Extraction d'un Environnement Photométrique et Géométrique pour la Réalité Augmentée

François Fouquet, Jean-Philippe Farrugia, Brice Michoud

AFIG 2008 / Toulouse
IRIT Presse

Extraction d'un Environnement Photométrique et Géométrique pour la Réalité Augmentée

François Fouquet[1,2,3], Jean-Philippe Farrugia[1,2,3], Brice Michoud[1,2,3]

[1]Université de Lyon
[2]Université Claude Bernard Lyon I
[3]LIRIS-CNRS
43 boulevard du 11 novembre 1918, 69622, Villeurbanne, France
francois.fouquet@liris.cnrs.fr

Abstract

Dans cet article, nous proposons une étude du passage en temps réel de deux méthodes d'acquisition, la reconstruction d'images HDR et le depth from focus, en vue de leur utilisation en réalité augmentée. Nous présentons les différentes contraintes de ces méthodes ainsi que les paramètres qui influencent les temps de calcul et la qualité des acquisitions obtenues. Nous décrivons les différents procédés mis en œuvre pour accélérer les calculs ainsi que leurs impacts sur la qualité. Enfin, nous donnons des pistes d'améliorations pour poursuivre l'adaptation de ces méthodes à la réalité augmentée.

Keywords: Imagerie HDR, Vision par ordinateur, Carte de profondeurs, Réalité augmentée, Rendu temps réel

1. Introduction

La réalité augmentée est un ensemble de techniques issues de l'imagerie qui consiste à mêler des éléments synthétiques et des éléments réels dans une même visualisation. Généralement, l'utilisateur peut se déplacer librement dans un environnement réel pour lequel il lui est proposé une visualisation augmentée d'objets synthétiques. Les applications de cette technique sont multiples allant des loisirs numériques (jeux vidéo) à la simulation de situations d'urgence pour la formation des personnels dans les industries à risques (ex : nucléaire). Très souvent la mise en place de la réalité augmentée nécessite du matériel spécialisé voire des lieux aménagés exclusivement pour cela (salle immersive). Dans le cadre de notre projet, un des objectifs est d'ouvrir au plus grand nombre l'accès à ces techniques. Nous avons donc pris le parti de n'utiliser que du matériel non spécifique et déjà très répandu (webcam, GPU...).

Augmenter la réalité signifie mettre en correspondance la représentation que l'on a de l'environnement et celle de l'objet synthétique qu'on souhaite y insérer. En fonction de l'application visée, cette fusion des représentations sera plus ou moins précise ou complète. Par exemple, pour une application d'assistance à la réparation de photocopieur, il faut placer les éléments virtuels très précisément (ie., alignés avec les différentes pièces mécaniques de la machine). En revanche, le ré-éclairage et la projection des ombres des éléments ne sont pas cruciaux pour le bon fonctionnement de l'application. A contrario, la reconstruction d'un monument historique détruit a besoin d'être visuellement convaincante. Dans ce cas, le placement peut être moins précis sans que cela dérange l'utilisateur. En revanche, une erreur d'occultation ou d'éclairage affectera grandement le réalisme.

Pour réaliser une réalité augmentée réaliste, nous nous plaçons dans ce deuxième cas et nous avons donc besoin d'acquérir plusieurs types d'informations sur l'environnement. Actuellement, nos travaux se focalisent sur l'acquisition de la géométrie et de l'éclairage car les deux sont nécessaires pour réaliser un ré-éclairage simple. Un grand nombre de méthodes ont été proposées pour acquérir ces informations à partir d'images d'un ou plusieurs points de vue mais peu d'entre elles sont adaptées à la réalité augmentée et à une contrainte de faible coût matériel.

Dans cet article, nous proposons de sélectionner deux méthodes, une pour la photométrie et l'autre pour la géométrie, et de tester l'influence de certains paramètres d'entrée sur le temps de calcul. L'objectif final est d'obtenir un bon compromis entre temps de calcul et qualité afin d'appliquer ces méthodes en temps réel.

La suite de cet article est composée de trois sections. La première présentera les travaux existants sur l'imagerie HDR et l'estimation de profondeur à partir d'images. La seconde décrira les tests et méthodes permettant une acquisition en temps interactif de la géométrie et la photométrie à partir

d'un flux vidéo. Enfin, nous conclurons en synthétisant les résultats importants puis nous présenterons certaines perspectives de ce travail.

2. Travaux existants

2.1. Acquisition de la photométrie

Les images obtenues avec un capteur classique (appareil photo ou caméra) fournissent généralement des informations suivant trois canaux (rouge, vert et bleu) qui correspondent d'une part à la quantité de lumière renvoyée au capteur par la zone de l'environnement située dans son champ de vision et d'autre part à la couleur de ces zones. Cependant, ces appareils ne permettent pas d'obtenir une image contenant toute la dynamique lumineuse de la scène en une seule prise de vue. Si la scène que l'on observe contient des zones de luminances très fortes (sources lumineuses) ou au contraire très faibles (ombres) alors les images obtenues présenteront des zones sur ou sous-exposées pour lesquelles la quantité de lumière réellement présente dans la scène n'est pas déterminable. Par ailleurs, leurs capteurs CCD ont généralement des fonctions de réponse non linéaires et la quantité de lumière qu'ils reçoivent va elle-même varier en fonction des paramètres de la caméra (temps d'exposition, ouverture...). Déterminer la luminance arrivant réellement au capteur à partir d'une image peut donc s'avérer complexe.

Pour lever ce problème nous utilisons des images HDR (*High Dynamic Range*) qui permettent de stocker l'ensemble des détails de luminance d'un environnement quelle que soit l'intensité de son éclairage. Ces images nécessitent généralement une opération de réduction de dynamique appelée reproduction de tons (*tone mapping*) pour être visualisées sur un écran. Il existe différentes techniques pour reconstruire les images HDR, reposant toutes sur une même idée introduite par Mann et Picard [MP95] d'une part et Madden [Mad93] d'autre part. Le principe est de combiner plusieurs images prises d'un même point de vue mais avec des expositions différentes afin de reconstruire une image contenant tous les détails de luminances des différentes expositions. Debevec et Malik [DM97] proposent d'utiliser des images pour lesquelles ils font uniquement varier le temps d'exposition. Ils déterminent dans un premier temps la fonction de réponse f de l'appareil puis ils reconstruisent ensuite l'image HDR en faisant une moyenne pondérée des images d'entrée transformées par f^{-1}. Mitsunaga et Nayar ont proposé une méthode similaire [MN99] mais faisant en plus l'hypothèse que f est un polynôme. Ils réalisent ensuite une recherche itérative de l'ordre et des coefficients du polynôme correspondant le mieux aux données observées. Cette deuxième méthode présente l'avantage de permettre en plus de la détermination de f, une détermination des ratios d'exposition entre les différentes images de départ. Les expositions de ces images peuvent donc être inconnues contrairement à la méthode de Debevec et Malik. La plupart des techniques actuelles d'imagerie HDR sont rassemblées dans l'ouvrage *"High dynamic range imaging"* [RWPD05]. Pour réaliser nos essais, nous avons choisi d'implémenter la méthode de Debevec et Malik telle qu'elle est décrite dans [DM97].

2.2. Acquisition de la géométrie

Il existe de nombreuses méthodes visant à retrouver une géométrie à partir d'images. Deux descriptions de cette géométrie sont possibles. La première est subjective, les données géométriques sont alors représentées sous la forme d'une carte de profondeurs (*depth map*) donnant les distances des objets depuis le point de vue de l'utilisateur. La deuxième qualifiée d'objective consiste à décrire la géométrie dans un repère global.

Cette représentation globale est utilisée par les méthodes se basant sur la notion d'enveloppe visuelle (*visual hull*) [Lau94]. Elles permettent de déterminer une enveloppe de la forme 3D d'un objet à partir de ses silhouettes 2D issues de plusieurs caméras calibrées. Le principal défaut de cette méthode dite de *shape from silhouette* est qu'il demeure des ambigüités quant aux placements des objets reconstruits dans l'espace. Elle nécessite également de pré-positionner et calibrer plusieurs appareils d'acquisition dans l'environnement afin d'obtenir des résultats ayant une bonne précision. Ceci en fait également une méthode difficile à envisager pour l'exploration d'un environnement inconnu. Parmi les techniques utilisant des images provenant de plusieurs points de vue, nous pouvons également citer l'ensemble des techniques de stéréo-vision qui présentent l'avantage d'être adaptées à une représentation subjective. À partir de deux images d'entrée ayant des points de vue proches, ces techniques vont mettre en correspondance des points de la première et de la deuxième image puis vont déduire la profondeur en fonction du déplacement de ces points entre les deux images et des paramètres intrinsèques et extrinsèques des caméras. Le principe de stéréo-vision et ses contraintes sont notamment décrits dans le livre de O.Faugeras [Fau93] et la plupart des méthodes dérivant de ce principe sont rassemblées dans l'article "*Structure from stereo : a review*" de Dhond et Aggarwal [DA89].

Les méthodes utilisant des images prises d'un même point de vue sont quant à elles mieux adaptées à une réalité augmentée subjective. Nous pouvons là encore distinguer plusieurs approches différentes. Certaines reposent sur un matériel particulier qui va donner de lui même un moyen de connaitre la profondeur des objets présents dans la scène. Parmi ces appareillages spéciaux, nous pouvons notamment citer les systèmes d'ouvertures codées (*coded aperture*) [LFDF07] et les caméras à temps de vol (*Time Of Flight Camera*) [RGY03]. Étant pour la plupart des modèles expérimentaux, ces appareils sortent pour l'instant du cadre de cette étude bien qu'ils présentent un grand intérêt. Une autre méthode, le *shape from shading* [ZTCS99], consiste à analyser les variations d'éclairement à la surface des objets 3D afin de déterminer leur géométrie. Cette méthode

nous est cependant difficile à mettre en œuvre car elle nécessite la connaissance (voire la maitrise) des sources de lumière présentes dans la scène. Enfin, récemment Saxena et al. ont proposé une méthode de détermination des cartes de profondeurs par apprentissage à l'aide de champs de Markov [SCN08]. Cette technique, qui fait écho à la méthode d'analyse de scène proposée par Hoiem [Hoi07], demande toutefois de définir et de calculer un grand nombre de descripteurs sur les images et nécessite une base d'apprentissage conséquente.

Enfin, il existe également des méthodes de reconstruction de la géométrie se basant sur la profondeur de champ de la caméra. En effet, les objets qui se situent dans le plan de focalisation sont nets tandis que ceux situés devant ou derrière ce plan sont flous. De plus, la quantité de flou présente en chaque point est directement dépendante de la distance des objets observés au plan de focalisation. Cette approche a été proposée dans un premier temps par Pentland [Pen87] qui donna également les bases de la méthode *depth from defocus*. Il y modélise la quantité de flou présente en chaque point par une fonction de diffusion (*Point Spread Function*) qu'il estime en analysant les images dans le domaine fréquentiel. La connaissance des paramètres de prise de vue (ouverture et distance focale) permet ensuite d'obtenir la distance du point dans l'espace. Cette méthode a été reprise par la suite dans des nombreux travaux, notamment ceux de Xiong et Shafer [XS93] qui propose l'utilisation de filtres de Gabor pour lever le problème lié au fenêtrage dans la transformée de Fourier locale. D'autres limitations apparaissent lors de l'utilisation de ces méthodes et notamment leur incapacité à déterminer la profondeur des surfaces ne présentant pas de contours marqués (ou de textures). Pour palier cette limitation, Pentland a suggéré l'utilisation d'une source de lumière structurée projetant ainsi la texture nécessaire à l'étude sur les zones uniformes. Cette idée a également été reprise par Nayar et al. [NWN95] qui proposent l'étude de structures de lumière particulières permettant d'estimer des cartes de profondeurs denses en temps réel.

Une dernière approche, appelée *depth from focus*, travaille à partir d'un jeu d'images acquises d'un même point de vue en faisant varier la distance de focalisation. Un estimateur local de focalisation est ensuite calculé sur les points des images afin de déterminer l'image sur laquelle chaque zone de la scène est la mieux focalisée ainsi la profondeur correspondante via le calibrage de la caméra. De nombreux travaux s'appuient sur cette méthode dont ceux de Grossmann [Gro87] et ceux de Nayar et Nakagawa [NN89]. Pour nos essais, nous avons choisi de nous baser sur cette approche.

3. Acquisition de l'environnement

Dans cette partie, nous allons développer les travaux effectués afin d'accélérer les calculs des deux méthodes que nous avons sélectionnées. Pour chacune, nous commencerons par en présenter les grandes lignes, puis nous décrirons les procédés d'accélération utilisés et leur influence sur les résultats obtenus. Enfin, nous proposerons quelques perspectives d'évolution suggérées par nos observations.

3.1. Acquisition de la photométrie

3.1.1. Implémentation de la méthode

La méthode que nous utilisons pour reconstruire des images HDR est celle présentée en 1997 par Debevec et Malik [DM97]. Elle se divise en deux parties distinctes, l'une développant une méthode permettant de déterminer la fonction de réponse de la caméra et l'autre présentant la manière de combiner les images d'entrée par l'intermédiaire de cette fonction. Debevec et Malik ne font varier que le paramètre de temps d'exposition sur les différentes prises de vue d'une même scène. Le jeu d'images ainsi obtenu contient alors toute la gamme des luminances qui s'y trouvent. Les prises de vue faites avec des temps d'exposition courts (ex : 1/1000) contiennent des informations sur les zones renvoyant une grande quantité d'énergie vers le capteur. Inversement, lorsqu'on prend des images avec de longues expositions (ex : 2s), elles contiennent des informations sur les zones de faibles énergie (zones d'ombre).

Pour déterminer la fonction de réponse, les auteurs commencent par sélectionner manuellement des points dans les images. Ensuite, ils recherchent la fonction de réponse passant le mieux par ces points. Ils construisent pour cela un système d'équations fortement surdéterminé. Ces équations prennent en compte à la fois les valeur de luminance récupérées sur chacune des images pour un même point de l'espace et les valeurs de temps d'exposition de ces images. Ce système est ensuite résolu sous sa forme matricielle en utilisant une décomposition en valeurs singulières afin d'obtenir un résultat approché minimisant l'erreur. Pour reconstruire l'image HDR, Debevec et Malik font la moyenne pondérée des valeurs de luminance obtenues en passant chaque image d'entrée par la fonction inverse de la fonction de réponse. L'utilisation d'une fonction de pondération permet de donner une importance plus grande aux valeurs des images d'entrée pour lesquelles les objets sont bien exposés et elle supprime les points pour lesquels les valeurs correspondent à des zones sur ou sous-exposées.

Telle que nous l'avons implémentée, cette méthode permet de reconstruire une image HDR en 4,4s en prenant 12 images de résolution 640×480 en entrée. Ce temps de calcul se répartit pour 3,2s sur la détermination de la fonction de réponse de la caméra et pour 1,2s sur la reconstruction de l'image HDR. Notons que pour éviter d'avoir à recaler les images, nous les prenons à partir d'un appareil posé sur pied. Nous avons également ajouté une sélection de points automatique. Notre étude portant sur l'évaluation de méthodes existantes en vue de leur adaptation à la réalité augmentée, nous avons ensuite cherché à identifier les différents limitations de cette méthode ainsi que les paramètres permettant

d'accélérer les calculs afin de nous rapprocher du temps réel en minimisant les pertes de qualité sur les reconstructions.

3.1.2. Réduction des temps de traitement

Dans un premier temps il est important de remarquer que la fonction de réponse de la caméra peut être considérée comme étant constante. Elle peut donc n'être estimée qu'une seule fois. Ensuite, nous pouvons remarquer que le nombre d'images prises en entrée est en lien direct avec le temps de calcul. Nous avons donc effectué un test pour estimer l'influence de ce paramètre sur les reconstructions. Pour commencer, la fonction de réponse de la caméra est pré-calculée sur le nombre maximum d'images à notre disposition (12 dans notre cas). Puis nous effectuons des reconstructions avec 2, 4 et 6 images en entrée ainsi qu'une reconstruction de référence utilisant les 12 images (figure 1). Les reconstructions sont enfin affichées avec un outil de visualisation par plages colorées où chaque demi ordre de grandeur de luminance correspond à une couleur. Cet outil permet d'apprécier rapidement la cohérence d'une reconstruction par rapport à la composition de la scène. Nous avons également proposé un outil permettant de comparer les différentes reconstructions avec la référence. Il renvoie une image où la couleur de chaque pixel correspond à la distance relative (exprimée en pourcentage) entre l'image et la référence. Les résultats présentés sur la figure 2 soulignent qu'avec la méthode utilisée, les reconstructions restent cohérentes quel que soit le nombre d'images en entrée et qu'en utilisant plus de 4 images les écarts à la référence sont limités.

La réduction du nombre d'images en entrée implique inévitablement de faire un choix dans les temps d'exposition sélectionnés pour faire la reconstruction. Bien que n'ayant aucune influence sur le temps de calcul, ce choix joue un rôle primordial dans la qualité des images HDR reconstruites. Un test similaire au précédent a permis de vérifier que le choix de temps d'exposition régulièrement répartis sur le domaine des temps d'exposition potentiellement porteurs d'informations offre les meilleurs résultats. Notons par ailleurs que seule une partie de ce domaine est intéressante pour la réalité augmentée : le ré-éclairage d'objets est fait à partir des sources de lumière de l'environnement qui sont observables sur les images à expositions courtes, les détails dans les zones de faible intensité ne sont donc pas importants pour cette application. Cela nous permet d'enlever les images à temps d'exposition élevé (qui sont donc très longues à acquérir) des données d'entrée. Ce test met également en évidence que les pertes de qualité que nous observons sont dans une certaine mesure prévisibles. En effet, les erreurs qui apparaissent se situent systématiquement dans des zones où nous avons enlevé l'image correspondant au temps d'exposition le plus porteur d'information (ex : si une image avec un temps d'exposition court est enlevée, il apparait des erreurs dans les zones de forte luminance).

3.1.3. Reconstruction sur des flux d'images

Pour tester la faisabilité d'une acquisition HDR en utilisant des webcams, nous avons évalué la vitesse de l'algorithme en lui envoyant des flux vidéo en entrée. Ce test montre que la reconstruction s'exécute à 10 frames par seconde (FPS) pour 2 flux en entrée, 7 FPS pour 4 flux et un peu plus de 5 FPS pour 6 flux. Nous avons utilisé des vidéos de résolution 320 x 240. Le temps de reconstruction étant directement dépendant du nombre de pixels des images à reconstruire, ceci divise d'emblée le temps de calcul par 4 par rapport à l'essai précédent. Il est important de noter que les résultats obtenus sont réalisés uniquement avec un CPU standard et sans optimisation algorithmique ou matérielle particulière. Enfin, il faut remarquer que lors de ce test, aucun recalage des images d'entrée n'est nécessaire. Or, dans une application réelle, l'utilisateur peut être en mouvement et le recalage devient alors indispensable. Cette phase ajoutera un temps de calcul impossible à négliger. En effet, les méthodes de recalage classiques basées sur les contours sont inadaptées car des images d'une même scène ayant des expositions différentes n'ont pas forcément de contours communs. La méthode de recalage *Mean Threshold Bitmap Alignment Technic* [War03] proposée pour répondre spécifiquement au problème précédent est également peu satisfaisante car elle est trop lente (environ 1s reconstruction) et ne tient pas compte d'une éventuelle rotation entre les images.

3.1.4. Évolutions de l'acquisition photométrique

Pour répondre au problème précédent, nous prévoyons de tester et développer d'autres méthodes de recalage dans nos travaux futurs. Nous envisageons également de faire des jeux de données étalons avec des caméras scientifiques afin d'évaluer les pertes dues à l'utilisation de webcams. Nous envisageons également de tester d'autres méthodes permettant de déterminer les fonctions de réponse des caméras car même si leur temps d'exécution n'est pas un facteur limitant pour la réalité augmentée, la précision de leur estimation est très importante pour la qualité des reconstructions.

3.2. Acquisition de la géométrie

3.2.1. Travail préliminaire

Comme nous l'avons dit précédemment, nous avons utilisé une méthode de "*depth from focus*" pour extraire des informations géométriques d'un jeu d'images. Le principe de cette méthode est donc de prendre une série d'images d'un même point de vue en faisant uniquement varier le paramètre de focalisation de la caméra. Ensuite, nous appliquons en chaque point de ces images un estimateur renvoyant une valeur représentative du niveau de flou. Nous reconstruisons ensuite une unique image contenant en chaque point la valeur du paramètre de focalisation de l'image présentant la meilleure focalisation (la plus nette). Enfin, cette image est convertie en carte de profondeurs en appliquant une table

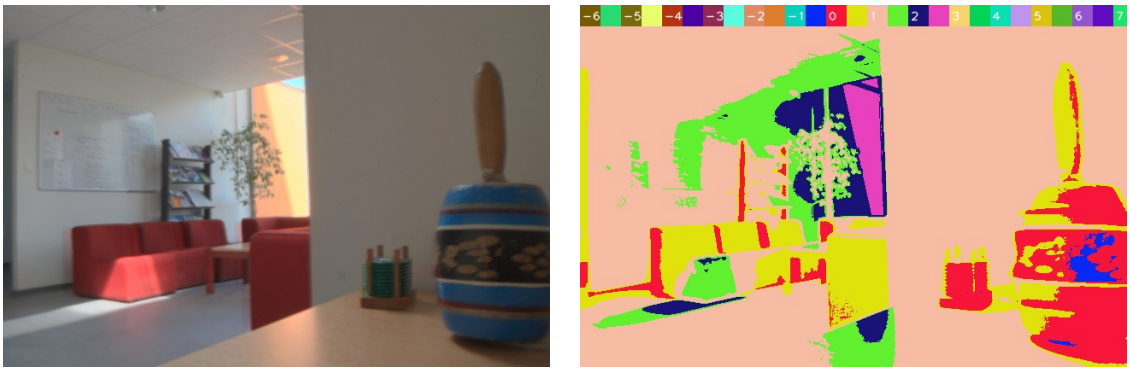
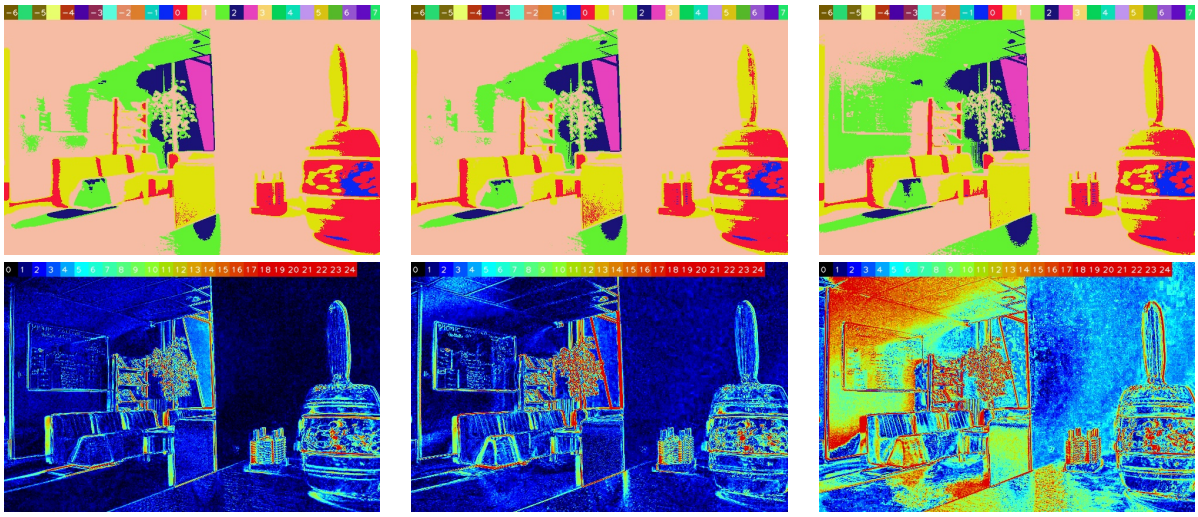


Figure 1: *Reconstruction de référence sur 12 images. A gauche : tone mapping de l'image HDR. A droite : visualisation en plages colorées.*



6 images 4 images 2 images

Figure 2: *Test du nombre d'images d'entrée pour la reconstruction. En haut : visualisation en plages colorées. En bas : comparaison avec l'image de référence. L'échelle s'étend du noir pour les pixels ayant des valeurs identiques jusqu'au rouge pour des pixels ayant 25% ou plus de différence.*

de correspondance (ou LUT : *Look Up Table*) calculée lors d'une étape de calibrage.

Choix de l'estimateur

Avant de choisir un estimateur de focalisation, il faut remarquer que, sur une image, le flou se traduit par une diffusion de l'énergie lumineuse d'un pixel sur ses voisins. Ceci implique une diminution des contrastes sur les contours. Si une scène présente des contours marqués, un moyen simple d'estimer si une image de cette scène est plus ou moins focalisée qu'une autre est, pour tout point des images, d'étudier les variations d'intensité le long des contours. Pour nos essais nous avons donc choisi la norme du laplacien comme estimateur de focalisation car elle est simple et rapide à évaluer. Plus cette norme prendra une valeur élevée et plus nous pourrons considérer que les variations de pentes des contours sont fortes et donc que l'image est nette en ce point. Cet estimateur sera cependant très sensible au bruit et il est possible d'utiliser d'autres estimateurs de focalisation plus complexes afin d'améliorer la qualité de l'estimation [NN89].

Calibrage

Maintenant que nous avons un moyen de comparer le niveau de flou en un point sur des images ayant des paramètres de focalisation différents, il nous faut associer chaque paramètre de focalisation à une distance dans l'espace. Pour notre application, nous avons utilisé une webcam logitech 9000Pro qui permet de faire varier le paramètre de focali-

sation sur une échelle de 256 valeurs. Nous avons réalisé un calibrage avec une mire afin d'obtenir une LUT entre ce paramètre et la distance de focalisation réelle. Lors du calibrage, nous avons placé la mire à une distance connue de la caméra puis nous avons pris un jeu d'images en faisant varier le paramètre de focalisation sur toute la gamme disponible (de 0 à 255). Si nous considèrons que tous les points du plan de la mire sont à la même distance de la caméra (ce qui est vrai si la mire est suffisamment loin de la caméra), nous pouvons calculer pour toutes les images la somme des estimateurs locaux de focalisation. Nous déterminons ensuite le maximum de cette somme en fonction du paramètre de focalisation. S'il est unique et aisé à déterminer, alors la valeur du paramètre de focalisation pour lequel il est atteint sera associée à la distance de la mire. Cette expérience est ensuite réitérée pour d'autres distances de mire afin de construire le tableau de correspondance entre la distance et le paramètre de focalisation.

3.2.2. Reconstruction des cartes de profondeurs

Pour reconstruire les cartes de profondeurs, nous avons procédé de la même manière que nous l'avions fait pour le calibrage. Nous commençons par réaliser un jeu d'images en faisant varier le paramètre de focalisation de la caméra. Il nous faut ensuite déterminer pour chaque point dans quelle image il est le mieux focalisé. Pour cela, nous calculons en chaque point de chaque image de l'ensemble d'entrée l'estimateur de focalisation puis nous ne conservons que la valeur du paramètre de focalisation de l'image pour lequel cet estimateur est maximum. Cette méthode présente cependant l'inconvénient majeur de calculer des estimations pour tous les points, y compris ceux situés sur des zones non texturées. L'information de contour recherchée n'étant pas présente en ces points, l'estimateur renvoie des valeurs très faibles pour toutes les images. C'est alors le paramètre de focalisation pour lequel le bruit est maximum qui sera conservé. De ce fait, nous avons décidé d'ajouter un seuillage sur l'estimateur qui permet d'éliminer les points sur lesquels l'attribution d'un plan de profondeur n'est pas fiable. Pour finir, nous reconstruisons la carte de profondeurs en remplaçant les valeurs des paramètres de focalisation enregistrés par les valeurs de distances correspondantes dans la LUT calculée lors du calibrage. La figure 3 présente un résultat obtenu avec cette méthode. Pour cette image, le temps de reconstruction est de 9s à partir de 256 images d'une résolution de 640×480.

Il est important de remarquer dans cet essai que la qualité de la carte de profondeurs obtenue est directement dépendante du seuil choisi et que le seuil "optimum" va varier d'un jeu d'images à l'autre. Il est donc nécessaire d'avoir un autre indicateur permettant d'avoir une idée du comportement du couple "jeu d'images/seuil". Nous avons choisi, pour faire cette vérification, de calculer le pourcentage de points déterminés. Il correspond au nombre de points ayant passé au moins une fois le seuil sur le nombre de points total de l'image. Cet indicateur va notamment nous permettre de comparer les images de profondeurs entre elles en considérant que deux images sont comparables si elles ont des taux de points conservés identiques. Même si cet indicateur est plus cohérent que la valeur de seuil pour l'évaluation des images, sa valeur optimale ne peut pas être déterminée une fois pour toute car elle dépend du pourcentage de points de contour présents dans les images d'entrée qui va lui-même grandement varier en fonction de la scène observée. Prendre un pourcentage de points conservés plus élevé que le pourcentage de points de contours fera alors apparaitre des "contours fantômes" à côté des contours réels ainsi que du bruit. Enfin, nous supposons actuellement qu'une autre source d'erreur de cette méthode pourrait être la présence de maxima locaux dus au bruit et qui produisent des maxima globaux à des endroits où ils ne devraient pas se trouver. Une méthode intéressante pour supprimer ce type d'erreur serait de prendre pour chaque point la valeur médiane de tous les paramètres de focalisation conservés lors du seuillage. Il faut néanmoins pour cela faire l'hypothèse que l'estimateur de focalisation suit approximativement une évolution gaussienne de part et d'autre du paramètre de focalisation le maximisant.

3.2.3. Réduction des temps de traitement

Réduction du nombre d'images

Comme pour l'acquisition photométrique, nous souhaitons adapter l'acquisition de la géométrie à la réalité augmentée et donc en réduire le temps de calcul. Lors du calibrage que nous avons réalisé, nous avons remarqué que, pour des raisons de limitation du matériel, les images dont les paramètres de focalisation sont aux extrêmes de la plage disponible n'apportent pas d'information pour la détermination de la carte de profondeurs. Le domaine des paramètres utiles s'étend en réalité entre les valeurs 50 et 250. La suppression de ces images du jeu de départ ainsi qu'une réduction plus globale du nombre d'images peuvent donc être envisagées. Pour tester les conséquences d'une telle réduction sur les résultats, nous avons refait 3 jeux de données en prenant régulièrement 1 image sur 10 pour le premier, 1 image sur 20 pour le deuxième et 1 image sur 40 pour le dernier. Notons que de par le principe de la méthode, les images de profondeurs obtenues auront autant de plans de profondeur différents qu'il y avait d'images différentes dans le jeu de départ (soit respectivement 20, 10 et 5). Par ailleurs, les images conservées dans les nouveaux jeux de données ne correspondent pas à un découpage régulier de l'espace des profondeurs. Ce découpage est cependant suffisant pour reconstruire des cartes de profondeurs et en vérifier leur cohérence. Les résultats en terme de temps de calcul sur des images 640×480 sont regroupés dans le tableau 1.

Comme le montrent ces résultats, la diminution du nombre d'images en entrée entraine une diminution du temps de calcul qui lui est quasiment proportionnelle. Les

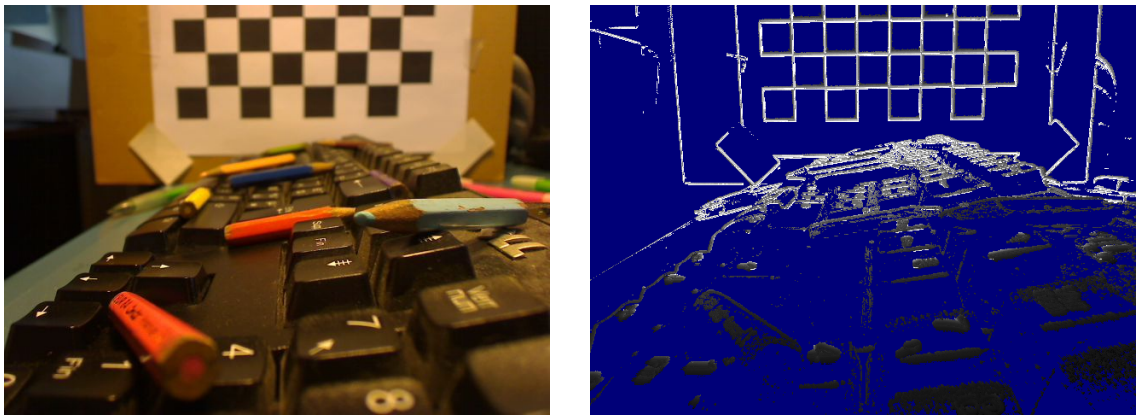


Figure 3: *Exemple de détermination de profondeurs. L'échelle de couleurs varie linéairement du noir (2cm ou moins) au blanc (60cm ou plus). Les pixels bleus correspondent aux points rejetés lors du seuillage.*

Nombre d'images en entrée	Toutes les images (256)	20 images	10 images	5 images
temps	9s	1,5s	0,8s	0,5s

Table 1: *Temps de calcul des cartes de profondeur en fonction du nombre d'image d'entrée*

cartes de profondeurs que nous reconstruisons restent visuellement cohérentes quel que soit le nombre d'images en entrée. La prise en compte d'un nombre moindre d'images en entrée semble donc un bon moyen de gagner du temps afin d'adapter cette méthode à la réalité augmentée.

Réduction de la résolution

Toujours dans l'optique de réduire les temps de calculs, nous avons réalisé un autre essai sur cette méthode consistant à réduire la résolution des images d'entrée. En effet, tout comme nous l'avons supposé pour la luminance, l'information de profondeur nécessaire à l'incrustation d'objets n'a pas besoin d'être très précise et il doit donc être possible d'acquérir des images de résolution moindre pour les applications de réalité augmentée. Nous avons donc refait un test de temps de calcul sur 2 jeux d'images construits en réduisant la résolution des images d'un des ensembles. Les deux nouveaux jeux comportent 256 images faisant respectivement 320×240 et 160×120. Les temps de calcul et images obtenus sont regroupés sur la figure 4.

De nouveau, nous observons une réduction de temps de calcul significative rendant le procédé attractif pour la réalité augmentée. En revanche, nous remarquons que si la réduction à 320×240 donne des cartes qui restent cohérentes, la réduction à 160x120 donne des résultats pour lesquels des erreurs sont très visibles. L'utilisation d'une réduction trop importante n'est donc peut être pas adaptée en l'état mais pourra être testée en implémentant des estimateurs de focalisation différents.

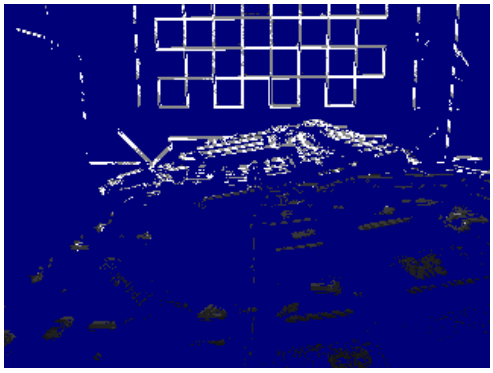
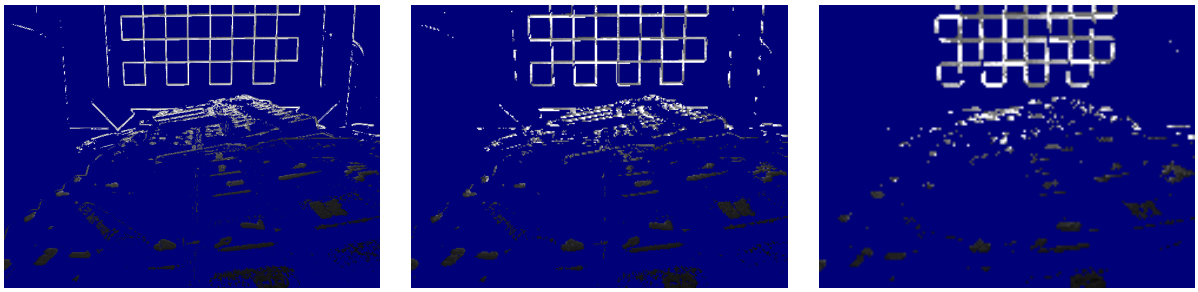


Figure 5: *Exemple de combinaison des réductions de la résolution et du nombre d'images. Échelle de couleurs linéaire du noir (2cm ou moins) au blanc (60cm ou plus). Les pixels bleus correspondent aux points rejetés lors du seuillage.*

Réductions combinées

Pour finir, nous avons voulu faire un essai en combinant les deux types de réduction. Nous avons donc réalisé une reconstruction sur une sélection de 5 images en 320×240. Le temps de calcul observé est de 170ms. L'image obtenue est présentée en figure 5. Ce résultat est très encourageant car il montre qu'actuellement les données peuvent être traitées à la même vitesse qu'elles sont acquises avec la caméra (30FPS) et il est déjà possible d'imaginer une reconstruction de cartes de profondeurs à la volée à une fréquence de 4 ou 5FPS.



640×480
9s

320×240
4,1s

160×120
1,1s

Figure 4: *Temps de calcul des cartes de profondeurs en fonction de la résolution des images. Échelle de couleurs linéaire du noir (2cm ou moins) au blanc (60cm ou plus). Les pixels bleus correspondent aux points rejetés lors du seuillage.*

3.2.4. Évolutions de l'acquisition géométrique

Les tests que nous avons réalisés sur la reconstruction de cartes de profondeur à partir d'images ayant des paramètres de focalisation différents nous permettent donc d'espérer faire des reconstructions en temps interactif avec une caméra. La prochaine étape de nos travaux sera donc d'essayer d'adapter cette reconstruction à des flux vidéos, avec toutes les contraintes nouvelles de recalage que ceci implique. Il serait également souhaitable d'envisager, comme cela a été fait pour les cartes de luminances, de créer une routine permettant de comparer les images résultats entre elles pour évaluer objectivement les pertes de précision lorsqu'on supprime des images. L'utilisation d'une carte de profondeurs de référence peut également s'avérer utile pour valider nos résultats. Nous remarquons enfin que, même si elles restent cohérentes, les reconstructions présentant peu de plans de profondeurs sont assez grossières et elles ne s'avèreront peut être pas suffisantes pour insérer convenablement des objets dans un environnement de réalité augmentée. De même, le fait que les cartes de profondeurs ne soient pas denses en limite l'utilisation et il faudra envisager l'étude de méthodes permettant d'obtenir directement des cartes de profondeurs denses. Un test d'incrustation d'objets permettra de répondre à ces questions de manière plus précise et est également prévu dans nos travaux futurs. Enfin, cette méthode devra être comparée à d'autres qui permettent également de retrouver des informations de profondeur telles que les méthodes de stéréo-vision.

4. Conclusion et perspectives

Les travaux que nous avons développés dans cet article ont permis de mettre en évidence qu'il existait une multitude de méthodes pour acquérir des informations photométriques et géométriques sur un environnement. Nous avons également vu qu'une bonne partie d'entre elles était inadaptée pour des applications de réalité augmentée car trop gourmande en calculs ou incompatible dans leur manière de représenter les données. Nous avons ensuite exposé les essais réalisés pour évaluer la capacité de certaines méthodes a priori plus adaptées à un contexte de réalité augmentée. Ces essais nous ont permis de voir que l'acquisition d'images HDR d'une part et des cartes de profondeurs d'autre part étaient envisageables en temps interactif (quelques frames par seconde) avec du matériel peu onéreux.

Nos travaux ouvrent également de nombreuses perspectives pour l'acquisition d'environnement. Dans un premier temps, il est possible d'envisager l'optimisation algorithmique et matérielle des méthodes existantes. L'acquisition de jeux de données étalons par l'intermédiaire de matériel spécialisé sera également nécessaire pour tester les méthodes que nous employons.

A l'avenir, nous envisageons également d'améliorer la qualité des rendus de notre réalité augmentée. Ceci passera par l'acquisition de nouvelles données sur l'environnement, telles que les caractéristiques de surface des matériaux composant les objets de la scène. Nous pourrons alors tenir compte de phénomènes optiques plus complexes qui ne sont actuellement pas gérés afin d'augmenter le réalisme des rendus. Notons par exemple que les données actuellement récupérées ne permettent pas de savoir s'il y a un miroir dans l'environnement et les calculs de réflexion de l'objet virtuel dans ce miroir ne seront donc pas faits lors du rendu, ce qui nuira grandement à son réalisme.

Enfin, il faut être conscient que les informations que nous récupérons à l'heure actuelle ne sont pas directement intégrables dans une application de réalité augmentée. En effet, les données photométriques reconstruites représentent l'ambiance lumineuse du point de vue de l'utilisateur. Or, pour ré-éclairer un objet, il faut connaitre la lumière arrivant au point de la scène où il doit être inséré. Dans la suite de nos travaux, il nous faudra donc nécessairement rechercher des méthodes permettant d'obtenir cette information à partir de celles que nous connaissons. De même pour les données géométriques, une étape de mise en correspondance entre l'échelle de l'environnement et de celle de l'ob-

jet sera inévitable. Cependant, le contexte de la réalité augmentée est également un avantage car il donne accès à un certains nombre d'informations complémentaires (la position de l'utilisateur, l'orientation de l'axe optique des caméras, de multiples prises de vues proches) sur lesquelles nous pourrons nous appuyer pour créer des méthodes d'acquisitions plus robustes et plus rapides.

5. Remerciements

Ces travaux sont soutenus par l'Agence Nationale pour la Recherche (ANR) dans le cadre du projet ANR-07-MDCO-001. Ils bénéficient également du soutien du projet LIMA du cluster ISLE de la région Rhône-Alpes.

References

[DA89] DHOND U., AGGARWAL J. : Structure from stereo : A review. *IEEE Transactions on System, Man and Cybernetics 19*, 6 (1989), 1489–1510.

[DM97] DEBEVEC P. E., MALIK J. : "recovering high dynamic range radiance maps from photographs". In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques (SIGGRAPH '97)* (1997), pp. 369–378.

[Fau93] FAUGERAS O. : *Three-Dimensional Computer Vision : A Geometric Viewpoint*. MIT Press, 1993.

[Gro87] GROSSMANN P. : Depth from focus. *Pattern Recognition Letters 5*, 1 (1987), 63–69.

[Hoi07] HOIEM D. : *Seeing the World Behind the Image : Spatial Layout for 3D Scene Understanding*. PhD thesis, Carnegie Mellon University, 2007.

[Lau94] LAURENTINI A. : The visual hull concept for silhouette-based image understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence 16*, 2 (1994), 150–162.

[LFDF07] LEVIN A., FERGUS R., DURAND F., FREEMAN W. T. : Image and depth from a conventional camera with a coded aperture. In *Proceedings of the 34th annual conference on Computer graphics and interactive techniques (SIGGRAPH '07)* (2007), pp. 70–78.

[Mad93] MADDEN B. : *Extended Intensity Range Image*. Tech. rep., University of Pennsylvania, 1993.

[MN99] MITSUNAGA T., NAYAR S. : Radiometric Self Calibration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '99)* (1999), vol. 1, pp. 374–380.

[MP95] MANN S., PICARD R. W. : Extending dynamic range by combining different exposed pictures. In *Proceedings of the 48th Annual Conference of The Society for Imaging Science and Technology (IS&T)* (1995), pp. 442–448.

[NN89] NAYAR S., NAKAGAWA Y. : *Shape from Focus*. Tech. rep., The Robotics Institute, Carnegie Mellon University, 1989.

[NWN95] NOGUCHI M., WATANABE M., NAYAR S. : Real-time focus range sensor. In *Procedings of the 5th International Conference on Computer Vision (ICCV '95)* (1995), pp. 995–1001.

[Pen87] PENTLAND A. P. : A new sense for depth of field. *IEEE Transactions on Pattern Analysis and Machine Intelligence 9*, 4 (1987), 523–531.

[RGY03] R. GVILI A. KAPLAN E. O., YAHAV G. : Depth keying. In *Proceedings of SPIE Electronic Imaging Conference* (2003), vol. 5006, pp. 564–574.

[RWPD05] REINHARD E., WARD G., PATTANAIK S., DEBEVEC P. : *High Dynamic Range Imaging : Acquisition, Display, and Image-Based Lighting*. Morgan Kaufmann Publishers Inc., 2005.

[SCN08] SAXENA A., CHUNG S., NG A. : 3-d depth reconstruction from a single still image. *International Journal of Computer Vision 76*, 1 (2008), 53–69.

[War03] WARD G. : Fast, robust image registration for compositing high dynamic range photographs from handheld exposures. *Journal of Graphics Tools 8*, 2 (2003), 17–30.

[XS93] XIONG Y., SHAFER S. : Depth from focusing and defocusing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '93)* (1993), pp. 68–73.

[ZTCS99] ZHANG R., TSAI P., CRYER J., SHAH M. : Shape from shading : A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence 21*, 8 (1999), 690–706.