Questions/réponses

Classe inversée

Cours Vision

10 octobre 2024

1 Remarques générales

- Si je n'ai pas répondu à votre question ou si la réponse n'est pas satisfaisante, vous avez le droit de la poser à nouveau.
- Si certaines questions n'apparaissent pas c'est simplement parce qu'elles sont redondantes avec d'autres.
- Autant que possible, l'ordre d'apparition des questions/réponses correspond au déroulé du cours.

2 Questions/Réponses

1. *Page 20* : Dans le polycopié, vous avez exprimé les paramètres intrinsèques et extrinsèques à partir de la matrice de projection perspective. On compte 10 paramètres indépendants (6 extrinsèques et 4 intrinsèques), qui apparaissent bien dans le système d'équation (1.17). Mais dans la partie 1.7 Calibrage d'une caméra, vous mentionnez 11 paramètres (sans compter le facteur multiplicatif près). Je ne comprends pas d'où provient le 11^{ème} paramètre. J'ai essayé de poser le système avec les calculs, je me suis demandé si ce 11^{ème} paramètre provient du terme m_{24} , qui est absent dans le système d'équation (1.17). Je n'ai pas l'impression qu'il y ait un lien, étant donné que $m_{24} = t_y * k_v * f + t_z * v_0$.

Réponse : Lorsqu'on considère la transformation avec les deux matrices distinctes, on peut effectivement compter 10 paramètres. En revanche, lorsqu'on considère la matrice de projection M , qui contient 12 paramètres, à un facteur multiplicatif près, nous identifions bien 11 paramètres. Ce n'est pas contradictoire puisque nous avons des formulations différentes.

2. *Page 32* : Pour la répétabilité, le cardinal est-t-il identique entre $\mathcal{P}_{D,1} \rightarrow \mathcal{P}_{D,2}$ et $\mathcal{P}_{D,2} \rightarrow \mathcal{P}_{D,1}$?

Réponse : Si nous n'imposons pas le nombre de points d'intérêt, alors le cardinal peut varier.

3. *Page 35* : Dans la détection de points d'intérêt multi-échelle (page 35), quelles sont les principales différences entre les détecteurs Harris-Laplace et Hessian-Laplace, et dans quels contextes l'un est préféré par rapport à l'autre ?

Réponse : Harris-Laplace est de premier ordre (on détecte plutôt des coins) alors que Hessian-Laplace est du second ordre (on détecte plutôt des fortes courbures). On préfère donc le premier lorsqu'on a une scène qui comporte des objets avec des angles droits alors que le seconde sera plus adapté avec des objets dont les extrémités sont plus arrondies (comme un papillon).

4. *Page 38* : Est-il possible d'utiliser les méthodes de mise en correspondances présentées en temps réel? (le tableau ne nous dit pas si les différentes méthodes sont utilisables en temps réel ou pas)
5. *Question proche* : Est ce que le détecteur SIFT a des cas d'applications pour le temps réels? Étant donné son descripteur à 128 dimensions... Sinon quels sont les cas d'applications typiques pour ce détecteur et quels sont les détecteurs les plus utilisés pour le temps réels?

Réponse : Si on considère les publications originales, je pense qu'aucun des auteurs n'a indiqué que l'opérateur proposé est applicable en temps réel. En revanche, certains détecteurs sont coûteux comme SIFT, les détecteurs basés courbures en multi-échelle, alors que d'autres sont moins coûteux comme Harris, SURF ou FAST (à favoriser pour approcher le temps réel). Quel que soit le détecteur choisi, il est toujours possible d'optimiser les calculs.

6. *Page 41* : Par rapport à Moravec, n'est-il pas plus pertinent de garder comme point d'intérêt la différence la plus importante entre deux pixels et donc le max de $E(x, y)$?

Réponse : Si on garde le maximum, cela signifie qu'on détecte autant les contours que les coins. En effet, si on garde le maximum, s'il existe une seule direction où la différence est élevée, on détectera un point d'intérêt. Dans l'approche initiale, on souhaitait écarter tous les points de contours et donc, il est judicieux dans ce cas de prendre le minimum.

7. *Page 63-64* : Comment est calculé le coût de lissage, que ce soit dans le coût des contraintes ou dans le coût global, et à quoi correspond la fonction de différence des niveaux de gris dans l'égalité (5.8)?

Réponse : Ce que nous appelons un coût de lissage, c'est la formalisation des contraintes expliquées dans le chapitre 7. Les différences, c'est ce qui est explicité dans le chapitre 6.

8. *Page 64-66* : Comment pourrait-on adapter les méthodes de mise en correspondance pour des images prises avec des caméras de différentes résolutions?

Réponse : On pourrait introduire un traitement multi-échelle, soit en utilisant une pyramide d'images, soit en effectuant des tests différentes tailles de fenêtre pour calculer le score de corrélation.

9. *Page 67* : Deux méthodes d'optimisation globale sont données, celles de l'affectation des plans, et celle du raffinement des paramètres. L'affectation des plans est plus efficace lorsque l'on veut corriger une mauvaise segmentation, mais quels sont les avantages du raffinement des paramètres? Dans quelles situations préfère-t-on l'un à l'autre?

Réponse : La première technique est dédiée au cas où l'on estime des plans, c'est-à-dire où le modèle de surface est un plan, alors que la seconde catégorie est plutôt dédiée à d'autres modèles de surface (comportant plus de paramètres) comme les B-Spline. Le choix de la technique d'optimisation est donc lié au modèle de surface utilisé. L'avantage d'une B-spline est de pouvoir modéliser des surfaces plus subtiles que de simples plans qui ne sont pas toujours une bonne approximation locale de la surface.

10. *Page 76* : Dans la section 5.3, les contraintes sont sous forme d'énergie alors que dans la section 7 les contraintes sont plutôt binaires (oui ou non, comme pour la contrainte de symétrie par exemple), comment incorporer ces contraintes dans l'équation 5.5?

Réponse : Cela signifie que dans l'équation proposée, le voisinage considéré est très basique, c'est un seul voisin.

11. *Question d'ouverture :* La mémorisation des points d'intérêt ainsi que leur mise en correspondance dans une séquence vidéo pourrait-elle être utilisée pour faire de la reconnaissance de lieux/position sur des séquences différentes ? Est-il possible d'utiliser les techniques vues dans ce cours ?

Réponse : Oui, les techniques vues en cours peuvent s'appliquer mais attention coût mémoire / temps de certaines techniques.

12. *Question d'ouverture :* Dans une vidéo comment la robustesse évolue-t-elle si un objet disparaît pendant quelques frames ? Cela garde-t-il en mémoire les caractéristiques de l'objet ou la mise en correspondance se fait image par image ?

Réponse : Si nous n'appliquons pas de contraintes, les points qui disparaissent vont continuer à être appariés mais ce sera faux. C'est la raison pour laquelle on peut utiliser des contraintes comme celles fournies dans le cours pour tenter d'identifier les points qui disparaissent, comme par exemple, en appliquant un seuil sur le score de corrélation. En effet, si le score de corrélation est faible, on peut en déduire que le point étudié n'a pas de correspondance.

13. *Question d'ouverture :* Dans un contexte de réalité augmentée ou de déplacement dans un environnement changeant (type véhicule autonome), comment les techniques de détection de mise en correspondance des points d'intérêts sont modifiées ou optimisées ?

Réponse : Il faut prendre en compte la dimension temporelle (images multiples) et exploiter le mouvement estimé précédemment dans le temps. Enfin, il faut considérer les approches multi-échelles, multi-résolution.

14. *Question d'ouverture :* Est-il pertinent d'utiliser des techniques d'estimation du flux optique pour en déduire un mouvement de caméra et réduire le champ de recherche pour l'appariement des points ? Cela supposerait un grand nombre d'images assez rapprochées (vidéo). Dans le cas d'une telle approche, quid de l'utilisation d'outils type accéléromètre pour la mesure du mouvement de caméra directement à la prise vidéo ? Ces techniques sont-elles pertinentes et/ou usitées ?

Réponse : Dans le contexte décrit et si les mouvements de la scène sont faibles, oui, c'est pertinent et c'est déjà fait dans la littérature. Et oui, c'est pertinent d'exploiter les informations données par un accéléromètre et ça se fait également.

15. *Question d'ouverture :* Dans le cadre de la mise en correspondance des points d'intérêt en stéréovision binoculaire, comment les contraintes d'unicité et d'ordre influencent-elles l'efficacité des algorithmes de mise en correspondance dans des scènes avec des objets se chevauchant ou des changements d'éclairage ? Existe-t-il des techniques pour adapter ces contraintes afin d'améliorer la robustesse dans ces conditions difficiles, tout en maintenant une précision élevée de la reconstruction 3D ?

Réponse : Ces contraintes sont justement là pour corriger les erreurs dues à des occultations et des changements d'éclairage. Elles n'ont donc pas besoin d'être adaptées.

16. *Question d'ouverture :* Vous avez comparé les différents types de contraintes (unicité, symétrie...) mais comment savoir celle que nous devons appliquer absolument, celles obligatoires et celles qui permettent "juste" d'affiner les résultats ?

Réponse : Dans la littérature, appliquer la contrainte de symétrie + un seuillage sur le score de corrélation, c'est la base. Ensuite, il faut agir en fonction du type de scène étudiée.