Interpolation de points de vue : approches directe et variationnelle

Sergi Pujades, Frédéric Devernay

Interpolation de points de vue : approches directe et variationnelle

Sergi Pujades Rocamora[1]
sergi.pujades-rocamora@inria.fr

Frédéric Devernay[1]
frederic.devernay@inria.fr

[1] Inria - Equipe PRIMA,
Univ. Grenoble Alpes, LIG, F-38000 Grenoble, France,
CNRS, LIG, F-38000 Grenoble, France.

Résumé

*Nous abordons la problématique d'interpolation de points de vue à partir d'une paire d'images stéréoscopique. Ces techniques comportent généralement 3 étapes : l'estimation des correspondances entre les vues, le dosage des contributions de chaque image dans la vue finale, et le rendu. D'un côté, tandis que l'état de l'art est très vaste dans l'estimation des correspondances, nous trouvons peu de travaux formels analysant quel est le "*bon*" dosage des contributions lors du mélange des images. D'un autre côté, concernant le rendu de nouveaux points de vue nous identifions deux groupes de méthodes bien distincts, les méthodes "*directes*", et les méthodes "*variationnelles*". Nous conduisons une étude pour analyser la performance des facteurs de dosage ainsi que l'impact de la méthode utilisée sur le résultat final obtenu. Nous évaluons ces méthodes sur des scènes lambertiennes et non-lambertiennes afin de voir, dans chaque cas, quel choix est le plus pertinent.*

Mots Clef

Interpolation de points de vue, rendu d'image basé image, poids de mélange, méthode variationnelle.

Abstract

*We address the topic of novel view synthesis from a stereoscopic pair of images. The techniques have mainly 3 stages : the reconstruction of correspondences between the views, the estimation of the blending factor of each view for the final view, and the rendering. On the one hand, the state of the art has mainly focused on the correspondence topic. Yet, little work addresses the question of which blending factors are best. On the other hand, we find two kinds of rendering methods : "*direct*" methods, defining the final image as a function of the original images, and "*variational*" methods, defining the image as the solution minimising an energy. We present experiments on lambertian and non-lambertian scenes to find out which blending factors perform best, and when a kind of method should be preferred.*

Keywords

Viewpoint interpolation, image-based rendering, blending factors, variational method.

1 Introduction

Nous abordons la problématique d'interpolation de points de vue à partir d'une paire d'images stéréoscopique. Ce domaine a été largement étudié, par exemple pour la génération de contenu pour la télévision 3D multi-points de vue sans lunettes.

Ces techniques comportent généralement 3 étapes : l'estimation des correspondances entre les vues disponibles et la nouvelle vue recherchée, le dosage des contributions de chaque image dans la vue finale, et enfin le rendu.

Tandis que l'état de l'art est très vaste dans l'estimation des correspondances, nous trouvons peu de travaux formels analysant quels sont les "*bons*" facteurs de dosage lors du mélange des images. Les facteurs les plus considérés sont la déformation locale de l'image lors du changement de point de vue et la distance entre le nouveau point de vue et les points de vue originaux. Plusieurs questions se posent : Comment justifier chacun de ces deux choix ? Y a-t-il une différence significative de résultat selon le dosage choisi ? Et dans ce cas, quel dosage doit être préféré ?

Concernant le rendu de nouveaux points de vue, nous identifions deux groupes de méthodes bien distincts : les méthodes "*directes*" et les méthodes "*variationnelles*". La plupart des méthodes de l'état de l'art sont directes : la couleur d'un pixel dans l'image rendue est une fonction des couleurs des pixels correspondants dans les images de référence.

Nous proposons de comparer les méthodes directes avec les méthodes variationnelles. Avec ces dernières, l'image rendue minimise une énergie correspondant à un maximum a posteriori, dérivé d'un modèle génératif. Grâce au formalisme bayésien, les poids de mélange entre les images sont déduits formellement. Ces optimisations sont plus lourdes en temps de calcul que les méthodes directes. Il est donc pertinent d'évaluer si les résultats de ces méthodes compensent leur complexité.

Dans cet article, nous conduisons une étude pour analy-

ser l'impact des facteurs de dosage et de la méthode utilisée sur le résultat final. Nous évaluons ces méthodes sur des scènes lambertiennes et non-lambertiennes afin de voir, dans chaque cas, quel choix est le plus pertinent.

Les expériences réalisées montrent que pour des scènes lambertiennes, le choix du poids de mélange a un impact négligeable. La qualité des résultats obtenus avec des méthodes directes et variationnelles est équivalente. Dans le cas non-lambertien, les poids de mélange considérant la variation angulaire produisent une amélioration de performance très légère et seulement visible sur la génération de points de vue très proches des images initiales. Toujours dans le cas non-lambertien, les méthodes variationnelles sont capables de mieux rendre la structure de l'image, donnant un meilleur résultat.

1.1 Travaux antérieurs

Les techniques d'interpolation de points de vue appartiennent au vaste domaine du rendu basé image [11]. *Unstructured Lumigraph* [2] introduit des propriétés souhaitables pour les méthodes de rendu d'images, notamment pour le dosage des images de référence dans l'image cible. Ces propriétés sont : "*utilisation d'un proxy géométrique*", "*gérer des vues en configuration arbitraire*", "*cohérence épipolaire*", "*distance angulaire minimale*", "*continuité*", "*considérer la résolution des images*", "*cohérence entre rayons optiques équivalents*", et finalement "*temps réel*". Les auteurs proposent aussi une méthode directe qui tient compte de toutes ces propriétés. La "*distance angulaire minimale*" est mesurée entre les rayons optiques de l'image cible et l'image de référence. La propriété "*considérer la résolution des images*" est évaluée avec une approximation de la jacobienne de l'homographie planaire reliant l'image cible avec l'image de référence. L'équilibre entre ces différents facteurs de dosage est ajusté selon la scène. Dans les faits, "*considérer la résolution des images*" a souvent un poids négligeable par rapport à la "*distance angulaire minimale*". Nous chercherons donc à mettre en lumière l'influence de ces poids sur l'image finale.

Parmi les méthodes de génération de points de vue intermédiaires à partir d'une paire stéréoscopique, nous retrouvons ces deux poids de mélange : d'une part certaines méthodes considèrent la distance normalisée α entre l'image cible et les vues références et assignent un poids de mélange α et $(1 - \alpha)$ [9, 13, 3]. D'autre part, des méthodes mesurent la déformation des images [8], avec la jacobienne de l'homographie planaire entre l'image cible et l'image de référence. Toutes ces méthodes ([2, 8, 9, 13, 3]) sont des méthodes directes.

Dans le domaine de la super-résolution, Wanner et Goldluecke [16] proposent une méthode variationnelle très générique pour la génération de nouveaux points de vue. Ils introduisent un modèle génératif décrivant la formation d'une image et établissent l'énergie correspondante au maximum a posteriori en utilisant le formalisme bayésien. Cette formalisation les conduit au poids de mélange

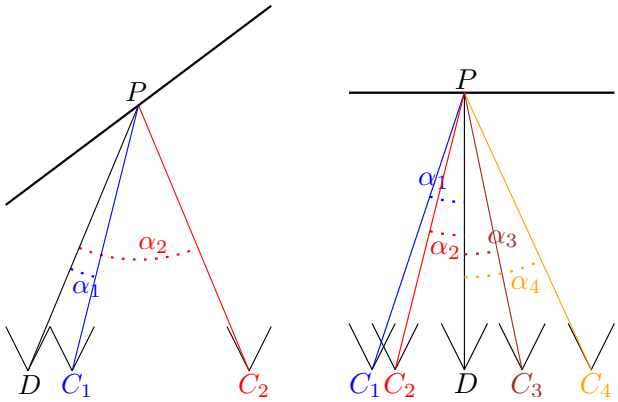

FIGURE 1 – La vue D est générée à partir des vues C_i en utilisant [16]. À gauche : à cause de la déformation de l'image, C_2 sera préférée à C_1, bien que la distance angulaire entre D et C_1 soit plus petite que celle entre D et C_2. À droite : on observe un plan fronto-parallèle. Bien que C_1, C_2, C_3 et C_4 ont des distances angulaires différentes par rapport à D, selon [16], toutes les C_i auront exactement la même contribution lors de la génération de D.

des images donné par le déterminant de la jacobienne de la transformation entre l'image cible et l'image de référence. Les poids de mélange obtenus avec [16] prennent en compte la déformation de l'image. La "*distance angulaire minimale*" n'apparaît pas dans les équations. Dans la Fig. 1 nous illustrons des configurations qui montrent des cas contradictoires avec les heuristiques de [2].

Les méthodes [17, 7] proposent une alternative basé sur la *transformation de maillage* de Gal et al.[4]. Son hypothèse est que les artefacts introduits par les déformations des maillages sont visuellement plus acceptables que ceux produits par le mélange de deux images. Or, dans le cas d'une déformation importante, il serait intéressant de pouvoir fusionner deux images générés avec ces méthodes. Dans ces travaux la question n'est pas évoquée.

Nous observons donc que, dans la littérature, différents poids de mélange sont utilisés, mais à notre connaissance il n'y a pas d'étude comparative qui les confronte pour évaluer leur performance.

2 Méthodes d'interpolation de points de vue

Les méthodes que nous présentons prennent en entrée une paire d'images stéréoscopique rectifiée, une carte de disparité qui donne les correspondances entre les images droite et gauche, et une valeur $\alpha \in [0, 1]$ correspondant à la position intermédiaire du nouveau point de vue à générer.

2.1 Méthodes directes

Les méthodes directes commencent par calculer les deux transformations inverses qui vont de l'image cible à chaque image de référence, en utilisant la carte de profondeur et la valeur α. Les zones n'ayant pas de correspondant sont mar-

quées comme invalides et n'auront pas d'information. La création de l'image cible se fait en regardant la couleur du pixel correspondant dans l'image de référence, en utilisant une interpolation linéaire.

2.2 Poids pour les méthodes directes

Nous avons choisi d'utiliser 4 poids de mélange différents dans nos expériences avec les méthodes directes. Tout d'abord nous considérons la pondération "*classique*" $((1-\alpha), \alpha)$ qui respecte la "*distance angulaire minimale*". Un deuxième choix qui respecte cette propriété est la pondération $((1-\alpha)^2, \alpha^2)$, normalisée par la somme des poids.
Le troisième choix ignore la résolution et la distance angulaire minimale. Elle donne le même poids aux deux images, indépendamment de α ou de la déformation locale introduite par la transformation des images.
Le dernier choix est d'utiliser un poids proportionnel à la déformation introduite par la transformation de l'image, comme décrit dans [16], $|\mathrm{det}D\tau|^{-1}$. Dans notre cas la transformation τ est donnée par la carte de disparité initiale et nous calculons sa déformation par des différences finies. D dénote l'opérateur différentiel et $D\tau$ est la matrice Jacobienne.
Dans tous les cas, nous normalisons les poids afin que leur somme soit 1. Dans le cas où une des deux images n'a pas d'information, effet souvent crée par le manque de visibilité d'un point de vue, nous donnons un poids 1 à l'image qui propose une information et 0 à celle qui n'en propose pas. Si aucune image de référence ne propose de valeur, nous considérons la couleur du pixel comme invalide.
Finalement, pour chaque pixel, nous faisons une somme pondérée de la couleur proposée par chaque vue de référence. Notons qu'à ce stade il y a des pixels qui n'ont pas d'information. Nous abordons cette problématique dans la section 2.5.

2.3 Méthodes variationnelles

Les méthodes variationnelles proposent de trouver l'image qui correspond au minimum d'une énergie. Ces énergies ont souvent deux termes : un terme d'attache aux données et un terme de régularisation.

$$E(u) = E_{\text{données}}(u) + \lambda E_{\text{a priori}}(u). \quad (1)$$

Dans le formalisme bayésien, on peut déduire les termes de ces énergies en modélisant le processus de formation d'image. Une hypothèse très répandue est de considérer le bruit d'observation comme étant gaussien. L'énergie associé au maximum a posteriori de ce modèle correspond au minimum aux sens des moindres carrés d'un système surdimensionné.
Nous notons l'image cible u , v_i l'image de référence numéro i, Ω_i le domaine de définition des images, m_i un opérateur binaire qui indique si le pixel est visible dans l'image finale et τ_i la transformation entre l'image de référence i et l'image cible. m_i et τ_i sont calculés à partir des images de référence (Sec. 2.6). ω_i est le poids de mélange associé aux pixels de l'image i et le centre d'intérêt de notre étude (Sec. 2.4). n est le nombre d'images. Dans notre cas $n = 2$. Le terme d'attache aux données est :

$$E_{\text{données}}(u) = \sum_{i=1}^{n} \frac{1}{2} \int_{\Omega_i} \omega_i m_i ((u \circ \tau_i) - v_i)^2 \quad (2)$$

2.4 Poids pour les méthodes variationnelles

Pour les méthodes variationnelles, nous avons utilisé deux poids différents.
Le premier poids considéré est celui proposé par [16] qui correspond à la déformation de l'image lors de sa transformation :

$$\omega_i = |\mathrm{det}D\tau_i|^{-1} \quad (3)$$

Le deuxième poids utilisé est la pondération "*classique*" $((1-\alpha),\ \alpha)$, qui prend en compte la "*distance angulaire minimale*". Bien que ce poids n'ait pas de modèle génératif associé connu, l'énergie reste bien définie et peut être minimisée (voir Sec. 2.5).
Les poids sont normalisés : $\sum_{i=1}^{n} \omega_i = 1$.

2.5 Modèle de connaissance *a priori*

L'hypothèse de connaissance *a priori* est nécessaire dans la formulation bayésienne afin de donner un arbitrage quand nous n'avons pas d'information sur certaines zones, ou quand nous avons plusieurs candidats pour une solution. Dans les méthodes directes ce phénomène apparaît aussi lorsqu'aucune image de référence n'apporte de l'information à l'image cible.
Précisons que, dans cet article, nous ne cherchons pas à obtenir les meilleures images possibles, mais à faire une comparaison raisonnable entre les méthodes "*directes*" et "*variationnelles*". La détermination de meilleures *hypothèses a priori* pour des images naturelles reste un sujet de recherche à part entière.
Une hypothèse *a priori* "*classique*" en traitement d'image est la variation totale [10] :

$$E_{\text{a priori}}(u) = \int_{\Omega} |\nabla u| \,. \quad (4)$$

Ω est le domaine de l'image recherché, ∇ est l'opérateur gradient et $|.|$ est la norme l_2. Ce terme de l'énergie a la propriété importante de définir une énergie convexe. Comme la dérivée du terme $E_{\text{données}}$ est Lipschitz continue [16], tous les éléments sont réunis pour l'utilisation de FISTA [1], une méthode d'optimisation convexe efficace de l'état de l'art.
Dans toutes nos expériences, le paramètre λ de l'éq. 1 a été choisi de façon empirique ($\lambda = 0.15$) afin de ne pas obtenir de zones sans information ni une image trop lisse par rapport à l'originale. Parmi les valeurs dans l'intervalle $[0.12, 0.18]$ nous n'avons pas observé de différence significative de comportement. En dehors de cette intervalle les résultats se dégradent.

Dans les méthodes "*directes*", nous utilisons une technique d'interpolation pour les zones où nous n'avons pas d'information. Cette technique propage la couleur du plus proche voisin. Elle a un comportement comparable au rôle de l'hypothèse *a priori* [10] lors de la minimisation de l'énergie. Elle nous a donc paru pertinente afin de pouvoir comparer les résultats des méthodes directes et variationnelles.

2.6 Estimation du proxy géométrique

Pour l'étude comparative, nous avons utilisé des mises en correspondance obtenues avec des méthodes de l'état de l'art [12, 5]. Les cartes de profondeur obtenues avec la méthode [5] ne sont pas denses. Nous avons complété ces cartes de profondeur avec la méthode d'interpolation utilisée dans [12]. Cette méthode propage la profondeur des régions voisines en générant des profondeurs cohérentes, c.-à-d. sans occlusion de régions existantes.

3 Expériences

3.1 Jeux de données

Pour comparer les différentes méthodes et l'impact des poids de mélange nous avons utilisé plusieurs jeux de données provenant de deux bases : le "*Middlebury Stereo Dataset*" [6] et le "*Stanford Lightfield Archive*" [14]. Dans le premier la nature des scènes est très lambertienne, c.-à-d. que la couleur d'un point de la scène ne dépend pas du point de vue. Chaque scène a 6 prises de vues alignées et nous avons utilisé les images extrêmes (1 et 6) pour la génération des points de vue restants.
La deuxième collection de jeux de données présente des scènes plus complexes, contenant des reflets spéculaires, des transparences et des inter-réflexions. Chaque jeu de données a 256 vues par scène, dans une configuration structurée de 16x16. Nous avons sélectionné une ligne intermédiaire contenant 16 images (la 8) et utilisé ses points de vue extrêmes (1 et 16) pour la génération des points de vue restants. De cette manière nous avons une vérité terrain avec laquelle nous pouvons évaluer la qualité des images interpolées. Dans la Fig. 2 nous montrons des exemples d'images utilisées lors des expériences.

3.2 Mesures de qualité d'image

Pour comparer les résultats obtenus avec la vérité terrain, nous avons utilisé deux mesures de l'état de l'art. Le "*Peak Signal to Noise Ratio*" (PSNR), exprimé en dB, est largement utilisé dans la littérature. Plus le PSNR est grand, meilleur est le signal. Sa valeur est représentative de la qualité du signal mais ne peut pas être considérée comme une mesure objective de la qualité visuelle d'une image, puisque le calcul se fait pixel à pixel. La mesure "*Structural SIMilarity*" [15] (SSIM) a été développée pour mesurer la similarité de structure entre deux images. L'hypothèse est que l'œil humain est plus sensible aux changement dans la structure de l'image, plutôt qu'à une différence pixel à pixel. Pour évaluer nos expériences nous utilisons une distance basée sur la SSIM : $\text{DSSIM} = \frac{1-\text{SSIM}}{2}$. Sa magnitude n'a pas d'unités et plus la valeur est petite, plus semblables sont les images.

FIGURE 2 – De gauche à droite : l'image "*aloe*" du jeux de données [6], puis "*tarot*", "*bracelet*" et "*amethyst*" du jeux de données [14].

3.3 Temps de calcul

La méthode directe peut être implémentée facilement sur une carte graphique et a un temps de calcul de l'ordre de la fréquence image d'une vidéo (1/30s) pour des images de taille 1024x1024. Elle peut donc être considérée comme temps-réel.
L'optimisation des méthodes variationnelles se fait en utilisant la méthode FISTA [1]. Elle est initialisée avec l'image constituée par le mélange direct des images de référence qui utilise les poids étudiés. Cette initialisation peut avoir des pixels sans information. Bien qu'elle soit aussi implémentée sur carte graphique, son temps de calcul est de l'ordre de la seconde pour des images de taille 1024x1024, et ne peut donc pas être considérée comme temps-réel.

3.4 Résultats

Nous analysons les résultats obtenus avec les jeux de données "*tarot*", "*bracelet*" et "*amethyst*" en utilisant les disparités calculées avec la méthode SGBM [5]. Les résultats obtenus sur les jeux de données "*aloe*" du "*Middlebury Dataset*", qui sont très lambertiens, sont très similaires à ceux de "*tarot*" et ne sont pas présentés. De plus, les résultats obtenus en utilisant les disparités calculées avec la méthode [12] montrent les mêmes tendances que ceux présentés avec la méthode SGBM [5].
Les images des jeux de donnes sont encodées dans l'espace sRGB. Il est conseillé de les convertir en RGB-linéaire lorsqu'on fait des opérations entre les pixels. Dans nos expériences, nous avons réalisé des tests comparatifs en utilisant la version sRGB et la version linéarisée. Pour les jeux de données utilisés, les résultats sont très similaires.

TAROT	*vue 2*		*vue 3*		*vue 8*		*vue 9*		*vue 14*		*vue 15*	
Méthodes directes												
$\alpha(1-\alpha)$	**33.65**	**25**	**31.76**	35	**30.19**	**42**	30.84	**36**	**32.60**	**29**	**33.81**	**24**
$\alpha^2(1-\alpha)^2$	33.49	26	31.47	*38*	30.21	**42**	30.78	**36**	32.38	31	33.58	26
Constants	32.20	27	31.24	**34**	30.16	**42**	**30.88**	**36**	31.61	33	32.64	26
Déformation	*31.81*	*29*	*30.94*	35	29.96	44	30.67	37	31.42	34	32.40	27
Méthodes variationnelles												
$\alpha(1-\alpha)$	33.60	**25**	31.65	35	29.95	*45*	30.45	*39*	31.96	33	33.10	27
Déformation ([16])	32.48	27	31.31	35	*29.78*	*45*	*30.29*	*39*	*31.22*	*36*	*32.33*	*28*

BRACELET	*vue 2*		*vue 3*		*vue 8*		*vue 9*		*vue 14*		*vue 15*	
Méthodes directes												
$\alpha(1-\alpha)$	36.12	**14**	32.87	28	**33.81**	**24**	33.69	**24**	33.71	**23**	35.67	**16**
$\alpha^2(1-\alpha)^2$	36.06	15	32.66	29	**33.81**	**24**	**33.71**	**24**	33.46	25	35.48	**16**
Constants	34.20	20	32.50	30	33.79	**24**	33.67	25	33.21	26	34.59	19
Déformation	*33.68*	*23*	*32.20*	*32*	33.46	25	33.40	26	*32.94*	*28*	*34.25*	*21*
Méthodes variationnelles												
$\alpha(1-\alpha)$	**36.27**	**14**	**32.92**	**27**	33.46	26	33.30	*27*	**33.80**	24	**35.77**	**16**
Déformation ([16])	34.52	19	32.44	30	*33.11*	*27*	*33.15*	*27*	33.06	27	34.47	*20*

TABLE 1 – Résultats numériques de la comparaison entre les images réelles et celles obtenues par les méthodes, sur les jeux de données "*tarot*" et "*bracelet*". Pour chaque vue et méthode, nous présentons les mesures de PSNR en dB (plus la valeur est grande, meilleur est le signal) et les mesures de DSSIM avec un facteur 10^{-4} (plus la valeur est petite, plus la similarité est grande). Les vues (2, 3, 14 et 15) sont des positions proches des vues de référence. Les vues (8 et 9) sont les points de vues centraux. La valeur de la meilleure méthode pour chaque vue et pour chaque jeu de données est en gras. La valeur de la méthode la moins bonne en italique.

La scène "*tarot*" comporte une boule avec des transparences qui viole le modèle lambertien, mais le reste de l'image ne présente pas de changement de couleur entre une vue et l'autre. La scène "*bracelet*" contient quelques reflets spéculaires sur le métal, mais la différence de couleur entre l'image droite et gauche est très faible. Finalement la scène "*amethyst*" est plus complexe. Elle présente des reflets et des inter-réflexions.

Dans le Tab. 1, si nous comparons les méthodes pour chaque vue, nous n'observons pas en général d'écart significatif entre les résultats obtenus par les différentes méthodes. Les valeurs de PSNR et DDSIM sont très proches. Pour la scène "*bracelet*", les résultats des vues 2 et 15 sont légèrement meilleurs avec les méthodes, directes ou variationnelles, qui prennent en compte la distance du point de vue $(\alpha, (1-\alpha))$ et $(\alpha^2\,(1-\alpha)^2)$. Ceci est cohérent car ces méthodes reconstruisent mieux les scènes avec des reflets. Cependant, notons que cette amélioration s'estompe rapidement lorsque nous passons aux vues suivantes (3 ou 14). Finalement, nous n'observons pas non plus d'écart entre les résultats obtenus avec les méthodes directes et les méthodes variationnelles.

Notons que dans le Tab. 1 ("*tarot*" et "*bracelet*") et le Tab. 2 ("*amethyste*") les valeurs de PSNR restent dans les mêmes magnitudes. Par contre les valeurs DSSIM sont beaucoup plus élevées dans le Tab. 2 ("*amethyste*") que les valeurs obtenues avec les autres scènes (Tab. 1). Ceci est dû à la complexité de la scène et à la capacité de la mesure DSSIM de mieux comparer la qualité visuelle. Vue par vue, nous observons des tendances similaires que dans le Tab.1. Par contre nous observons un écart significatif entre les valeurs DSSIM obtenues avec les méthodes directes et variationnelles. Dans des scènes plus complexes, les méthodes variationnelles sont capables de mieux reconstruire la structure de l'image.

3.5 Discussion

Lorsque la scène est lambertienne, les résultats montrent que le choix du poids de mélange est superflu. Il n'y a pas de raison d'en préférer un à un autre. La vue choisie importe peu, puisque les couleurs des deux images sont très semblables. Les erreurs de mise en correspondance, qui ont des différences de couleur plus importantes, ont un effet marginal sur la différence de qualité globale de l'image finale. Toujours dans le cas lambertien, nous observons que les méthodes "*directes*" et les méthodes "*variationnelles*" produisent des résultats de qualité équivalente. Nous pouvons donc conclure, que dans le cas lambertien, les méthodes variationnelles ne représentent pas un avantage qualitatif vis-à-vis des méthodes directes, alors qu'elles sont plus complexes.

Dans le cas où la scène n'est pas lambertienne, pour les vues très proches des images de référence, nous obtenons de meilleurs résultats lorsque le mélange prend en compte la distance du point de vue. Cet effet est léger mais visible. Par contre il s'estompe rapidement dès que nous nous éloignons de la vue de référence. Nous observons cet effet pour les méthodes "*directes*", comme pour les "*variationnelles*". Pour les scènes complexes, nous observons que les résultats des mesures de qualité structurelle obtenus avec les méthodes "*variationnelles*" sont supérieurs à ceux des méthodes "*directes*". Les méthodes variationnelles, bien que

AMETHYST	*vue 2*		*vue 3*		*vue 8*		*vue 9*		*vue 14*		*vue 15*	
Méthodes directes												
$\alpha(1-\alpha)$	**33.84**	1247	**32.67**	1265	30.66	1292	30.88	1299	33.60	1305	34.86	1305
$\alpha^2(1-\alpha)^2$	33.81	*1255*	32.61	*1282*	**30.69**	1294	30.84	1300	33.42	*1325*	34.79	*1313*
Constants	32.16	1218	31.62	1239	30.63	1288	**30.91**	1294	32.67	1273	33.13	1274
Déformation	*31.56*	1239	31.07	1259	*30.27*	*1307*	*30.54*	*1313*	*32.00*	1293	*32.36*	1296
Méthodes variationnelles												
$\alpha(1-\alpha)$	32.80	**1093**	31.39	**1116**	30.30	**1160**	30.66	**1168**	**33.77**	**1142**	**34.98**	**1141**
Déformation ([16])	31.87	1107	*30.96*	1126	*30.27*	1163	30.63	1170	33.01	1150	33.68	1152

TABLE 2 – Résultats numériques de la comparaison entre les images réelles et celles obtenues par les méthodes, sur le jeu de données "*amethyst*". Pour chaque vue et méthode, nous présentons les mesures de PSNR en dB (plus la valeur est grande, meilleur est le signal) et les mesures de DSSIM avec un facteur 10^{-4} (plus la valeur est petite, plus la similarité est grande). Les vues (2, 3, 14 et 15) sont des positions proches des vues de référence. Les vues (8 et 9) sont les points de vues centraux. La valeur de la meilleure méthode pour chaque vue et pour chaque jeu de données est en gras. La valeur de la méthode la moins bonne en italique.

plus lentes, sont capables de mieux reconstruire la structure de la scène dans le cas non-lambertien. La variante proposée avec les poids $((1-\alpha),\alpha)$ a un léger avantage qualitatif lorsque l'image cible est proche d'une vue de référence. Bien qu'elle n'ait pas de "*justification bayésienne*", elle est légèrement plus performante.

4 Conclusions et perspectives

Dans cet article nous présentons une étude sur l'impact du choix des poids de mélange dans les méthodes d'interpolation de points de vue. Nous comparons plusieurs poids de mélange ainsi que deux familles de méthodes ("*directes*" et "*variationnelles*") pour la génération de points de vue.

Lorsque la scène est lambertienne, le choix du poids de mélange est superflu. Lorsque la scène est presque lambertienne, la méthode directe avec le poids $(\alpha,(1-\alpha))$ est légèrement préférable, grâce à sa simplicité. Dans ces deux cas il n'y a pas de raison de devoir appliquer une méthode variationnelle avec une minimisation plus lente.

Dans le cas de scènes plus complexes, les méthodes variationnelles ont un intérêt vis-à-vis de leur meilleure capacité à reconstruire la structure de l'image. Leur coût supplémentaire de calcul est compensé par une meilleure qualité d'image. Dans le cadre des expériences présentées, avec 2 vues de référence, la méthode proposée (variationnelle avec $(\alpha,(1-\alpha))$) améliore les résultats de la méthode variationnelle [16].

Concernant la suite des recherches, une perspective intéressante est d'analyser jusqu'à quelle quantité de données *non lambertiennes* les méthodes variationnelles restent compétitives. Le choix de la méthode peut-il être déduit des images en entrée de façon robuste ?

Dans cet article nous avons mis l'accent sur l'interpolation de points de vue à partir d'une paire d'images stéréoscopique. Il serait intéressant de mener cette étude comparative sur des configurations plus génériques, c.-à-d. avec un nombre quelconque de caméras et une position arbitraire d'image cible.

La généralisation naïve de $(\alpha,(1-\alpha))$ pour le cas avec plus de 2 images devient difficile à justifier, puisque la distance des centres optiques ne représentent plus la "*distance angulaire minimale*". Plusieurs questions se posent : Dans le cas de multiples images de référence, les méthodes variationnelles peuvent-elles apporter une qualité supérieure ? Est-il possible de trouver un nouveau modèle génératif qui déduit formellement l'utilisation de la distance angulaire dans les poids de mélange ? Ce modèle devrait-il explicitement considérer les scènes non-lambertiennes ?

Remerciements

Ces travaux sont réalisés dans le cadre du projet collaboratif Action3DS financé par la *Caisse de dépôts*.

Références

[1] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1) :183–202, 2009.

[2] C. Buehler, M. Bosse, L. McMillan, S. Gortler, and M. Cohen. Unstructured Lumigraph rendering. In *Proc. SIGGRAPH*, pages 425–432. ACM, 2001.

[3] F. Devernay and A. R. Peon. Novel view synthesis for stereoscopic cinema : detecting and removing artifacts. In *Proc. 3DVP Workshop*, pages 25–30. ACM, 2010.

[4] R. Gal, O. Sorkine, and D. Cohen-Or. Feature-aware texturing. In *Proc. Eurographics conference on Rendering Techniques*, pages 297–303. Eurographics Association, 2006.

[5] H. Hirschmuller. Accurate and efficient stereo processing by semi-global matching and mutual information. In *Proc. CVPR*, volume 2, pages 807–814. IEEE, 2005.

[6] H. Hirschmuller and D. Scharstein. Evaluation of cost functions for stereo matching. In *Proc. CVPR*, pages 1–8. IEEE, 2007.

[7] M. Lang, A. Hornung, O. Wang, S. Poulakos, A. Smolic, and M. Gross. Nonlinear disparity mapping for stereoscopic 3d. *Proc. TOG*, 29(4) :75, 2010.

[8] M. Lhuillier and L. Quan. Image interpolation by joint view triangulation. In *Proc. CVPR*, volume 2. IEEE, 1999.

[9] J. H. Park and H. W. Park. Fast view interpolation of stereo images using image gradient and disparity triangulation. *Signal Processing : Image Communication*, 18(5) :401–416, 2003.

[10] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D : Nonlinear Phenomena*, 60(1) :259–268, 1992.

[11] H.-Y. Shum, S.-C. Chan, and S. B. Kang. *Image-based rendering*. Springer, 2007.

[12] M. Sizintsev and R. P. Wildes. Coarse-to-fine stereo vision with accurate 3d boundaries. *Image and Vision Computing*, 28(3) :352–366, 2010.

[13] A. Smolic, K. Muller, K. Dix, P. Merkle, P. Kauff, and T. Wiegand. Intermediate view interpolation based on multiview video plus depth for advanced 3D video systems. In *Proc. ICIP*, pages 2448–2451. IEEE, 2008.

[14] V. Vaish and A. Adams. The (New) Stanford Light Field Archive. `http://lightfield.stanford.edu`, 2008.

[15] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment : from error visibility to structural similarity. *Image Processing*, 13(4) :600–612, 2004.

[16] S. Wanner and B. Goldluecke. Spatial and angular variational super-resolution of 4D light fields. In *Proc. ECCV*, pages 608–621. Springer, 2012.

[17] T. Yan, R. W. Lau, Y. Xu, and L. Huang. Depth mapping for stereoscopic videos. *IJCV*, pages 1–15, 2013.