



Détection de différence de mise au point lors des prises de vues stéréoscopiques

Sergi Pujades, Frédéric Devernay

► To cite this version:

Sergi Pujades, Frédéric Devernay. Détection de différence de mise au point lors des prises de vues stéréoscopiques. *Traitement du Signal*, 2015, 32 (1), pp.65-86. 10.3166/TS.32.65-86 . hal-01352806

HAL Id: hal-01352806

<https://inria.hal.science/hal-01352806>

Submitted on 7 Sep 2016

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.

Détection de différence de mise au point lors des prises de vues stéréoscopiques

Sergi Pujades, Frédéric Devernay

Inria - Equipe PRIMA,
Univ. Grenoble Alpes, LIG, F-38000 Grenoble, France
CNRS, LIG, F-38000 Grenoble, France
{sergi.pujades-rocamora, frederic.devernay}@inria.fr

RÉSUMÉ. La production d'images stéréoscopiques nécessite un rig stéréoscopique avec deux caméras parfaitement synchronisées et alignées. La plupart des imprécisions de ce montage peuvent être corrigées en direct ou en post-production. Par contre, une différence de distance de mise au point ou de profondeur de champ entre les caméras produira des dégradations irrécupérables dans les images. Dans cet article nous détectons des différences de mise au point entre les deux vues d'une paire stéréoscopique en quatre étapes. D'abord nous calculons une carte de disparité dense. Ensuite nous mesurons la netteté dans chaque image et nous comparons ces mesures. Puis, avec des méthodes statistiques robustes, nous identifions les zones de l'image qui présentent des différences. Finalement, nous proposons une méthode de visualisation sur les images originales pour informer l'opérateur des problèmes, et lui donner des indices pour les résoudre.

ABSTRACT. Live-action stereoscopic content production requires a stereo rig with two cameras precisely matched and aligned. While most deviations from this perfect setup can be corrected either live or in post-production, a difference in the focus distance or focus range between the two cameras will lead to unrecoverable degradations of the stereoscopic footage. In this paper we detect focus mismatch between views of a stereoscopic pair in four steps. First, we compute a dense disparity map. Then, we use a measure to compare focus in both images. After this, we use robust statistics to find which images' zones have a different focus. Finally, to give useful feedback, we show the results on the original images and give hints on how to solve the focus mismatch.

MOTS-CLÉS : cinéma stéréoscopique, flou, netteté, distance de mise au point, depth from focus, depth from defocus.

KEYWORDS: stereoscopic cinema, focus mismatch, depth from focus, depth from defocus.

DOI:10.3166/TS.32.65-86 © 2015 Lavoisier

Extended Abstract

Live-action stereoscopic content production requires a stereo rig with two cameras precisely matched, temporally synchronized and geometrically aligned. Most deviations from this perfect setup can be corrected either live or in post-production, for example with the rectification of both stereoscopic streams. However a difference in the focus distance or focus range between the two cameras will lead to unrecoverable degradation of the stereoscopic footage. It is fundamental to have precise and reliable methods to detect any focus difference and fix it on-set. In this paper we propose a method to detect focus mismatch between views of a stereoscopic pair. Once the differences are detected, we indicate the stereoscopic operator which settings in the optics should be considered in order to fix them.

In this paper we study the relationship between the optical blur and the image blur. The optical blur model has two parameters: the focal distance and the aperture. The image blur is measured from the intensity variations in the sensor. One of the contribution of this paper is to establish the relationship between those two measures. Moreover we study in detail the models of image and optic blur difference and their relationship. A key point of our method is the hypothesis that the difference of optical blur and the difference of image blur have the same sign. This hypothesis allows to detect optical blur differences by only observing the images.

The proposed approach has four steps. First, we compute a dense disparity map with a multi-scale method, which is able to handle blur differences between the input images. Then, we use a blur measure in each individual image, and we compare both measures using the disparity map between the images. After this, we use robust statistics to find which images' zones have a different focus. Finally, to give useful feedback, we propose a visualization method showing the results on the original images. Moreover, thanks to the relationship between the image blur and optical blur, we can give hints to the operator on how to solve the focus mismatch.

To validate our method we conducted experiments on two types of datasets: synthetic ray-traced images, and actual images. The results on the synthetic data allow to validate the hypothesis stating that the sign of the optical blur difference is the same as the sign of the image blur difference. Moreover, the visualization method and the hints given to the operator show a very good performance. The results obtained with the actual images show the limitations of the approach to give the correct hint to the operator. However, the visualization method still has a good performance and allows to give a reliable feedback to the operator.

1. Introduction

Dans cet article nous proposons une méthode pour détecter des différences de *mise au point* (MAP) dans des images stéréoscopiques. Ces différences apparaissent quand la *distance de mise au point* (DMP) ou la *profondeur de champ* (PC) ne sont pas identiques pour les deux caméras.

La production d'images stéréoscopiques nécessite un rig stéréoscopique avec deux caméras parfaitement synchronisées et alignées. Ces réglages se font sur le lieu du tournage et prennent du temps. La plupart des imprécisions de ce montage peuvent être corrigées en direct ou en post-production. Par contre, une différence de DMP ou de PC entre les caméras produira des dégradations irrécupérables dans les images. Il est donc important de détecter ces incohérences rapidement pour que l'opérateur puisse les corriger sur le tournage.

Le lien entre la différence de MAP entre les deux images et la fatigue cérébrale ne fait pas l'unanimité. Les études se sont concentrées sur la visualisation d'une image nette avec une image floue (Lambooy *et al.*, 2007 ; Kooi, Toet, 2004 ; Seuntings *et al.*, 2006 ; Stelmach *et al.*, 2000), mais les effets de deux images ayant la même PC et des DMP différentes n'ont pas été étudiés à notre connaissance. Dans la TV-3D et le cinéma 3D, la MAP joue aussi un rôle artistique, et la plupart des réalisateurs croient qu'une dé-synchronisation de MAP peut interférer dans l'effet relief (Mendiburu, 2012).

Nous voulons détecter cette différence de MAP avec l'information provenant exclusivement des images capturées. Nous considérons que nous n'avons aucune information provenant des optiques : ni leur focale, ni leur DMP, ni la taille d'ouverture de leur iris. Bien qu'il existe des protocoles bien établis pour l'obtention de ces informations, toutes les optiques ne sont pas compatibles avec ces standards. Dans la plupart des cas les informations obtenues sont imprécises, puisque la calibration est faite en usine alors que les valeurs réelles dépendent de la position de l'optique par rapport au capteur de la caméra.

Mesurer le *flou dû à l'optique* (FO) à partir du *flou dans l'image* (FI) est un problème mal posé. Si nous avons une zone de couleur homogène dans la scène, nous ne pouvons pas distinguer une image nette d'une image floue car les deux seront fortement similaires. En outre, l'image nette d'un poster montrant une image floue apparaîtra floue. Dans notre cas, nous ne voulons pas mesurer le FO : nous voulons détecter une *différence* entre les deux images, puisque c'est cette incohérence qui risque de perturber la perception stéréoscopique. Dans nos travaux, nous ne traitons pas non plus la problématique de la restauration d'une image floue pour la rendre nette. Notre but se limite à détecter s'il y a une différence entre la DMP et/ou la PC des deux caméras. Nous imposons une contrainte de rapidité à notre méthode, afin d'être utilisable pendant le tournage. Nous voulons aussi être capables de donner de l'information au chef opérateur ou au réalisateur sur la nature du problème. Est-ce que la PC est la même pour les deux caméras ? Les deux caméras ont-elles la même DMP ? Ce sont des questions auxquelles nous voulons répondre.

2. Modèle de flou, différence de flou et le flou dans les images

2.1. Modèle de flou dû à l'optique

Le flou dans une image est généralement modélisé comme une convolution d'une image parfaitement nette avec une fonction de dispersion (PSF). Dans notre cas, nous supposons que cette fonction est en relation avec le FO, et nous ignorons les effets qui pourraient y contribuer de façon mineure (aberrations optiques, diffraction, etc).

Plusieurs modèles de fonction de dispersion ont été proposés dans la littérature (FitzGerrell *et al.*, 1997 ; Subbarao, 1988), bien que le modèle le plus utilisé soit celui d'une gaussienne 2D (Mennucci, Soatto, 1999 ; Pentland, 1987 ; Subbarao, 1988 ; Lin, Gu, 2007) (la fonction de dispersion réelle dépend, entre autres, de la forme de l'iris).

La taille du FO, σ , dépend de la distance du point 3D imagé au plan du diaphragme. On appellera cette distance la profondeur. Il ne faut pas la confondre avec la distance entre le point 3D imagé et le centre optique (point au centre du diaphragme). Dans notre cas, les surfaces de points 3D qui se retrouvent à la même profondeur ne forment pas des hémisphères mais des plans. Si on a deux caméras en configuration stéréoscopique, les axes optiques sont presque parallèles et orthogonaux au segment reliant les deux centres optiques. Dans ce cas, la profondeur est la même pour les deux caméras (les deux plans des diaphragmes coïncident) et la disparité stéréoscopique dépend simplement de la profondeur.

Une propriété importante démontrée par (Rajagopalan *et al.*, 2004 ; Schechner, Kiryati, 2000) est que la taille de FO a une relation linéaire avec la disparité stéréoscopique, comme l'illustre la figure 1. L'intuition derrière cette relation est qu'une image prise avec une PC limitée, correspondant à un diaphragme donné et une DMP, peut être interprétée comme la moyenne d'une infinité d'images prises avec une PC infinie (un diaphragme infiniment petit réduit au centre optique), et toutes convergentes à la DMP. Cette infinité de centres optiques couvrent complètement la surface du diaphragme. La taille du FO, à une profondeur donnée, correspond à la disparité stéréoscopique entre les deux positions extrêmes de ces centres optiques sur l'axe horizontal. Le modèle de FO a seulement deux paramètres : la *distance de mise au point* (DMP) et la *profondeur de champ* (PC).

On peut considérer qu'un élément de la scène est *net* s'il est à une disparité qui a une taille de FO plus petite qu'un seuil fixe, établi selon le critère subjectif de l'opérateur. Les éléments à une disparité avec une taille de FO plus grande seront considérés comme *flous*.

2.2. Différence de deux modèles de flou dû à l'optique

Comme nous avons deux fonctions de FO, une par caméra, nous considérons leur différence. Nous proposons de ranger chaque paramètre (DMP et PC) du modèle en trois catégories : il peut être plus grand à gauche qu'à droite, égal, ou plus petit à

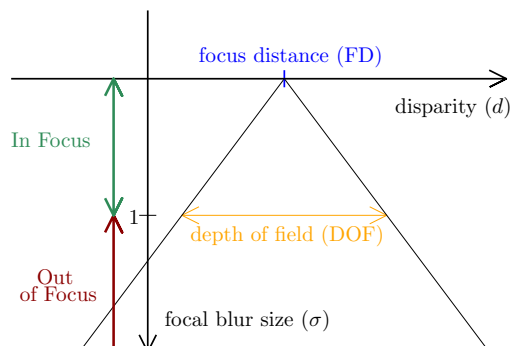


Figure 1. Le modèle de flou dû à l'optique : la taille du FO est une fonction de la disparité stéréoscopique linéaire par morceaux. Elle est nulle à la disparité correspondante à la distance de mise au point (DMP) et sa pente est inverse à la profondeur de champ (PC)

gauche qu'à droite. Une translation globale ou un facteur d'échelle appliqués aux paramètres ne changera pas la forme de la courbe. Sur la figure 2 nous représentons graphiquement les neuf cas, issus des combinaisons de trois possibilités pour chaque paramètre.

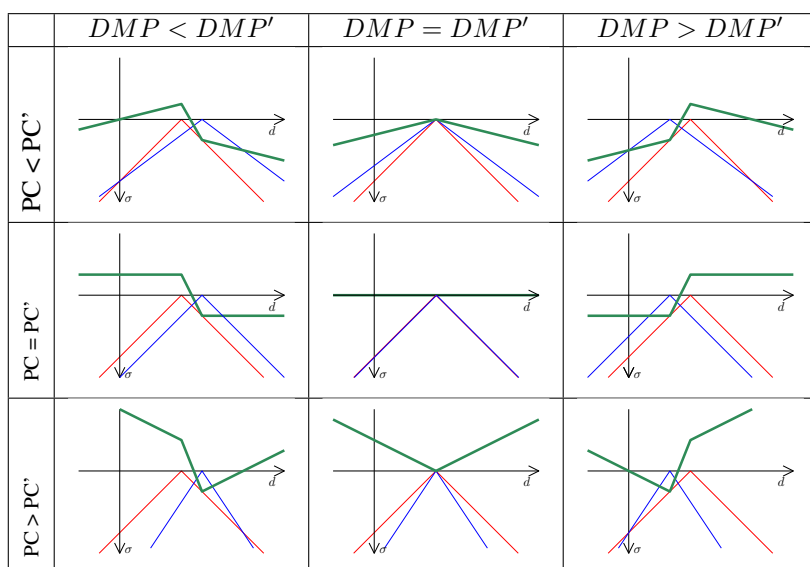


Figure 2. Graphiques de flou dû à l'optique et leur différence. Chaque case contient les graphiques des fonctions gauche en rouge, droite en bleu, puis leur différence en vert. Les paramètres de l'image gauche sont DMP et PC. DMP' et PC' sont les paramètres de l'image droite

2.3. Flou dans l'image

Nous venons d'introduire le modèle de FO. Mais dans notre méthode, nous avons seulement les images en entrée, et pas de connaissance des réglages des caméras. Ceci implique que nous ne pouvons pas mesurer le FO directement, mais seulement le *flou dans l'image* (FI). Une hypothèse clé dans notre méthode est de supposer que, pour chaque pixel, la différence de FI a le même signe que la différence de FO. Nous allons donc utiliser le signe de la différence de FI pour détecter une différence dans le FO. La figure 3 montre les graphiques correspondant au signe du FO en orange. Bien que le flou dans l'image dépende aussi de la texture de la scène, la courbe du signe de la différence de FI devrait être égal à la courbe du signe de la différence de FO. Cet ensemble de courbes constitue le *catalogue* de formes admissibles pour les graphiques de signe du FI.

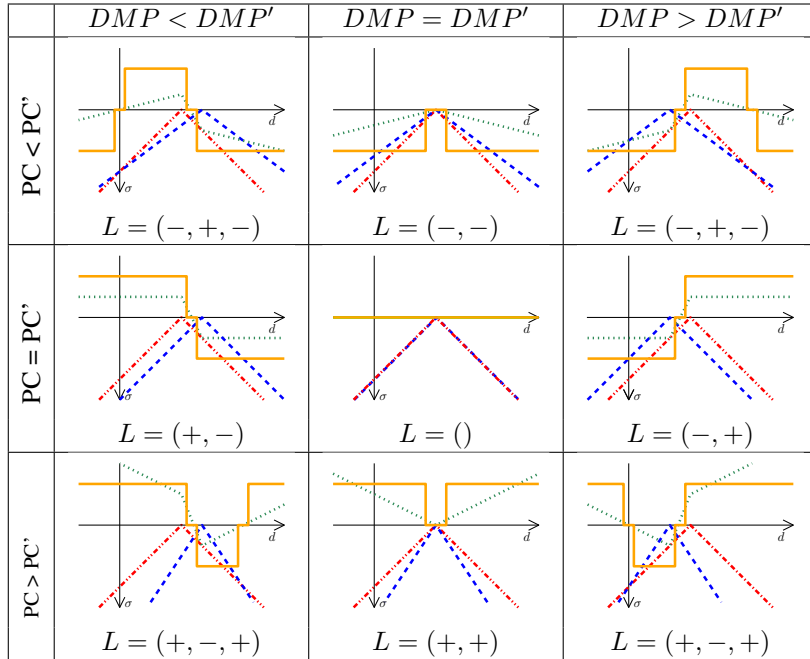


Figure 3. Catalogue des types de courbes possibles pour chaque configuration. En orange, le signe de la différence de FO. En pointillées les fonctions de FO gauche (rouge), FO droite (bleu) et leur différence (vert). Pour chaque courbe nous donnons aussi la liste L de signes que nous utiliserons dans la méthode de classification (section 3.5).

Notons que si nous considérons seulement la forme des courbes dans la figure 3, nous sommes incapables de différencier les cas ($PC < PC'$, $DMP < DMP'$) et ($PC < PC'$, $DMP > DMP'$). De façon analogue ($PC > PC'$, $DMP < DMP'$) a la même courbe

que ($PC > PC'$, $DMP > DMP'$). Dans la section 3.5, nous explicitons sous quelles conditions nous pouvons obtenir des informations utiles dans ces cas-là.

3. Mesure de différence de mise au point

Nous détaillons ci-après l'algorithme pour calculer la différence de MAP entre les images gauche et droite, exprimée en termes de sa DMP et de sa PC. Nous commençons par calculer une carte dense de disparité qui donne les correspondances entre les pixels des images, en utilisant une méthode multi-résolution peu sensible à la différence de flou (Sizintsev, Wildes, 2010). Ensuite, nous mesurons le FI et le signe de cette différence pour chaque paire de pixels mis en relation (par la carte de disparité). Nous estimons sur ces mesures bruitées un modèle de *différence de FO*. Finalement, nous essayons d'extraire de l'information utile à partir du modèle estimé et de donner un retour à l'opérateur. Les étapes de l'algorithme sont :

1. Calculer une carte de disparité dense
2. Mesurer la différence de flou dans l'image pour chaque pixel, et additionner ces valeurs par disparité
3. Estimer un modèle sur ces mesures
4. Classifier le modèle obtenu
5. Donner un retour à l'opérateur

3.1. Calcul des cartes de disparité

Nous devons d'abord calculer la disparité stéréoscopique qui permet de relier les images gauche et droite. Notons $d(x, y)$ pour la relation gauche-droite et $d'(x, y)$ pour la relation droite-gauche. Le calcul de la disparité stéréoscopique est en lui-même un sujet de recherche dans le domaine de la vision par ordinateur. Des avancées récentes montrent que, dans la plupart des cas, des disparités précises peuvent être calculées rapidement, voire à cadence vidéo. Par contre, dans des situations difficiles, comme une profondeur de champ réduite, des zones manquant de texture, des discontinuités de profondeur, des patrons répétitifs, des transparences ou des réflexions spéculaires, la disparité reste encore très difficile à estimer. Ces situations sont souvent la cause d'erreurs locales dans la plupart des méthodes.

Nous pouvons supposer que les images originales sont rectifiées. Elles n'ont pas de disparité verticale : un point (x, y) dans l'image gauche correspond à un point sur la même ordonnée y , ayant des coordonnées (x', y) dans l'image droite.

Pour notre problématique, nous avons besoin d'une méthode de calcul de disparité rapide qui se comporte bien dans le cas de PC réduite. Nous utilisons donc la méthode (Sizintsev, Wildes, 2010) capable de trouver des bonnes valeurs de disparité entre une image nette et une autre floue. Cette méthode n'est pas la seule de l'état de l'art qui remplisse nos contraintes et nous pourrions en utiliser d'autres.

Par ailleurs, nous voulons ignorer les éléments des images qui sont demi-occlus, puisqu'ils n'ont pas de correspondant sur l'autre image. Pour de détecter ces zones, nous considérons $d(x, y)$, la disparité gauche-droite du pixel $p = (x, y)$. Ce pixel correspond aussi au pixel $(x', y) = (x + d(x, y), y)$ sur l'image droite. Si les disparités ne sont pas proches ($|d(x, y) - d'(x', y)| < D_T$), un pixel en cache un autre et nous éliminerons du processus le pixel le plus lointain. Typiquement, la valeur de D_T est 1.

3.2. Mesure du flou dans l'image

Pour comparer le flou dans l'image entre des points correspondants sur les images, nous utilisons l'opérateur SML (*Sum of Modified Laplacian* (Huang, Jing, 2007 ; Nayar, Nakagawa, 1994)). Cet opérateur fut conçu pour des applications de *depth-from-focus* :

$$\text{SML}(p) = \sum_{i=x-N}^{x+N} \sum_{j=y-N}^{y+N} \nabla_{ML}^2 I(i, j), \text{ pour } \nabla_{ML}^2 I(i, j) \geq T, \text{ où}$$

$$\begin{aligned} \nabla_{ML}^2 I(x, y) = & |2I(x, y) - I(x - s, y) - I(x + s, y)| \\ & + |2I(x, y) - I(x, y - s) - I(x, y + s)| \end{aligned}$$

T est un seuil discriminant : des valeurs inférieures à T ne sont pas additionnées. N est la taille de la fenêtre utilisée pour le calcul du SML, et s est la taille du pas pour le calcul du laplacien. Dans nos expérimentations les paramètres étaient $T = 5$ (images 8-bits), $N = 1$ et $s = 1$. Pour un pixel $p = (x, y)$ dans l'image *gauche*, $\text{SML}(p)$ est sa réponse à l'opérateur SML. $\text{SML}'(p)$ est la réponse de l'opérateur sur son pixel correspondant sur l'image *droite*.

Nous écrirons $S(p)$ pour désigner le signe de la différence des SML entre deux pixels :

$$S(p) = \text{signe} (\text{SML}(p) - \text{SML}'(p)) \quad (1)$$

Elle est positive si l'image gauche est plus nette que la droite, et elle est négative si la droite est plus nette que la gauche. Elle est nulle si $\text{SML}(p) = \text{SML}'(p)$. Typiquement, des zones sans texture ont des valeurs SML nulles.

Notons qu'une différence importante de contraste entre les images peut fortement perturber ces mesures. Une zone floue et très contrastée risque d'être considérée comme plus nette qu'une zone nette mais peu contrastée. Néanmoins la différence de contraste entre les images d'une paire stéréoscopique peut être calculée facilement, puis corrigée par une transformation linéaire des valeurs de l'image moins contrastée. Nous considérons que les images en entrée ont été normalisées, mais que des différences locales de contraste peuvent subsister, introduisant un léger biais sur les mesures.

Pour avoir une indication sur la fiabilité de la différence de flou dans les images, nous définissons :

$$m(p) = \max (|\text{SML}(p)|, |\text{SML}'(p)|)$$

Cette valeur sera élevée si l'image est texturée et nette.

Pour chaque valeur de disparité dans la scène, $M(d)$ sera la moyenne des $S(p)$ pour les pixels qui sont à disparité d , pondérés par $m(p)$. $w(d)$ sera la somme de poids correspondante. Donc

$$M(d) = \frac{\sum_{p|d(p)=d} m(p)S(p)}{\sum_{p|d(p)=d} m(p)} \quad \text{et} \quad w(d) = \sum_{p|d(p)=d} m(p)$$

Ces éléments nous donnent une estimation de quelle image est plus nette pour les pixels de même disparité.

Évidemment cette estimation peut être bruitée. Les sources d'erreur les plus importantes que nous avons déjà évoquées sont :

- La valeur de disparité peut être peu précise, impliquant une erreur de mise en correspondance entre (x, y) et (x', y) . $S(p)$ risque d'être faux.
- Des différences locales de contraste peuvent biaiser $S(p)$, même si la disparité est bien estimée.
- Le FO dans des zones de l'avant-plan peut «*baver*» à travers les discontinuités de profondeur : des zones de l'arrière-plan peuvent être mesurées à tort comme floues.
- Le nombre de pixels disponibles dans chaque disparité dépend du contenu de la scène. L'information pour certaines valeurs de profondeur peut être inexistante.

Pour compenser ces erreurs nous utiliserons des méthodes robustes pour l'estimation du modèle.

3.3. Estimation du modèle de flou dû à l'optique à partir du signe de la différence de flou dans l'image

Bien que les mesures soient bruitées, le signe de la différence de FI, en fonction de la disparité, doit avoir une des formes présentées dans la figure 3. Dans cette étape nous estimons un modèle qui respecte ce *catalogue* (figure 3).

En entrée nous avons les mesures $M(d)$ et $w(d)$ et nous cherchons une fonction $C(d)$ qui minimise l'énergie :

$$E(C) = \sum_d w(d)E_{\text{données}}(d) + \lambda E_{\text{régul}}(d), \text{ avec } \begin{cases} E_{\text{données}}(d) = |M(d) - C(d)| \\ E_{\text{régul}}(d) = |C(d-1) - C(d)| \end{cases} \quad (2)$$

$E_{\text{données}}$ est le terme d'attache aux données. Comme le modèle que nous cherchons est continu par morceaux en d , nous introduisons le terme $E_{\text{régul}}$ comme régularisateur. Pour chaque disparité d , la fonction $C(d)$ peut prendre un des états suivants : “*Gauche plus nette que droite*”, “*Droite plus nette que gauche*” et “*Les deux images ont la*”

même netteté”. Nous représentons ces états en valeurs numériques $+1, 0, -1$. Nous voulons que notre implantation du calcul soit robuste au bruit, donc nous permettons à $C(d)$ de prendre 5 valeurs possibles :

- Droite plus nette (-0.7)
- Droite peut-être plus nette (-0.3)
- Même netteté (0)
- Gauche peut-être plus nette (0.3)
- Gauche plus nette (0.7)

Le fait que la courbe $C(d)$ doive prendre une des formes correspondantes à la figure 3 nous rajoute une dernière contrainte : le nombre de changements de signe est limité. Pour chaque passage d’une valeur positive ou négative vers zéro, nous ajoutons une transition. Par contre si on passe d’une valeur positive à une négative directement (ou l’inverse), nous comptons deux transitions. En observant les formes de la figure 3 on peut voir que le nombre maximal de transitions permis est de quatre.

Néanmoins, cette condition est nécessaire mais pas suffisante. Certaines courbes ont quatre transitions ou moins mais ne correspondent à aucune courbe de notre *catalogue*. Ces courbes ne correspondent pas non plus à une partie (un sous-ensemble) de nos modèles (plus d’information sur les modèles incomplets dans la section 3.5.2). La figure 4 montre des formes de courbes qui sont interdites. Ces formes interdites seront éliminées de nos candidates pendant l’optimisation.

Nous pouvons exprimer l’énergie de l’équation 2 par récurrence sur d :

$$\begin{aligned} E^{d+1}(C) &= \sum_{k=d_{\min}}^{d+1} w(k)E_{\text{données}}(k) + \lambda E_{\text{régul}}(k) = \\ &= E^d(C) + w(d+1) |M(d+1) - C(d+1)| + |C(d) - C(d+1)| \end{aligned}$$

Le minimum de l’énergie totale $E(C) = E^{d_{\max}}(C)$ peut donc s’obtenir en un temps linéaire (en $O(d_{\max} - d_{\min})$) par programmation dynamique. Le résultat est une courbe qui, pour chaque disparité, nous dit si la gauche est plus nette, si les deux ont la même netteté ou si la droite est plus nette.

3.4. Visualisation de la différence de MAP

A cette étape nous sommes déjà capables de visualiser le résultat brut de notre méthode. Nous proposons de dessiner des zébrures sur les zones de l’image qui sont moins nettes que les autres. Nous pourrions faire ce dessin par pixel, c’est-à-dire en utilisant l’information de $S(p)$ de l’équation 1, mais nous nous sommes aperçus que le résultat était trop bruité et difficile à interpréter (voir figure 9).

Nous pouvons obtenir une bien meilleure visualisation en utilisant la fonction $C(d)$. Nous avons la disparité d pour chaque pixel, et nous pouvons dessiner les zé-

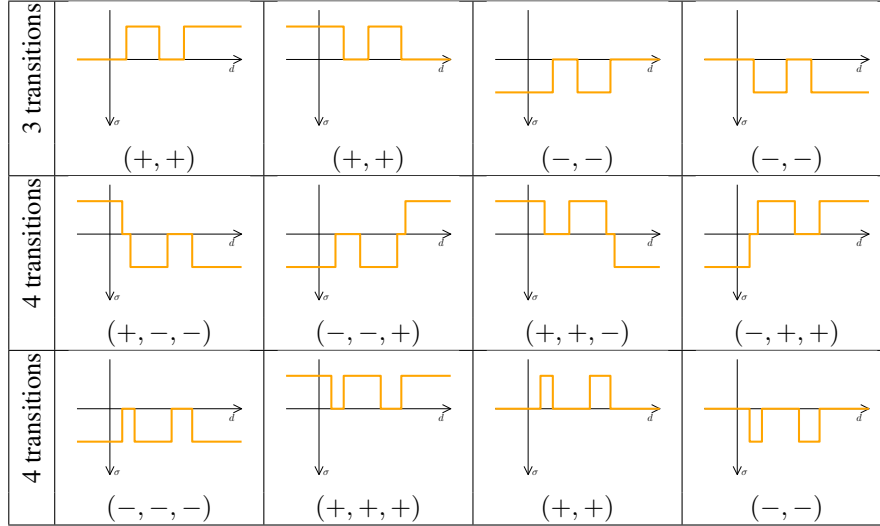


Figure 4. Courbes interdites : elles ne correspondent pas aux courbes du catalogue ni à des sous-ensembles. Pour chaque courbe nous donnons aussi la liste des signes utilisés dans l'étape de classification (voir section 3.5)

brures à partir du modèle $C(d)$ obtenu. Nous dessinons des zébrures dans un sens si une image est *peut être moins nette* (c.-à-d. $C(d) = -0.3$ pour l'image gauche). Et nous dessinons des zébrures dans les deux sens si nous sommes sûrs que l'image est *moins nette* (c.-à-d. $C(d) = -0.7$ toujours pour l'image gauche). De cette façon nous évitons de montrer le bruit présent dans la mesure $S(p)$, et nous pouvons interpréter facilement les images obtenues (figure 11).

Nous donnons donc une information visuelle indiquant quelle caméra devrait être réglée, et nous pouvons laisser la liberté à l'opérateur de décider quelle est la cause de cette différence. Dans la plupart des cas, spécialement s'il y a assez de mesures autour de la zone de DMP des deux caméras, nous sommes capables de classifier $C(d)$ et donner des indications plus précises sur la nature du problème.

3.5. Classification de la fonction du signe de la différence de flou des images

Nous voulons classer la fonction $C(d)$ dans une des cases de notre *catalogue* de la figure 3. Pour cela, nous construisons une liste L constituée des parties positives et négatives de la fonction, ordonnées par disparité croissante. Dans les figures 3 et 4 nous montrons, pour chaque forme, sa liste L . Une fonction $C(d)$ autorisée peut avoir un maximum de quatre transitions de signe, donc notre liste n'aura au maximum que trois éléments, et aucun au minimum.

3.5.1. Cas où des mesures sont disponibles pour toutes les disparités

Si nous avons des mesures pour *toutes les valeurs de disparité*, correspondante aux profondeurs de zéro à l'infini, nous pourrions classer facilement la fonction $C(d)$ en utilisant la liste L et notre *catalogue*. Comme nous avons rejeté les configurations interdites dans les étapes précédentes, les seules configurations possibles sont : $L \in \{(), (+), (-), (+, +), (-, -), (+, -), (-, +), (+, -, +), (-, +, -)\}$.

Chaque configuration, exceptées $L = (+)$ et $L = (-)$, peut être mise en relation avec un élément de la figure 3 et nous donner des indications sur la cause du problème. Néanmoins, la supposition que toutes les mesures sont disponibles pour toutes les disparités n'est pas très réaliste. Nous allons donc revoir cette supposition.

3.5.2. Un cas plus réaliste : des mesures clairsemées

Selon la géométrie de la scène, nous n'aurons de l'information que sur quelques valeurs de disparité. Ces mesures nous donneront une vision partielle du modèle complet de la différence de flou. Dans quelques cas, des visions partielles de modèles différents donneront lieu à la même mesure, ce qui rendra difficile le choix du modèle qui a généré les mesures. Nous présentons un exemple, montrant la difficulté d'estimer le bon modèle à partir de mesures partielles. Soit d_0 une disparité fixe quelconque.

$$\text{Soit } C(d) = \begin{cases} +1 & \text{si } d \leq d_0, \\ -1 & \text{si } d > d_0. \end{cases} \quad \text{et} \quad L = (+, -).$$

Les modèles candidats sont montrés dans la figure 5.

$DMP < DMP'$			$DMP > DMP'$	
$PC < PC'$	$PC = PC'$	$PC > PC'$	$PC < PC'$	$PC > PC'$

Figure 5. Candidats possibles pour l'exemple où $C(d)$ change de positive à négative dans la partie visible de la disparité (intervalle non-rayé) :
5 configurations possibles

$DMP < DMP'$		
$PC < PC'$	$PC = PC'$	$PC > PC'$

Figure 6. Candidats possibles si on suppose que les disparités correspondant aux DMP des deux caméras sont dans l'intervalle de disparité mesurable (intervalle non-rayé) : plus que 3 configurations possibles

Dans cet exemple, un grand ensemble de modèles candidats correspondent à notre fonction $C(d)$. Pour réduire leur nombre, une supposition raisonnable pourrait être de considérer que l’opérateur a fait le point d’une caméra sur un objet de la scène, et que l’autre DMP est aussi dans l’intervalle de disparité observable. Cette hypothèse réduit le nombre de candidats possibles pour notre modèle. Il ne nous reste que les candidats de la figure 6, qui partagent la relation $DMP < DMP'$.

Nous pouvons affirmer que si nous avons des mesures sur les disparités correspondantes aux DMP, nous pouvons *toujours* répondre à la question : *quelle distance de mise au point est la plus proche ?* Une fois que l’opérateur a corrigé les distances de mise au point ($DMP = DMP'$), nous pouvons *toujours* répondre à la question : *quelle profondeur de champ est la plus grande ?*

4. Résultats

Afin de valider notre méthode nous avons réalisé des tests avec des images générés par lancer de rayons et avec des images réelles issues de prises de vues. Cette section est composée de cinq parties. Dans les deux premières nous expliquons la génération d’images test avec la méthode de lancer de rayons et illustrons les étapes de la méthode avec un exemple. Ensuite nous analysons les résultats obtenus avec la méthode proposée et le premier jeu d’images générées. Finalement, dans les deux dernières parties nous présentons la méthode pour l’acquisition d’un jeu de données réelles et analysons les résultats obtenus par la méthode sur ce jeu de données.



Figure 7. Exemple d’images et cartes de profondeur des scènes “Patio” et “Office”

4.1. Images synthétiques générées par lancer de rayons

Pour valider notre méthode, nous avons généré deux jeux d’images de test. Pour avoir un contrôle total sur les paramètres de la caméra nous avons utilisé un moteur de rendu qui utilise le lancer de rayons (POV-Ray (Young, 1992)) et nous avons généré des images de la scène “Patio” et “Office” de Jaime Vives Piqueres (Piqueres, 2003) avec une caméra de modèle perspective (Devernay, 2014a ; 2014b). POV-Ray peut simuler le FO en calculant l’ensemble de rayons optiques voisins du pixel rendu en faisant sa moyenne. La taille de ce voisinage correspond à la taille du diaphragme. Ce modèle de génération de flou est donc similaire au nôtre (section 2.1).

Nous avons généré des images avec des paramètres différents de PC et de DMP. Pour chaque vue (gauche et droite) nous avons généré 5 images : *Tout*, *Près*, *Près2*,

Loin et *Loin2*. Les images *Tout* ont une PC infinie. Dans le jeu “Patio” les images *Loin** ont une DMP à $d_L = 14,2$ (point vert sur la figure 7) et les images *Près** ont une DMP à $d_P = 10,2$ (point bleu sur la figure 7). Dans le jeu “Office” les images *Loin** ont une DMP à $d_L = 0,85$ (point vert sur la figure 7) et les images *Près** ont une DMP à $d_P = -11,7$ (point bleu sur la figure 7).

Les images *Loin* et *Près* ont un diaphragme de valeur 5, qui génère une profondeur de champ finie. Les images *Loin2* et *Près2* ont un diaphragme de valeur 10 (double de *Loin* et *Près*), ayant donc une PC plus réduite.

4.2. Résultats intermédiaires

Nous allons dérouler l'exemple (*GLoin, DPrès*). Dans la figure 8 nous avons un détail de ces images : à gauche la fontaine est floue, tandis qu'à l'arrière, les fleurs sont nettes. À droite la fontaine est nette, tandis que les fleurs sont floues.



Figure 8. Détail de la paire “Patio” (*GLoin, DPrès*)

Dans la figure 9 nous observons $d(p)$: la carte de disparité calculée; $S(p)$: le signe de la différence de SML; et $m(p)$: le maximum de SML et SML'. L'image de $m(p)$ a été égalisée afin de montrer les détails dans le milieu de l'échelle de valeurs. Les mesures $S(p)$ sont, comme nous l'attendions, très bruitées. Mais une dominance du signe est visible pour les éléments à disparité d_P et d_L . Ces valeurs correspondent aux points bleus et verts sur la carte de disparité (figure 9a). L'image $m(p)$ montre quelle information est plus fiable : elle correspond aux zones texturées et nettes sur une image. Dans la figure 10 nous pouvons voir les données que l'étape d'estimation de modèle prend (section 3.3): $M(d)$, $w(d)$ et $C(d)$. Les valeurs minimale et maximale de $C(d)$ obtenues sont très proches des valeurs correctes des distances de mises au point. Ces valeurs indiquent un fort décalage entre les mises au point des caméras. Nous pouvons donc facilement répondre à la question la plus importante : *Y a-t-il une différence de MAP?* Avec le modèle estimé et le *catalogue* nous pouvons conclure que $DMP > DMP'$ et $PC = PC'$. Éventuellement, $PC = PC'$ pourrait ne pas être vrai si nous n'avions qu'une vision partielle du modèle (intervalle de disparité réduit). $PC > PC'$ ou $PC < PC'$ sont aussi possibles.

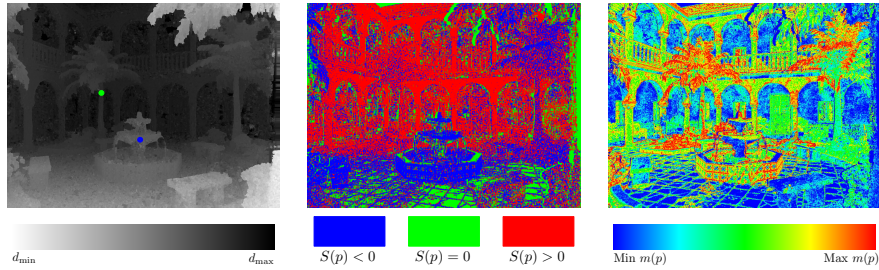


Figure 9. Résultats intermédiaires pour la paire (GLoin, DPrès). De gauche à droite : $d(p)$: Disparité calculée. $S(p)$: Signe de la différence de SML. $m(p)$: Maximum des SML. Les points bleu et vert dans la carte de disparité correspondent aux DMP Près et Loin

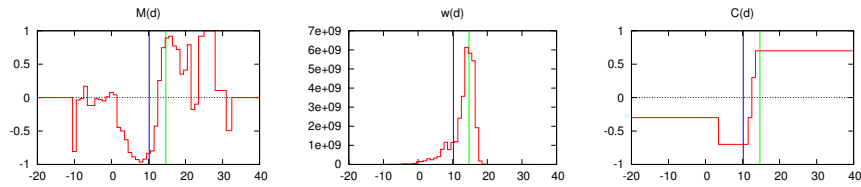


Figure 10. Résultats pour la paire (GLoin, DPrès). De gauche à droite : $M(d)$: moyenne pondérée du signe de la différence de SML. $w(d)$: somme du maximum des valeurs de SML. $C(d)$: Modèle estimé. Les marques bleue et verte correspondent respectivement aux disparités Près et Loin

Finalement, dans la figure 11 nous montrons les zébrures obtenues sur les images originales (GLoin, DPrès) (la gauche avec mise au point sur l'arbre à l'arrière, la droite avec la mise au point sur la fontaine). Notons que les bonnes zones de chaque image ont été correctement identifiées.



Figure 11. Images originales (GLoin, DPrès) avec les zébrures : la zone sur l'image gauche (c'est-à-dire la fontaine) est moins nette. La zone sur l'image droite (c'est-à-dire le fond) est moins nette.

4.3. Discussion des résultats sur images synthétiques

Afin de valider notre méthode nous l'avons testée avec des cartes de disparité vérité terrain et des cartes calculées. Nous avons réalisé les calculs entre chaque paire possible formée par nos 5 images originales. Nous avons donc 25 résultats que nous présentons dans les figures 12,13 et 14. Nous aimerions répondre aux questions suivantes, qui sont ordonnées par difficulté :

1. Les DMP et PC sont-elles identiques ?
2. Les DMP sont-elles identiques ? Laquelle est la plus grande ?
3. Les PC sont-elles identiques ? Laquelle est la plus grande ?

Dans tous les cas, nous pouvons répondre correctement à la première question : les éléments de la diagonale valent zéro, et les autres ont des valeurs non nulles.

Les quatre graphiques en bas à gauche et les autres quatre en haut à droite de la figure 12 montrent que nous pouvons répondre avec précision à la deuxième question quand les DMP sont différentes (c'est-à-dire *Près vs. Loin* ou *Loin vs. Près*). Pour les quatre premiers, nous obtenons la même forme que dans l'exemple de la section 4.2, et nous déduisons que $DMP > DMP'$. Pour le deuxième groupe, nous obtenons les courbes symétriques sur l'axe y, ce qui nous permet de conclure que $DMP < DMP'$. Par contre, dans ces huit cas, nous sommes incapables de répondre à la troisième question et d'affirmer s'il y a une différence de PC entre les caméras (voir section 4.2).

Les douze graphiques restants correspondent à une différence de PC entre les caméras. Nous pouvons observer que seulement deux d'entre eux ont la forme que notre modèle nous avait prédit. Pour ces cas, nous répondons correctement à toutes les questions : (*GLoin*, *DLoin2*) et (*GLoin2*, *DLoin*). Dans les dix restants, bien que nous ne puissions pas dire si la PC est différente à partir des courbes, nous pouvons toujours répondre à la question *Quelle image est la plus nette ?*

Lorsque nous utilisons les disparités calculées nous obtenons des résultats très similaires à ceux obtenus en utilisant la disparité vérité terrain. Notons qu'il n'y a pas de différence significative entre la figure 12 et les figures 13, 14.

4.4. Prises de vues d'images réelles

Pour valider notre méthode sur des prises de vues réelles nous avons acquis des images avec un appareil photo numérique (Nikon D80). Nous appellerons ce jeu de données "Desk". Pour la prise de vues stéréoscopique nous avons utilisé un rail de travelling que nous avons réglé au mieux pour avoir l'axe de déplacement perpendiculaire au capteur de l'appareil. Nous avons créé un jeu d'images avec deux DMP

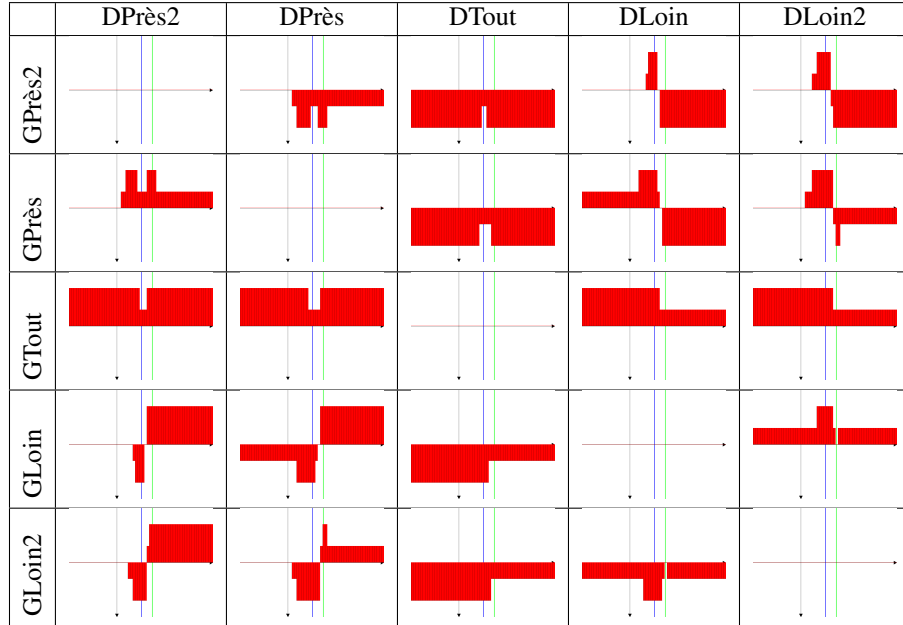


Figure 12. Modèles estimés avec la disparité vérité terrain pour la scène “Patio”

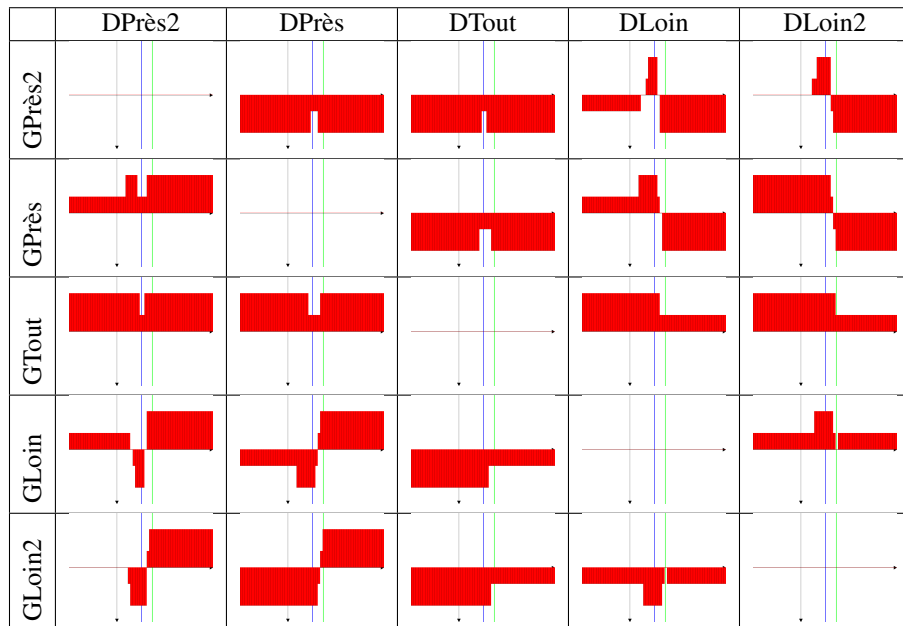


Figure 13. Modèles estimés avec la disparité calculée pour la scène “Patio”. Les marques bleue et verte correspondent respectivement aux disparités Près et Loin

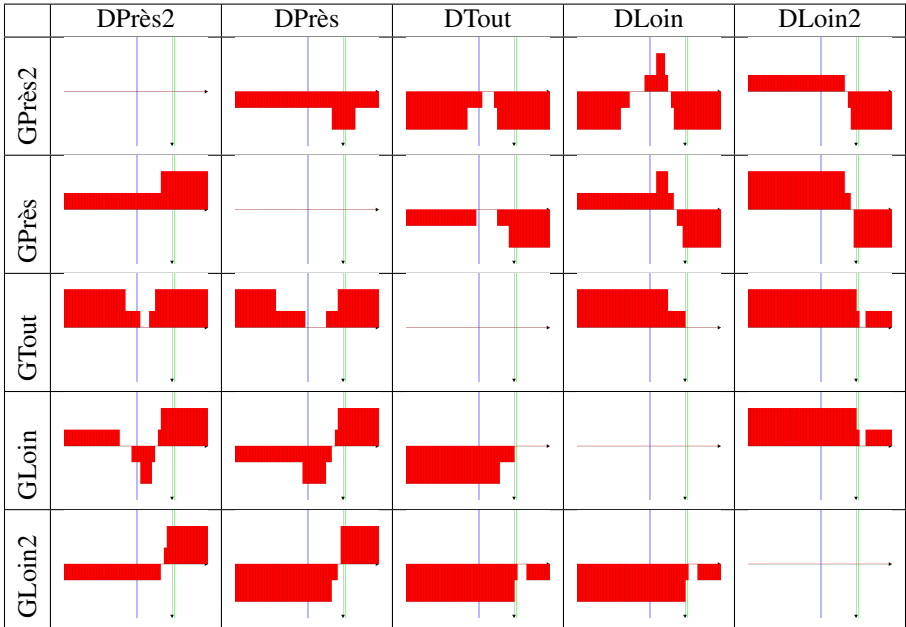


Figure 14. Modèles estimés avec la disparité calculée pour la scène “Office”. Les marques bleue et verte correspondent respectivement aux disparités Près et Loin



Figure 15. La paire (GPrès2, DTout) du jeu de données “Desk”

différentes (Près et Loin) et trois PC avec des ouvertures $f/5.6$ (Près2, Loin2), $f/11$ (Près, Loin) et $f/32$ (Tout). Dans la figure 15 nous montrons la paire (GPrès2, DTout).

Une difficulté de la génération d’un bon jeu de données est le fait que le changement de DMP pendant la prise de vues modifie légèrement la focale de l’appareil. On se retrouve donc avec des cadres légèrement différents entre les images. Ces différences font que lors du calcul de la carte de disparité les valeurs varieront selon la paire sélectionnée et les résultats seront difficilement comparables. Afin d’obtenir des résultats comparables entre toutes les combinaisons de paires possibles nous avons d’abord corrigé les distorsions dues à l’optique, puis stabilisé chaque série d’images



Figure 16. Détail de l'image GLoin2 du jeu "Desk" et sa carte de profondeur calculée correspondante. Le stylo rouge et toute sa zone de flou autour sont étiquetés à la profondeur "Près", alors que les mesures de flou correspondent à des valeurs de "Loin"

du même point de vue vers l'image la plus nette (*Tout*). Pour la correction des distorsions nous avons utilisé un logiciel commercial, et pour la rectification nous avons calculé des homographies.

Une fois les images stabilisées, nous avons calculé la rectification entre les images les plus nettes, et nous avons appliquée cette rectification à toutes les images. On se retrouve ainsi avec un jeu d'images qui aura des valeurs de disparité comparables entre n'importe quelle paire choisie.

4.5. Discussion des résultats sur images réelles

Les résultats obtenus avec le jeu de données "Desk" sont présentés dans la figure 17. Nous pouvons clairement observer les limitations de la méthode lorsque nous utilisons des images réelles. Bien qu'il y ait des courbes correctes (par ex. (GPrès, GPrès2) ou (GLoin, DPrès)), beaucoup d'entre elles n'ont pas la forme idéale attendue. Par exemple, aucun des courbes de la diagonale n'est strictement zéro.

Le facteur le plus important qui introduit des erreurs est l'estimation de la carte de correspondances. Dans le cas réel où le flou est plus important que dans le cas des images synthétiques générées, des pixels flous au bord des objets sont considérés comme étant "devant", alors qu'on voit à travers eux des pixels nets du fond. Dans la figure 16, nous observons un exemple de ce phénomène. Ces imprécisions ajoutent du bruit dans nos mesures, et empêchent la bonne estimation du modèle.

Malgré cette limitation, nous sommes toujours capables de donner une indication visuelle à l'opérateur sous la forme de zébrures. Bien que le modèle ne soit pas correctement estimé, les zébrures sur les images donnent une très bonne idée de la différence de flou, comme le montre la figure 18.

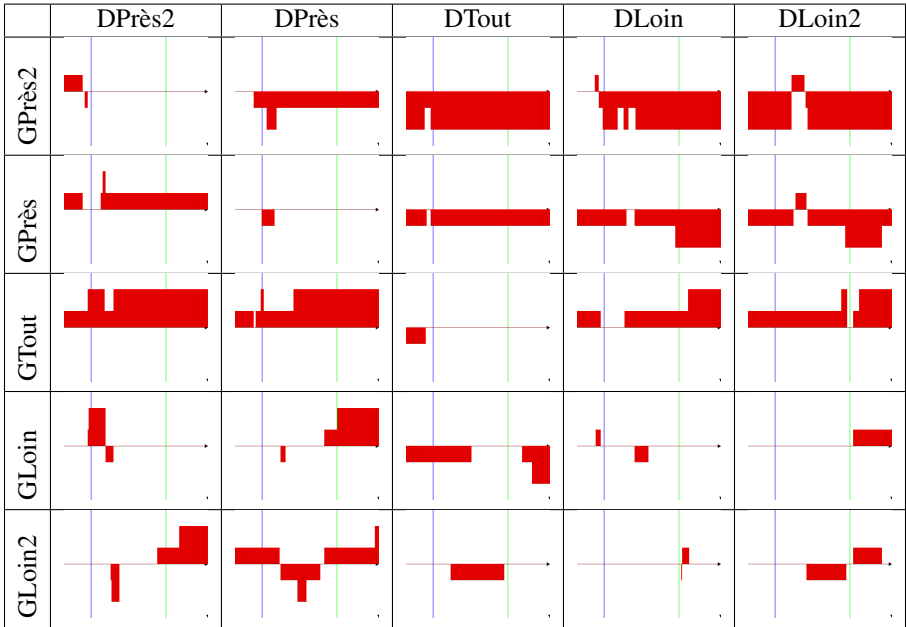


Figure 17. Modèles estimés avec la disparité calculée pour la scène “Desk”. Les marques bleue et verte correspondent respectivement aux disparités Près et Loin

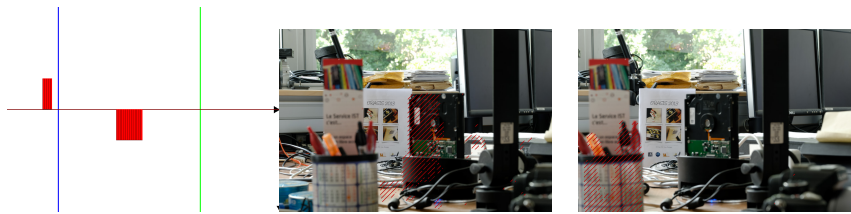


Figure 18. Modèle estimé et zébrures sur les images pour la paire (GLoin, DLoin). Malgré l’erreur dans le modèle final obtenu, l’opérateur peut être rassuré sur sa mise au point sur l’élément “Loin”. Les zébrures simples (pot de crayon et disque dur) indiquent un doute sur la correspondance de flou entre les zones “Près”

5. Conclusions et perspectives

Nous avons présenté une méthode complète pour détecter une différence de MAP (mise au point) entre deux images d’une paire stéréoscopique. Comme il est impossible de récupérer l’information perdue par le flou après le tournage, il est très important que les DMP (distances de mise au point) et les PC (profondeurs de champ) soient identiques.

Les résultats obtenus sur les images générées par lancer de rayons sont très bons. Ils permettent de valider notre hypothèse concernant l’égalité entre le signe de la diffé-

rence de FO (flou dû à l'optique) et FI (flou dans l'image) exposé dans la section 2.3. Tous les éléments de la diagonale des figures 13 et 14 valent zéro, ce qui veut dire que notre méthode est très précise pour le cas le plus important : les deux réglages sont identiques. Un autre point positif est que nous n'avons aucun *faux positif* : toutes les configurations présentant un problème sont détectées correctement. Nous pouvons affirmer que notre méthode a un taux de réussite de 100 % lorsque l'on veut répondre à la question : *Les DMP et les PC sont-elles identiques ?* Dans les configurations où les DMP sont différentes, notre méthode les détecte correctement et nous sommes capables de dire quelle DMP est plus petite. Dans le cas où il y a une différence de PC, nous n'arrivons pas toujours à la détecter. Néanmoins nous pouvons toujours indiquer quelle caméra est la plus nette.

Nos résultats obtenus avec les disparités *vérité terrain* et les disparités *calculées* pour les images synthétiques sont très similaires. Nous pouvons donc affirmer que la méthode choisie pour le calcul des disparités s'adapte à notre besoin (images floues en entrée).

En revanche, les résultats sur les images réelles montrent les limitations de notre méthode. Des études plus approfondies devraient en élucider la cause. Le modèle de flou théorique doit-il être adapté à celui d'une vraie caméra ? Jusqu'à quel point la méthode proposée est-elle robuste à la dégradation de la mise en correspondance ? Nous pouvons donc conclure que la méthode proposée n'est pas encore capable de guider un système automatique de contrôle de mise au point. Néanmoins, le retour à l'utilisateur sous la forme de zébrures sur les images permet de lever des alertes et de guider l'intervention de l'opérateur.

Toutes les étapes de la méthode peuvent être calculées en temps-réel : les cartes de disparité, la différence de SML, l'estimation du modèle, sa classification et le retour sous forme de zébrures. La méthode finale pourrait donc être implantée en temps-réel et intégrée dans des outils de tournage stéréoscopique pour aider les opérateurs à résoudre les différences de MAP sur le terrain.

Des travaux sont en cours pour améliorer la méthode, en introduisant un modèle théorique plus précis de la différence de SML sur des images naturelles, basé sur le fait que le spectre de puissance des images nettes est approximativement en $1/f^2$.

Remerciements :

Ces travaux ont été réalisés dans le cadre du projet Action3DS financé par la Caisse des Dépôts et Consignations.

Bibliographie

- Devernay F. (2014a). *Office focus dataset*. <http://devernay.free.fr/vision/focus/office>. (Acces-
sed: 2013-09-11)
- Devernay F. (2014b). *Patio focus dataset*. <http://devernay.free.fr/vision/focus/patio>. (Accessed:
2013-09-11)

- FitzGerrell A. R., Dowski Jr. E. R., Cathey W. T. (1997). Defocus transfer function for circularly symmetric pupils. *Applied Optics*, vol. 36, n° 23, p. 5796–5804.
- Huang W., Jing Z. (2007). Evaluation of focus measures in multi-focus image fusion. *Pattern Recognition Letters*, vol. 28, n° 4, p. 493–500.
- Kooi F. L., Toet A. (2004). Visual comfort of binocular and 3D displays. *Displays*, vol. 25, n° 2, p. 99–108.
- Lambooij M. T. M., IJsselsteijn W. A., Heynderickx I. (2007). Visual discomfort in stereoscopic displays: a review. *Proc. SPIE Stereoscopic Displays and Virtual Reality Systems XIV*, vol. 6490, n° 1.
- Lin H.-Y., Gu K.-D. (2007). Photo-realistic depth-of-field effects synthesis based on real camera parameters. In *Advances in visual computing*, p. 298–309. Springer.
- Mendiburu B. (2012). *3D TV and 3D cinema: tools and processes for creative stereoscopy*. Focal Press.
- Mennucci A., Soatto S. (1999). On observing shape from defocused images. In *International conference on image analysis and processing*, p. 550–555.
- Nayar S. K., Nakagawa Y. (1994). Shape from focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, n° 8, p. 824–831.
- Pentland A. P. (1987). A new sense for depth of field. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, n° 4, p. 523–531.
- Piqueres J. V. (2003). *The persistence of ignorance*. <http://www.ignorancia.org>. (Accessed: 2013-09-11)
- Rajagopalan A., Chaudhuri S., Mudénagudi U. (2004). Depth estimation and image restoration using defocused stereo pairs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, n° 11, p. 1521–1525.
- Schechner Y. Y., Kiryati N. (2000). Depth from defocus vs. stereo: How different really are they? *International Journal of Computer Vision*, vol. 39, n° 2, p. 141–162.
- Seuntjens P., Meesters L., IJsselsteijn W. (2006). Perceived quality of compressed stereoscopic images: Effects of symmetric and asymmetric jpeg coding and camera separation. *ACM Transactions on Applied Perception (TAP)*, vol. 3, n° 2, p. 95–109.
- Sizintsev M., Wildes R. P. (2010). Coarse-to-fine stereo vision with accurate 3d boundaries. *Image and Vision Computing*, vol. 28, n° 3, p. 352–366.
- Stelmach L. B., Tam W. J., Meegan D. V., Vincent A., Coriveau P. (2000). Human perception of mismatched stereoscopic 3D inputs. In *Proceedings international conference on image processing (icip)*, vol. 1, p. 5–8.
- Subbarao M. (1988). Parallel depth recovery by changing camera parameters. In *Proceeding of the international conference on computer vision (iccv)*, p. 149–155.
- Young C. (1992). *Pov-ray - the persistence of vision raytracer*. www.povray.org. (Accessed: 2013-09-11)

Article soumis le 11 octobre 2013.

Accepté le 4 avril 2014.