

Placement optimal de caméras contraintes pour la synthèse de nouvelles vues

Grégoire Nieto

Frédéric Devernay

James Crowley

INRIA Grenoble Rhône-Alpes - équipe PRIMA
LIG Laboratoire d'informatique de Grenoble

gregoire.nieto@inria.fr

Résumé

Nous étudions le problème du placement optimal sous contraintes, de plusieurs caméras, pour la synthèse de nouvelles vues. Une telle configuration optimale est définie comme celle qui minimise l'incertitude de projection des pixels des caméras de prise de vue sur la vue à synthétiser. Le rendu de cette vue est souvent précédé d'une phase de reconstruction 3D approximative. Nous dérivons la matrice de covariance associée à l'incertitude sur la géométrie, puis nous propageons l'erreur sur le plan de la nouvelle vue. Nous observons l'influence de l'interoculaire et de la distance focale des caméras sur l'erreur projetée, pour des distributions de points aléatoires à diverses profondeurs.

Mots Clef

Reconstruction 3D, synthèse de nouvelle vue, rendu basé image, propagation d'erreur.

Abstract

We address the problem of optimal viewpoint placement of several constrained cameras for novel view synthesis. Such an optimal configuration should minimize the uncertainty of the projection of pixels from the shooting cameras onto the novel view. The image-based rendering step is often preceded by a 3D reconstruction step. We first derive the expression of the geometric error covariance matrix, then we project the distribution onto the new sensor plane. We study how the baseline and the focal length of the cameras influence the newly-derived error measure, for randomly distributed points at miscellaneous depths.

Keywords

3D Reconstruction, Novel View Synthesis, Image-Based Rendering, Error Propagation.

1 Introduction

Au cinéma ou à la télévision, nous sommes souvent contraints de placer les caméras à bonne distance de la scène, et par conséquent d'utiliser des longues focales pour se « rapprocher » de l'objet filmé. Celles-ci présentent cependant des inconvénients majeurs comme la déformation de l'arrière plan et la perte de relief, qui rendent difficiles les

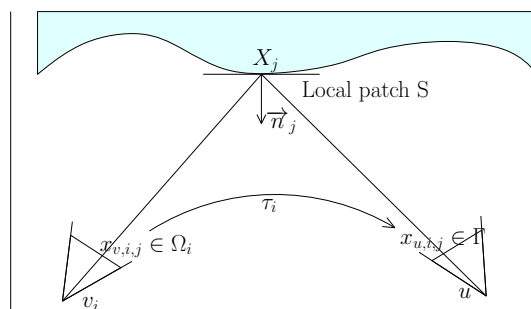


FIGURE 1 – La vue u est synthétisée par les vues v_i . Chaque point image $x_{v,i,j}$ est transformé en point $x_{u,i,j}$ par le warp τ_i .

prises de vues stéréoscopiques avec un relief suffisant. Une solution à ce problème consiste à créer un nouveau point de vue à partir d'un ensemble de prises de vue [2]. L'objectif de notre travail est de trouver la meilleure configuration de caméras de prise de vue, sous contraintes, pour la synthèse d'une vue virtuellement « plus proche » de la scène. Si le problème a été maintes fois traité pour la reconstruction 3D, il est nouveau dans le domaine de l'IBR (*Image-Based Rendering*).

2 Travaux antérieurs

Le rendu basé image a été l'objet de nombreux travaux [1]. Pujades et al. [2] proposent une méthode inspirée du *lumigraph* [3] pour synthétiser de nouveaux points de vue à partir d'un ensemble non structuré de caméras. Les images qu'elles fournissent permettent l'estimation grossière de *proxy* géométriques. La nouvelle image est ensuite obtenue par minimisation d'une fonction de coût dérivée d'une estimation MAP (Maximum a posteriori). La contribution de chaque caméra est fonction de l'incertitude sur la géométrie et du bruit perturbant la mesure des capteurs.

Bien que la recherche du placement optimal pour la reconstruction 3D ne soit pas nouvelle [4, 5], déterminer les paramètres des caméras de prise de vue pour le rendu basé image est inédit.

3 Approche proposée

La synthèse de point de vue par *lumigraph* nécessite une première étape de reconstruction grossière de la géométrie de la scène. Le maillage approximatif ainsi obtenu (ensemble de points 3D X_j et de normales à la surface \vec{n}_j), appelé *proxy*, permet de calculer les *forward warps* $\tau_i : \Omega_i \rightarrow \Gamma$ qui à chaque point image d'une camera de prise de vue $x_{v,i,j} \in \Omega_i$ associe le point image correspondant $x_{u,i,j} \in \Gamma$ dans la vue à synthétiser (Fig. 1).

La mesure du point image $x_{v,i,j}$ n'est pas exacte : on suppose pour simplifier qu'elle est distribuée selon une loi normale centrée sur $\bar{x}_{v,i,j}$ et de covariance $C_{v,i,j}$. De ce fait la projection de ce point image $\tau(x_{v,i,j}) = x_{u,i,j}$ sur la vue à synthétiser suivra aussi une loi normale, car on linéarise le *warp* en l'approximant par son développement en série de Taylor au premier ordre. La covariance de la distribution obtenue est représentée par une ellipse sur le plan image, qui dépend des paramètres des caméras de prise de vue et de la géométrie. Ces ellipses nous serviront dans un premier temps à définir une mesure $\mu(P)$ de la qualité de la nouvelle vue en fonction des paramètres P des caméras de prise de vue : positions et distances focales. Enfin, un algorithme d'optimisation nous permettra de trouver les paramètres qui minimisent globalement μ suivant des contraintes de position et de cadrage.

4 Mesure d'incertitude

Pour un point de l'espace X_j et une camera de prise de vue v_i , le point image correspondant $x_{v,i,j}$ est projeté sur la nouvelle vue u en $x_{u,i,j}$. On définit la mesure d'incertitude par $\mu_{i,j}(P) = \text{trace}(C_{u,i,j})$, $C_{u,i,j}$ étant la covariance de la distribution de $x_{u,i,j}$ que nous souhaitons calculer. La mesure globale de l'image à synthétiser que nous voulons minimiser est $\sum_{i,j} \mu_{i,j}(P)$. Le calcul de $C_{u,i,j}$ nous montre qu'elle peut s'écrire comme la somme de deux termes : un terme dit de résolution et un terme dit géométrique.

4.1 Le terme de résolution

Le terme de résolution est une covariance qui résulte de la propagation de $C_{v,i,j}$ pour la géométrie estimée (\bar{X}_j , \vec{n}_j) d'après la formule de propagation d'erreur [5] (Fig. 2 - gauche). Ce terme pénalise les caméras de prise de vue ayant une courte focale car elle restituent moins de détail de l'objet qui est filmé. En effet plus la distance focale de la caméra est courte, plus le "cône d'incertitude" projeté sera large. Le terme de résolution pénalise aussi les caméras qui regardent la surface de biais car la projection de l'incertitude sur la surface est grande.

4.2 Le terme de géométrie

Le terme de géométrie est une matrice de covariance issue de la projection de l'incertitude du point 3D reconstruit X_j sur la nouvelle vue (Fig. 2 - droite). Il est proportionnel au produit scalaire entre le rayon incident et la normale à la surface, et pénalise donc les caméras éloignées de la vue à synthétiser. Cependant c'est aussi une fonction croissante

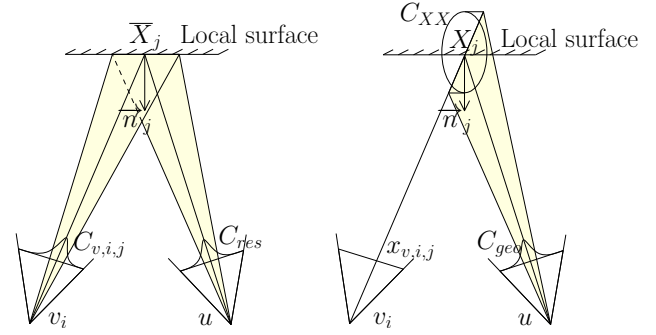


FIGURE 2 – Gauche : le terme de résolution C_{res} est la propagation de $C_{v,i,j}$ pour la surface estimée. Droite : le terme de géométrie C_{geo} est la projection sur la nouvelle vue u de l'incertitude du point d'intersection du rayon incident avec la surface locale, calculée grâce à la covariance C_{XX} du point 3D reconstruit.

de l'incertitude sur la géométrie, qui est faible lorsque les caméras de prise de vue sont bien écartées [4]. La minimisation de ce terme revient à trouver un compromis entre une bonne reconstruction de la géométrie et une faible re-projection de l'incertitude.

5 Minimisation de la mesure d'incertitude

On cherchera d'abord les minima de μ pour des caméras contraintes à un axe situé à une distance donnée de la scène, puis pour des configurations plus originales (dans un plan ou en arc de cercle par exemple). Les caméras, que l'on prendra rectifiées, devront avoir au moins la totalité de la scène dans leur champ, et chaque point 3D devra être vu par au moins deux caméras pour être reconstruit.

La recherche du minimum de $\mu(P)$ se fera grâce à un algorithme d'optimisation classique (Gauss-Newton, Levenberg-Marquardt) ou un algorithme génétique [5].

Références

- [1] S. B. Kang, Y. Li, X. Tong, and H. Y. Shum, Image-Based Rendering, *Foundations and Trends® in Computer Graphics and Vision* 2006, 173 – 258.
- [2] S. Pujades, F. Devernay, and B. Goldluecke, Bayesian View Synthesis and Image-Based Rendering Principles, *CVPR* 2014, 3906 – 3913.
- [3] C. Buehler, M. Bosse, L. McMillan, S. Gortler and M. Cohen, Unstructured Lumigraph Rendering, *SIGGRAPH*, 2001, 425–432.
- [4] C. Beder and R. Steffen, Determining an Initial Image Pair for Fixing the Scale of a 3D Reconstruction from an Image Sequence, *Pattern Recognition, Lecture Notes in Computer Science*, 2006, 657 – 666.
- [5] G. Olague and R. Mohr, Optimal camera placement for accurate reconstruction, *Pattern Recognition*, 2002, 927–944.