



Synthèse de vues pour l'initialisation de pose

Pierre Rolin

► **To cite this version:**

Pierre Rolin. Synthèse de vues pour l'initialisation de pose. Vision par ordinateur et reconnaissance de formes [cs.CV]. Université de Lorraine, 2017. Français. <NNT : 2017LORR0025>. <tel-01536649>

HAL Id: tel-01536649

<https://hal.archives-ouvertes.fr/tel-01536649>

Submitted on 24 Jul 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Résumé

La localisation est un problème récurrent de la vision par ordinateur, avec des applications dans des domaines multiples tels que la robotique ou la réalité augmentée. Dans cette thèse on considère en particulier le problème d'initialisation de la pose, c'est-à-dire la localisation sans information a priori sur la position de la caméra. Nous nous intéressons à la localisation à partir d'une image monoculaire et d'un nuage de points reconstruit à partir d'une séquence d'images. Puisque nous n'avons pas d'a priori sur la position de la caméra, l'estimation de la pose s'appuie sur la recherche de correspondances entre des points de l'image et des points du modèle de la scène. Cette mise en correspondance est difficile en raison de sa combinatoire élevée. Elle peut être mise en défaut lorsque l'image dont on cherche la pose est très différente de celles ayant servi à la construction du modèle, en particulier en présence de forts changements de point de vue.

Cette thèse développe une approche permettant la mise en correspondance image-modèle dans ces situations complexes. Elle consiste à synthétiser localement l'apparence de la scène à partir de points de vue virtuels puis à ajouter au modèle des descripteurs extraits des images synthétisées. Comme le modèle de scène est un nuage de points, la synthèse n'est pas faite par rendu 3D mais utilise des transformations 2D locales des observations connues de la scène. Les contributions suivantes sont apportées. Nous étudions différents modèles de transformation possibles et montrons que la synthèse par homographie est la plus adaptée pour ce type d'application. Nous définissons une méthode de positionnement des points de vue virtuels par rapport à une segmentation de la scène en patchs plans. Nous assurons l'efficacité de l'approche proposée en ne synthétisant que des vues utiles : elles sont éloignées de celles existantes et elles ne se recouvrent pas. Nous vérifions également que la scène est visible à partir des points des vue virtuels pour ne pas produire des vues aberrantes à cause d'occultations. Enfin, nous proposons une méthode de recherche de correspondances image-modèle qui est à la fois rapide et robuste. Cette méthode exploite la répartition non-uniforme des correspondances correctes dans le modèle, ce qui permet de guider leur recherche. Les résultats expérimentaux montrent que la méthode proposée permet de calculer des poses dans des configurations défavorables où les approches standard échouent. De façon générale la précision des poses obtenues augmente significativement lorsque la synthèse de vue est utilisée. Enfin nous montrons que, en facilitant la mise en correspondance image-modèle, cette méthode accélère le calcul de pose.

Mots-clés: Calcul de pose, mise en correspondance, synthèse de vues

Abstract

Localisation is a central problem of computer vision which has numerous applications such as robotics or augmented reality. In this thesis we consider the problem of pose initialisation, which is pose computation without prior knowledge on the camera position. We are interested in pose computation from a single image and a point cloud that has been reconstructed from a set of images. As we do not have prior knowledge on the camera position, pose estimation entirely rely on finding correspondences between the image and the model. The search for these correspondences is a difficult problem because of its high combinatorial complexity. It can fail if the image is very different from the ones we used to construct the model, in particular when there is a large viewpoint change between them.

This thesis proposes an approach to make matching possible in such difficult scenarios. It consists in synthesising locally the appearance of the scene from virtual viewpoints and add descriptors extracted from these synthetic views to the model. Because the scene model is a point cloud, the synthesis is not a 3D rendering but a local 2D transform of existing observations of the scene. The following contributions have been proposed. We study different transform models and show that homographic transformations are the best suited for this application. We define a method to position the virtual viewpoints with respect to a planar segmentation of the scene model. We ensure time efficiency by only synthesising useful views, i.e. views that are far from the existing one and don't overlap. Furthermore we verify that the synthesized surface is visible from the virtual viewpoint to avoid producing aberrant views due to occlusions. Finally, we propose a robust and time efficient method to research image-model correspondences. It uses geometric cues in a guided matching framework to efficiently identify sets of correct correspondences. Experimental results show that the proposed approach makes possible pose computation in situation where standard methods fail. In general the precision and repeatability of computed poses is significantly improved by the use of view synthesis. We also show that it also reduce the pose computation times by making image-model matching easier.

Keywords: Pose computation, matching, view synthesis