156] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, pp. 185–203, 1981.
- [157] S. Negahdaripour, "Revised definition of optical flow: integration of radiometric and geometric cues for dynamic scene analysis," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20, pp. 961–979, sep 1998.
- [158] C. Fermüller, D. Shulman, and Y. Aloimonos, "The statistics of optical flow," *Computer Vision and Image Understanding*, vol. 82, no. 1, pp. 1–32, 2001.
- [159] J. L. Barron, D. J. Fleet, and S. S. Beauchemin, "Performance of optical flow techniques," *International Journal of Computer Vision*, vol. 12, pp. 43–77, 1994.
- [160] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," 1981.
- [161] J. Y. Bouguet, "Pyramidal implementation of the affine lucas kanade feature tracker description of the algorithm," 2001.
- [162] P. Anandan, "A computational framework and an algorithm for the measurement of visual motion," *International Journal of Computer Vision*, vol. 2, pp. 283–310, 1989.
- [163] S. Zhu and K. K. Ma, "A new diamond search algorithm for fast block-matching motion estimation," *Image Processing, IEEE Transactions on*, vol. 9, pp. 287–290, feb. 2000.
- [164] Y. Nie and K. K. Ma, "Adaptive rood pattern search for fast block-matching motion estimation," *Image Processing, IEEE Transactions on*, vol. 11, pp. 1442–1449, dec. 2002.
- [165] Y. W. Huang, C. Y. Chen, C. H. Tsai, C. F. Shen, and L. G. Chen, "Survey on block matching motion estimation algorithms and architectures with new results," *The Journal of VLSI Signal Processing*, vol. 42, pp. 297–320, 2006.
- [166] C. Schmid, Appariement d'images par invariants locaux de niveaux de gris. Application à l'indexation d'une base d'objets. PhD thesis, INPG, 1996.
- [167] F. Comby, Estimation du mouvement apparent majoritaire dans une séquence d'images vidéo par accumulation de votes bimodaux sur un histogramme approché. PhD thesis, Université Montpellier II, 2001.
- [168] J. Martin and J. L. Crowley, "Experimental comparison of correlation techniques," in *IAS-4, International Conference on Intelligent Autonomous Systems*, 1995.
- [169] S. Chambon and A. Crouzil, "Mesures de corrélation robustes aux occultations," in *ORASIS 2003 - Journées jeunes chercheurs en vision par ordinateur*, (Gérardmer, France), May 2003.
- [170] S. C. Cheng and H. M. Hang, "A comparison of block-matching algorithms mapped to systolic-array implementation," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 7, pp. 741–757, oct 1997.
- [171] Kodak Lossless True Color Image Suite, <http://r0k.us/graphics/kodak/>.
- [172] G. Agranov, R. Mauritzson, S. Barna, J. Jiang, A. Dokoutchaev, X. Fan, and X. Li, "Super small, sub 2um pixels for novel cmos image sensors," 2005.
- [173] Learn the difference between a DSLR camera and compact And DSLR camera - you substitute ?, <http://camerausa.blogspot.fr/2012/03/learn-difference-between-dslr-camera.html>.

- [174] “Sony ericsson cyber-shot s006 camera module,” tech. rep., System Plus Consulting, 2011.
- [175] Le Stabilisateur, <http://www.commentcamarche.net/contents/l-objectif/le-stabilisateur>.
- [176] SteadyShot INSIDE - Body-integrated Image Stabilization System, http://www.sony.net/SonyInfo/technology/technology/theme/alpha_01.html.

Annexe 1 Instructions d'expérience de différence juste visible

Ce protocole d'instruction est celui utilisé pour faire passer l'expérience de différence juste visible aux différents observateurs. Cette expérience permet de déterminer l'évolution de la perception visuelle en fonction d'une dégradation. Ces différentes instructions permettent à l'observateur d'être totalement autonome pendant le test. Préalablement, un test d'acuité visuelle a été réalisé pour s'assurer que la personne ait une vision d'au moins 10/10 avec les deux yeux.

- Bonjour, tout d'abord, merci d'avoir accepté de prendre du temps pour passer ce test de comparaison de qualité d'image. Merci de vous mettre à l'aise, est de vous assoir sur cette chaise adossé au dossier afin de ne pas bouger pendant tout le long de l'expérience. Dans cette expérience, vous allez évaluer la qualité globale de la scène par comparaison d'image. Ces images ont subits différentes dégradations au niveau d'un flou de bougé linéaire. Ceci va impacter la qualité globale des images. Deux images vont être affichées sur l'écran avec deux flous différents. Le but sera de voter pour l'image ayant la meilleure qualité. En d'autre terme, vous voterez pour l'image que vous voudriez garder si vous devez en choisir une seule des deux. A chaque fois les deux images ont des flous différents mais peuvent être très proche en qualité. Même si il n'est pas évident de choisir, il faut tout de même se décider pour choisir l'image de meilleure qualité. N'oubliez pas qu'il n'y a pas de bonne ou de mauvaise réponse car c'est votre perception visuelle qui va déterminer qu'elle est l'image de meilleure qualité. Chaque image est affichée à gauche ou à droite de façon aléatoire de même pour les différents flous, ils changent de manière aléatoire. A la fin de l'expérience, le programme se fermera tout seul et enregistrera automatiquement les résultats.

- Regardons maintenant un exemple : Couple d'image #1. (Deux images ayant des flous différents facilement perceptibles) Par exemple, il apparait d'une façon claire que cette image est de meilleure qualité. Donc votons pour celle-ci justement. (Vote) Regardons encore un autre exemple : Couple d'image #2. (Deux images ayant des flous différents mais peu perceptible) Ici par exemple, il est difficile de choisir qu'elle image choisir, alors il faut voter pour celle qui vous semble de meilleur qualité. (Vote) Est-ce que vous avez des questions sur comment choisir l'image de meilleur qualité ?

Puis l'observateur test lui-même les 10 couples d'images suivant en guise d'entraînement. (Vote)

Avez-vous d'autres questions avant de commencer le test ? Réponse aux questions, puis démarrage du test.

Annexe 2 Instructions pour l'expérience de règle de qualité d'image ajustable

Ce protocole d'instruction est celui utilisé pour faire passer l'expérience de règle de qualité d'image ajustable aux différents observateurs. Cette expérience permet de quantifier la qualité d'une image grâce à une règle de mesure analogue à la règle classique. Ces instructions permettent à l'observateur d'être totalement autonome pendant le test. Préalablement, un test d'acuité visuelle a été réalisé pour s'assurer que la personne ait une vision d'au moins 10/10 avec les deux yeux.

- Bonjour, tout d'abord, merci d'avoir accepté de prendre du temps pour passer ce test de mise en correspondance de qualité d'image. Merci de vous mettre à l'aise, est de vous assoir sur cette chaise adossé au dossier afin de ne pas bouger pendant tout le long de l'expérience. Dans cette expérience, vous allez ajuster la qualité d'une image par rapport à une image à quantifier en termes de qualité. Les images à quantifier qui vont vous être présentés ont subits différentes dégradations au niveau d'un quelconque flou de bougé. Ceci va impacter la qualité globale de l'image. Deux images vont être affichées sur l'écran, l'une avec un flou fixe, l'image à quantifier et l'autre avec un flou que vous allez pouvoir faire varier. Le but sera d'ajuster l'image de flou variable de la même façon et de la meilleure manière que l'image avec un flou fixe. Ainsi par votre ajustement, vous jugerez que les deux images ont exactement la même qualité. En d'autre terme après ajustement, il vous sera impossible de dire laquelle est de meilleures qualités. Puis vous passerez à l'image suivante. A chaque fois l'image à quantifier sera différentes, alors il vous faudra réajuster l'image de qualité variable. A la fin de l'expérience, le programme se fermera tout seul et enregistrera automatiquement les résultats.

- Regardons maintenant un exemple : Couple d'image #1, avec une image avec un flou fixe, l'autre avec le flou variable. Grâce à l'ascenseur, il est simple de faire varier la qualité de l'image avec le flou variable. Une des techniques est de faire varier l'image de façon importante dans les deux sens pour borner (en termes de qualité) l'image à quantifier puis d'ajuster de manière plus fine jusqu'à avoir les deux images de qualité similaire.

- Puis l'observateur ajuste lui-même 3 autres couples d'images suivant en guise d'entraînement.

- Avez-vous d'autres questions avant de commencer le test ? Réponse aux questions.
- Faire passer le test de calibrage.
- Puis sans donner les résultats du test de calibrage, choisir le bon calibre puis faire passer le test final.

Liste des Publications

Conférence Internationale

Fabien Gavant, Laurent Alacoque, Antoine Dupret, and Dominique David, “A physiological camera shake model for image stabilization systems,” in *Sensors, 2011 IEEE*, pp. 1461–1464, oct. 2011.

Fabien Gavant, Laurent Alacoque, Antoine Dupret, Tien Ho Phuoc, and Dominique David, “Perceptual image quality assessment metric that handles arbitrary motion blur,” in *Proc. SPIE 8293, Image Quality and System Performance IX*, 829314, jan. 2012.

Brevet

En cours de dépôt : « Technique de stabilisation d’images garantissant leur netteté ».

TITRE

Contribution à des architectures de stabilisation d'images basées sur la perception visuelle et la physiologie du tremblement humain

RESUME

Avec l'intégration des appareils photos dans les appareils mobiles, leur démocratisation et la réduction de la taille de l'imageur, de l'optique et de la taille pixels, les photos sont de plus en plus sujettes au flou de bougé dû aux tremblements de la main. À cette tendance s'ajoute un accroissement constaté dans l'exigence de qualité d'image de la part des utilisateurs. Pour réduire ce flou, des systèmes de stabilisation d'image ont été développés. Néanmoins ceux-ci ne permettent pas de garantir la qualité de netteté des images et souffrent parfois d'une intégration limitée.

En réponse à ces limitations, ces travaux de recherche proposent, d'une part, un modèle de tremblement physiologique permettant de simuler de manière fidèle les flous de bougé et, d'autre part, une étude sur la perception visuelle du flou permettant le développement d'une métrique de qualité. Enfin des architectures de stabilisations, exploitant ces nouveaux outils, sont proposées. Ces nouvelles architectures permettent de réduire le nombre de composants externes ainsi que de garantir la netteté des images stabilisées.

Mots clés : *architectures de stabilisation d'image, flou de bougé, modèle de perception humaine du flou, imagerie numérique*

TITLE

Architectures for image sensors stabilization based on visual perception and on the physiology of hand tremor; a contribution

ABSTRACT

With the integration of cameras in mobile devices, their democratization and the reduction of the imager's size, the optical system dimensions and the pixels miniaturization, the photos become more and more subject to motion blur due to the hand tremor. In addition, the requirements in terms of image quality become higher and higher. Hence, in order to reduce this blur, several image stabilization systems have been developed. Nevertheless, they cannot guarantee the sharpness quality of resulting images and in some cases, they show integration difficulties.

In order to overcome these limitations, the research work presented in this thesis proposes, first of all, a physiological tremor model that aims to simulate realistic camera shake and secondly, presents a study on visual perception of blur. This study enables the development of a quality metric. Finally, stabilization algorithms and architectures exploiting these new tools are presented. These new architectures reduce the number of external components and ensure sharp stabilized images.

Keywords: *image stabilization architectures, motion blur, human perception blur model, digital imaging*

Thèse préparée au CEA-Leti dans le laboratoire L3I (Laboratoire de conception circuit Intégré, Intelligent pour l'Image), MINATEC Campus, 17 rue des Martyrs, 38054 Grenoble Cedex 9, France.