Filtrage de la profondeur pour le plaquage de texture dans la télévision 3D

I. Daribo[1] C. Tillier[1] B. Pesquet-Popescu[1]

[1] Laboratoire LTCI, UMR CNRS 5141

Ecole Nationale Supérieure des Télécommunications (ENST)
Département de Traitement du Signal et des Images (TSI)
46 rue Barrault, 75634 Paris Cedex 13, France

{daribo, tillier, pesquet}@enst.fr

Résumé

Le rendu à base d'images de profondeur est le procédé permettant de synthétiser de nouvelles vues virtuelles à partir d'une vue réelle et de l'information de profondeur associée pour chaque pixel. Le principal problème dans cette étape de traitement est de gérer les zones nouvellement exposées (trous) qui apparaissent dans les images virtuelles. Une manière classique de diminuer le nombre de ces trous est de prétraiter la carte de profondeur, avant le plaquage de texture. Dans cet article, nous présentons une nouvelle technique de filtrage pour le rendu à base d'images de profondeur. Nous pouvons remarquer en effet qu'il est nécessaire pour diminuer le nombre de trous d'appliquer un filtrage lisseur dans la carte de profondeur près des contours des objets, mais qu'il est inutile de filtrer les zones déjà relativement lisses. Notre solution est basée sur un filtre gaussien pondéré prenant en compte la distance d'un pixel aux contours. De cette manière, nous diminuons les distorsions géométriques et la complexité par rapport à un filtrage uniforme de la carte de profondeur. Les résultats présents dans ce papier se situent dans le contexte de la création de vue stéréoscopique pour la télévision 3D.

Mots clefs

image de profondeur, plaquage de texture, télévision tridimensionnelle.

1 Introduction

Après la haute définition (HD), la prochaine révolution de la télévision sera d'intégrer la perception de la vidéo en trois dimensions. D'importantes recherches en vision stéréoscopique et en perception tridimensionnelle ont permis le développement de technologies connues à ce jour, dans des domaines d'applications telles que le cinéma digital, les salles IMAX, la médecine. L'un des inconvénients liés à ces technologies est la nécessité d'utiliser des dispositifs spéciaux supplémentaires pour l'utilisateur (généralement des lunettes). Néanmoins, le développement de la télévision numérique et des supports autostéréoscopiques permettent aujourd'hui d'imaginer l'introduction de la 3D dans des applications broadcast telles que la télévision. Il n'en reste pas moins que l'acquisition et la transmission du contenu autostéréoscopique doivent respecter les contraintes qu'impose le broadcasting, et principalement deux d'entre elles : la possibilité de s'adapter aux différents types de récepteur (en terme de taille, de nombre de vues, etc..), et la rétro-compatiblitié permettant de visualiser un contenu 2D sur un écran non-autostéréoscopique classique.

Deux approches sont généralement utilisées pour générer du contenu stéréoscopique. La première méthode utilise une paire de cameras traditionnelles, positionnées de manière à reproduire le système visuel humain, chaque camera correspondant à un oeil. Les deux séquences monoscopiques capturées sont ainsi transmises à l'utilisateur. Ce procédé a l'avantage de transmettre l'information telle qu'elle aurait été perçue dans la vie réelle. La deuxième solution consiste à transmettre seulement une vidéo de couleur monoscopique accompagnée d'une information de profondeur associée à chaque pixel. Dans ce cas, une ou plusieurs vues "virtuelles" peuvent être synthétisées du coté utilisateur au moyen d'algorithmes de rendu à base d'images de profondeur.

La première solution a l'avantage de fournir une parfaite vue à chaque oeil, mais on peut constater au moins deux défauts. Premièrement, ce système est optimisé, durant l'acquisition de la scène, pour une configuration précise du récepteur (taille et nombre de vues) ne permettant pas un ajustement facile de l'effet de profondeur. En second lieu, la quantité d'information à transmettre (deux vidéos de couleur monoscopiques) est assez importante. En revanche, la deuxième solution consistant à transmettre une vidéo couleur et sa vidéo de profondeur associée, ne permet pas de générer de parfaites vues virtuelles, mais présente d'autres avantages. Premièrement, la bande passante totale utilisée pour la transmission est réduite. Des études expérimentales sur le codage d'une vidéo de profondeur [1] ont montré qu'une bonne sensation de relief pouvait être obtenue en n'utilisant pour le codage de la vidéo de profondeur que 20% du débit utilisé pour la vidéo de couleur.

De plus, l'adaptation nécessaire pour les différents types de récepteur est automatiquement obtenue par la possibilité de synthetiser toutes les vues "virtuelles" désirées.
C'est pour ces avantages que la transmission d'une vidéo de couleur et de la vidéo de profondeur correspondante est parfaitement adaptée à la télévision 3D, et a été déjà particulièrement étudié au sein du projet européen 'Advanced Three-Dimensional Television System Technologies" (ATTEST) [2].
L'inconvénient le plus important de cette solution se situe dans le rendu à base d'images de profondeur, aussi appelé "3D warping" dans la littérature de l'informatique graphique [3]. Le principe est d'effectuer un changement de repère successif pour chaque point de l'image de la camera de départ. En premier lieu, on effectue une projection du plan 2D de la caméra de départ vers un repère global 3D, puis le point 3D obtenu est ensuite projeté sur le plan 2D de la caméra virtuelle voulue. En raison des fortes discontinuités présentes dans la carte de profondeur, le processus de projection de pixels révèle des zones cachées dans la vue d'origine qui deviennent visibles dans les vues virtuelles. Pour traiter ce problème, nous pouvons utiliser des filtres lisseurs ou des techniques d'interpolation plus complexes [4] pour remplir ces zones nouvelles.
Dans ce papier, nous présentons une nouvelle technique de filtrage pour le rendu à base d'images, permettant de réduire ou d'enlever complètement les trous générés sans dégrader toute la carte de profondeur. Cette solution utilise la distance aux contours des objets pour pondérer un filtre gaussien. Les travaux précédents sont introduits dans la section 2, suivie de la description de notre méthode dans la section 3. La section 4 présente nos résultats expérimentaux et nous finirons par une conclusion dans la section 5.

2 Travaux précédents

Dans le contexte de la télévision 3D, la vidéo couleur monoscopique accompagnée de la vidéo de profondeur doivent être compressées, transmises via les infrastructures traditionnelles 2D de la télévision numérique, et être rendue au niveau du récepteur au moyen d'algorithme de rendu à base d'images de profondeur comme illustré sur la Figure 1.

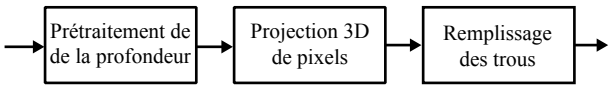


Figure 1 – *Rendu à base d'images de profondeur*

2.1 Prétraitement de la carte de profondeur

Un faible changement de point de vue, après une projection de pixel, crée des trous dans l'image virtuelle. Des prétraitements [4], [5], [6], [7] ont été proposés de manière à réduire ces trous. Nous présentons quelques-uns d'entre eux ci-dessous.

Lissage de la carte de profondeur. De manière générale, la carte de profondeur est filtrée comme dans [4] au moyen de lisseurs connus comme le filtre moyenneur, le filtre gaussien, ou encore le filtre médian. Ces filtres ont la particularité de réduire les fortes discontinuités, et donc les artefacts produits au voisinage des contours des objets lors du processus de projection de pixels. Le filtre gaussien est le plus utilisé et des expériences [5] ont montré que la perception de la qualité d'image stéréoscopique augmente avec la force du lissage appliquée sur la carte de profondeur.

Filtrage asymétrique de la carte de profondeur. Appliquer un lissage à la carte de profondeur permet de réduire aisément les fortes transitions mais l'utilisation des filtres symétriques introduisent des distorsions géométriques : des contours verticaux se courbent. Une manière de diminuer l'influence de ce phénomène est d'utiliser des filtres asymétriques [6], en appliquant un lissage plus important dans la direction verticale. Ainsi, les distorsions géométriques sont réduites, mais toute la carte de profondeur est lissée alors qu'un traitement localisé sur les contours aurait été suffisant.

Lissage localisé sur les contours de la carte de profondeur. Un filtre a été proposé dans [7] permettant de localiser le filtrage uniquement sur les contours, plutôt que de lisser toute la carte de profondeur, diminuant ainsi la dégradation sur l'information de profondeur. De ce fait, les trous identifiés comme gros sont réduits en taille, permettant une interpolation plus facile des zones à combler.

2.2 Plaquage de texture

Nous pouvons distinguer 2 rôles différents pour la vidéo monoscopique couleur. Le premier est de la considérer comme la vue centrale, et dans ce cas-là nous créons les vues virtuelles de gauche et de droite en effectuant une translation du point de vue et une rotation de cette vue centrale. Le second rôle que peut jouer la vidéo monoscopique couleur est d'être directement la vue de gauche ou celle de droite. Dans ce cas, au lieu de générer 2 vues virtuelles, nous n'avons plus besoin que d'en créer une seule puisque nous avons déjà l'autre. Dans la suite de ce papier, nous considérons que nous transmettrons la vue droite. Dans ce cas, la vue virtuelle de gauche est générée à partir d'une translation du point de vue de longueur double, ce qui crée des trous plus gros et plus nombreux. Malgré tout, la qualité de la vue de droite n'a pas été réduite et certaines expériences [8] ont déjà mis en évidence que la perception binoculaire est mieux supportée et la sensation de profondeur est mieux ressentie avec une qualité asymétrique qu'avec une réduction de la qualité des deux vues.
Considérons un système de caméras avec des paramètres connus pour générer le contenu stéréoscopique, dans une configuration où le centre optique est placé à l'infini (projection parallèle) (Fig. 2). La vue projetée est alors obtenue par une première projection, une translation horizontale et une nouvelle projection des pixels. La transformation

qui définit les nouvelles coordonnées de la vue de gauche (x_l, y) à partir des coordonnées de l'image de référence (x_r, y) en prenant en compte la composante de relief Z est la suivante :

$$x_l = x_r + \frac{t_x \times f}{Z} \quad (1)$$

où t_x représente la translation de la caméra horizontale, généralement prise égale à la distance moyenne interoculaire humaine, et f la focale de la caméra de référence.

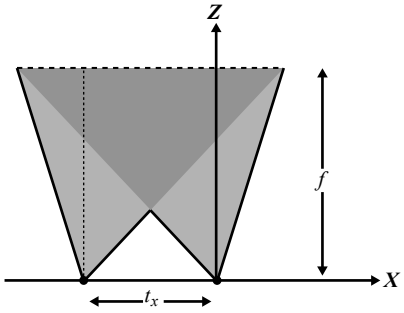


Figure 2 – *Configuration des caméras*

2.3 Remplissage des trous

Le fait de pré-traiter la carte de profondeur permet de réduite le nombre et la taille des trous créés par le plaquage. Néanmoins il arrive que l'opération de filtrage ne suffise pas à supprimer en totalité les trous. Une dernière étape est nécessaire pour traiter ces derniers, en effectuant une interpolation des valeurs manquantes.

3 Lissage de la carte de profondeur en fonction de la carte de distance

Dans le but de réduire ou même d'enlever dans les vues virtuelles générées les nouvelles zones révélées, il est nécessaire de traiter préalablement la carte de profondeur. De manière générale un lissage est effectué, avec un filtre gaussien. Au lieu de lisser entièrement la carte de profondeur, nous proposons un filtre adaptatif prenant en compte la distance aux contours. Un traitement préliminaire est nécessaire afin d'obtenir cette information de distance qui nous permettra d'extraire une région d'intérêt. Par la suite, ces données géométriques nous permettront de calculer les coefficients de pondération nécessaires à l'opération de filtrage.

3.1 Extraction de la région d'intérêt

Nous avons vu plus haut que l'opération de projection 3D révèle des zones de la scène pour lesquelles aucune information n'est présente dans l'image d'origine, comme illustré par la Figure 4.
La carte de disparité est obtenue à partir de la carte de profondeur à l'aide de l'expression suivante

$$\Delta x = \frac{t_x \times f}{Z} \quad (2)$$

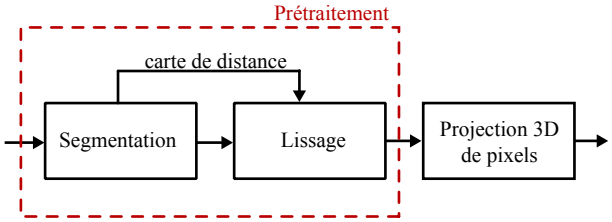


Figure 3 – *Prétaitement de la carte de profondeur suivie de la projection de pixels*

Figure 4 – *En rose : les zones révélées (image provenant de la 1ère image de la séquence Interview)*

où t_x, f et Z ont déjà été définis par (1). Notons que le déplacement des objets au premier plan est susceptible de masquer celui de l'arrière plan. De ce fait, la carte de disparité est calculée dans l'ordre de profondeur, de l'arrière-plan vers le premier plan.
La région d'intérêt est générée à partir de la carte de disparité en appliquant un détecteur de contours directionnel. La carte binaire ainsi générée révèle les déplacements de pixels les plus importants, et ainsi, les zones où il est essentiel de filtrer fortement, permettant une diminution, voire une suppression, des trous présents dans les vues virtuelles.

3.2 Génération de la carte de distance

Le calcul de la distance a pour but d'obtenir pour tout point sa distance à un objet. Dans nos travaux, nous nous servons de cette information de manière à pondérer une adaptation d'un filtre dans le domaine discret. Une valeur nulle dans la carte de distance représente un pixel appartenant à la région d'intérêt. Une valeur non nulle représente la plus petite distance à la région d'intérêt. Parmi toutes les définitions de distance dans l'espace discret, nous avons utilisé la 4-distance. Cette distance est définie pour deux pixels $A(x_A, y_A)$ et $B(x_B, y_B)$ par

$$D(A,B) = |x_A - x_B| - |y_A - y_B|\,, \quad \text{où } A, B \in \mathbb{Z}^2 \quad (3)$$

En prenant en compte la propagation spatiale de la distance, il est possible de calculer la carte de distance successivement d'un pixel à ses voisins, avec un coût de calcul raisonnable. Cette propagation de la distance relève de l'hypothèse qu'on puisse déduire la distance d'un pixel

à partir des valeurs voisines, ce qui est particulièrement adaptée aux algorithmes séquentiels et à la parallélisation. Un exemple de carte de distance est représenté par la Figure 5.

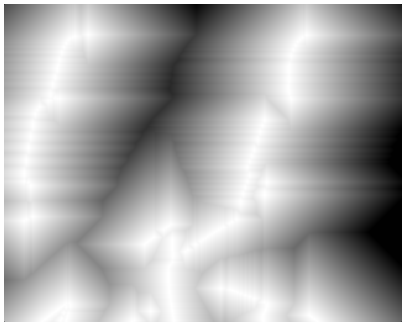


Figure 5 – *Carte de distance utilisant la 4-distance (image provenant de la 1ère image de la séquence Interview)*

3.3 Lissage de la carte de profondeur

Nous avons réuni tout au long des précédentes sous-sections les éléments nécessaires à l'obtention d'un filtre gaussien adaptatif. L'information de distance par pixel rend possible un fort lissage au voisinage d'un contour, et un faible lissage dès qu'on s'éloigne d'un contour. La nouvelle valeur de profondeur de la carte de distance I du pixel $A(x_A, y_A)$ est définie par

$$\alpha(x_A, y_A) \times I(x_A, y_A) + \Big(1 - \alpha(x_A, y_A)\Big) \times g_\sigma(I)(x_A, y_A) \tag{4}$$

où $x_A, y_A \in \mathbb{Z}$, et avec

$$\alpha(x_A, y_A) = \begin{cases} \frac{D(A,B)}{D_{max}} & \text{if } D(A,B) < D_{max} \\ 1 & \text{else} \end{cases} \tag{5}$$

où B, représente le pixel situé sur le contour le plus proche et $\alpha \in [0, 1]$, normalisé par la distance maximum D_{max}, contrôle la force du lissage utilisée au moyen de la carte de distance D. Ainsi, la qualité des zones éloignées des contours présents dans la carte de distance est préservée. La convolution gaussienne g_σ est définie par

$$g_\sigma(I)(x, y) = \sum_{v=\frac{-w}{2}}^{\frac{w}{2}} \sum_{u=\frac{-w}{2}}^{\frac{w}{2}} I(x-u, y-v) G_{2D,\sigma}(u, v) \tag{6}$$

$$G_{2D,\sigma}(x, y) = \frac{1}{2\pi\sigma^2} \exp\left\{-\frac{x^2+y^2}{2\sigma^2}\right\} \tag{7}$$

où la convolution discrète dépend de la taille de la fenêtre w et de la largeur du noyau Gaussien σ. A l'instar de [6], nous avons fixé w égale à 3σ.

4 Résultats expérimentaux

Dans nos travaux, nous avons utilisé la séquence stéréoscopique Interview. Les paramètres expérimentaux de la caméra utilisés sont $t_x = 48$ mm pour la distance inter-caméra et $f = 200$ mm pour la focale. Les paramètres de lissage σ et D_{max} sont respectivement 20 et 55. Nous pouvons observer dans la Figure 6 l'erreur introduite dans les cartes de profondeur filtrées par un procédé de lissage intégral [1] (à gauche) et avec notre méthode (à droite).

(a) Lissage intégrale (b) Lissage en fonction de la distance aux contours

Figure 6 – *Erreur introduite dans les cartes de profondeur*

Nous avons choisi le domaine de la carte de profondeur pour illustrer l'erreur (Figure 6), car notre solution est appliquée dans celui-ci. Cependant cette erreur n'est pas visualisée directement par l'utilisateur. Par conséquent nous avons choisi d'organiser notre modèle de comparaison PSNR en deux parties comme illustré par la Figure 7. La première évaluation de qualité mesure la distorsion de la carte de profondeur juste avant l'étape de rendu à base d'images de profondeur. La seconde évaluation mesure la qualité des images reconstruites après le rendu à base d'images de profondeur.

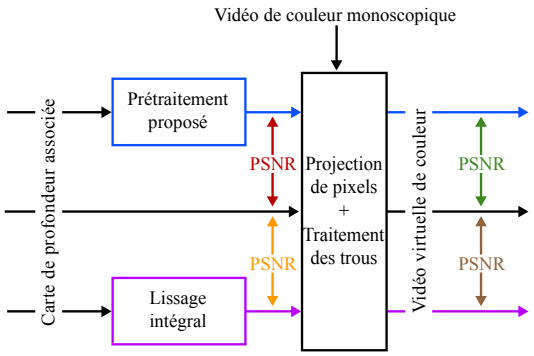
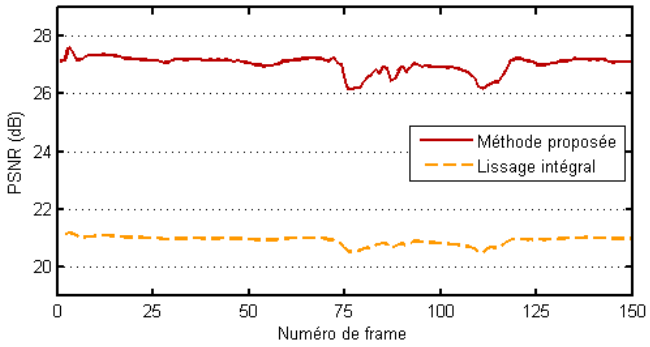


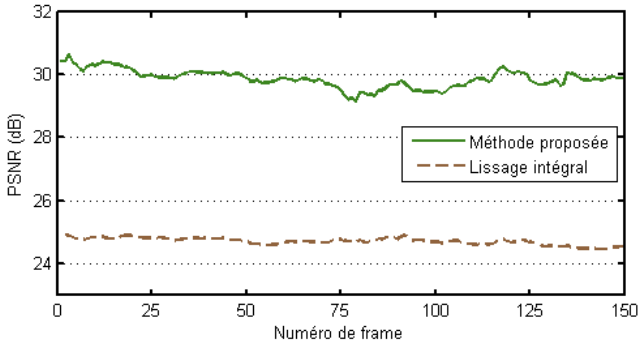
Figure 7 – *Disposition pratique des deux calculs de PSNR*

Dans les deux cas nous comparons notre solution à la méthode classique de lissage intégral, qui consiste à lisser avec un filtre gaussien l'image dans son intégralité. Nous pouvons observer sur la Figure 8 une amélioration importante de la qualité avec notre méthode. De plus, subjectivement nous constatons moins de dégradation sur les images

générées, ce qui s'explique par une meilleure préservation des détails de notre méthode dans la carte de profondeur.



(a) Comparaison PSNR entre les cartes de profondeur



(b) Comparaison PSNR entre les vues générées

Figure 8 – *Comparaison PSNR sur les séquences Interview*

5 Conclusion

Dans ce papier, nous avons introduit un nouveau filtre adaptatif pour le rendu à base d'images de profondeur, prenant en compte la distance aux contours des objets. La méthode proposée a l'avantage de ne pas introduire de distorsion non nécessaire dans la carte de profondeur. Nos résultats expérimentaux montrent un gain important en efficacité de notre méthode. Dans nos prochains travaux nous nous concentrerons sur les caractéristiques géométriques des objets au voisinage des contours.

Références

[1] Christoph FEHN : A 3D-TV approach using depth-image-based rendering (DIBR). *In Proceedings of VIIP 03*, Benalmadena, Spain, septembre 2003.

[2] Christoph FEHN : A 3D-TV system based on video plus depth information. *In Proceedings of Thirty-Seventh Asilomar Conference on Signals, Systems, and Computers*, volume 2, pages 1529–1533, novembre 2003.

[3] R. I. HARTLEY et A. ZISSERMAN : *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN : 0521540518, second édition, 2004.

[4] Christoph FEHN : Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV. *In Proceedings of the SPIE Stereoscopic Displays and Virtual Reality Systems XI*, pages 93–104, San Jose, CA, USA, janvier 2004.

[5] Wa James TAM, Guillaume ALAIN, Liang ZHANG, Taali MARTIN et Ronald RENAUD : Smoothing depth maps for improved steroscopic image quality. *In* Bahram JAVIDI et Fumio OKANO, éditeurs : *Three-Dimensional TV, Video, and Display III*, volume 5599, pages 162–172, 2004.

[6] Liang ZHANG et W.J. TAM : Stereoscopic image generation based on depth images for 3D TV. *IEEE Transactions on Broadcasting*, 51(2):191–199, juin 2005.

[7] Wan-Yu CHEN, Yu-Lin CHANG, Shyh-Feng LIN, Li-Fu DING et Liang-Gee CHEN : Efficient depth image based rendering with edge dependent depth filter and interpolation. *In ICME 2005 : IEEE International Conference on Multimedia and Expo*, pages 1314–1317, 6-8 2005.

[8] Pieter SEUNTIENS, Lydia MEESTERS et Wijnand IJSSELSTEIJN : Perceived quality of compressed stereoscopic images : Effects of symmetric and asymmetric JPEG coding and camera separation. *ACM Transactions on Applied Perception*, 3(2):95–109, janvier 2006.