Questions/réponses
Classe inversée
Cours Vision

11 octobre 2023

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 18 :* pouvez-vous nous réexpliquer pourquoi entre la la figure 1.2 et la 1.3/4 on passe de 3 coordonnées à 4 ? De ce que nous avions compris, les coordonnées homogènes d'un point c'était passer de (x, y, z) à $(x, y, 1)$ mais dans le poly ils passent de (x, y, z) à $(x, y, z, 1)$ donc nous ne comprenons pas.

Réponse : En coordonnées homogènes, en 2D, on obtient $(x, y, 1)$, en 3D, on obtient $(x, y, z, 1)$.

2. *Page 23, section 2.1 :* Est-ce que le choix de la position relative des deux points de vue de la stéréovision binoculaire (si les deux sont éloignés ou rapprochés) affecte-t-il le processus de reconstruction 3D et la précision du calcul de la profondeur ? Est-il préférable de les rapprocher ou de les éloigner, ou il n'y a pas beaucoup de différence ?

Réponse : Oui, plus les points de vue sont éloignés et plus il y a d'écart entre l'apparence des éléments en communs. Plus l'apparence diffère et plus l'étape de mise en correspondance est difficile. De plus, plus les points de vue sont éloignés et moins il y a de recouvrement entre les deux images. Il y a donc une plus grande difficulté à trouver les correspondances de points.

3. *Page 25 :* Pouvez-vous expliquer comment la rectification épipolaire peut être réalisée dans des scénarios où les paramètres de calibrage sont partiellement connus ou incertains ? Y a-t-il des techniques de rectification adaptées pour traiter des cas où les informations de calibrage ne sont que partiellement disponibles, et comment peuvent-elles être mises en œuvre pour améliorer la précision de la rectification stéréoscopique dans de telles situations ?"

Réponse : Pour réaliser la rectification épipolaire, sans calibrage, on s'appuie sur la matrice fondamentale. C'est en dehors de ce cours de développer ce sujet mais je vous propose un article sur moodle pour compléter ma réponse.

4. *Page 31 :* Dans la figure 3.3, en quoi consiste le calcul de la réponse afin d'obtenir l'image (b) ? Et qu'est-ce qu'une réponse "élevée" ?
5. Comment choisir le seuil pour la détection des points d'intérêt, et quelles sont les considérations à prendre en compte dans ce processus ?

Réponse : Il s'agit du coeur des différentes techniques exposées dans ce cours : c'est le calcul de la probabilité que le point correspondent à un point d'intérêt. Une réponse élevée correspond à une probabilité élevée que le point soit caractéristique. Pour le choix du seuil, ce sera souvent réalisé en fonction du nombre de points attendus et de la répartition.

6. *Page 31 :* Quel est l'intérêt de définir la répétabilité d'une primitive pour au final étudier le taux de répétabilité entre 2 images ?

Réponse : Si on cherche à mettre en correspondance des points d'une image à une autre mais que le détecteur de points d'intérêt ne permet pas de détecter plusieurs fois le même point dans différentes images, on n'y arrivera pas. Pour résumer le résultat en une seule valeur, on s'appuie sur le taux de répétabilité. Il faut aussi retenir qu'on ne peut pas faire ce travail d'évaluation, si nous n'avons pas la vérité terrain, c'est-à-dire la connaissance des correspondances exactes.

7. *Page 33 :* Outre le fait que les approches de détection du premier ordre utilisent les dérivées premières de l'image et que les approches du second ordre se basent sur les dérivées secondes, quelles sont les différences entre deux approches, notamment en terme d'utilisation ?

Réponse : On ne s'intéresse pas aux mêmes éléments caractéristiques dans l'image. Dans le cas des approches du premier ordre, on cible plutôt les coins et les contours alors que pour les approches du second ordre, on utilise plutôt la notion de courbure et les brusques changements de courbure.

8. *Page 43, section 4.5.2 :* Comment la méthode Harris-Laplace détermine-t-elle les échelles pour le calcul de la réponse de Harris, et comment ces échelles sont utilisées pour estimer les extrema locaux lors de la première étape du processus ?
9. Pour l'algorithme Harris-Laplace : comment l'ajustement du facteur d'échelle c peut influencer la détection d'un point d'intérêt ? Y'a t'il des valeurs de c qui correspondent à des applications spécifiques ?

Réponse : Les extremas locaux sont sélectionnés comme avec l'approche SIFT. Dans un voisinage donné, en échelle et en espace. Ensuite, il faut effectivement choisir c le facteur d'échelle et c'est plutôt empirique. Comme pour SIFT où on doit choisir le nombre d'échelles et d'octaves. Je n'ai pas de valeurs spécifiques pour des applications spécifiques à proposer.

10. *Page 47, section 4.4.3 :* Comment la sensibilité du détecteur MSER aux images floues peut avoir impact sur sa performance dans des tâches pratiques de traitement d'images ? Existe-t-il des moyens d'atténuer cette sensibilité ?

Réponse : Si une image est flou, les frontières entres les objets sont moins nettes est un simple seuillage aura des difficultés à séparer correctement les régions, mêmes en faisant varier le seuil.

11. *Page 49, section 4.4.7 :* Comment le détecteur MSER, qui se base sur les régions maximales extrema et stables, diffère-t-il fondamentalement du détecteur PCBR, qui lui aussi utilise la notion des lignes de partage des eaux pour la détection de régions d'intérêt ?

Réponse : MSER s'appuie sur les régions détectées grâce à une détection de contours basées ligne de partage des eaux alors que le détecteur PBCR, on calcule la courbure des contours (qui sont effectectivement détéctées avec des contours.

12. *Page 54 :* Nous avons du mal à comprendre le principe de SURF, notamment le schéma 4.12.

Réponse : La figure ne correspond pas à expliquer SURF mais à expliquer le principe des images intégrales. Cette figure explique la formule (4.33).

13. *Page 61, dans l'équation (5.2) :* Pour la minimisation de l'Energie, comment choisir λ ? Pourquoi privilégier un coût à un autre ? On ne prend pas 1/2 par défaut ? On privilégie une valeur de lambda plutôt qu'une autre en fonction du cas d'étude ?

Réponse : C'est effectivement un hyperparamètre difficile à choisr. Une valeur classique est effectivement $\frac{1}{2}$. Une autre façon classique d'aborder le problème est de faire varier λ pendant le processus d'optimisation. Pour le choix du coût en général, c'est toujours en fonction du type de scènes étudiées. Est-ce que les images sont plus ou moins texturées, contrastées. Est-ce qu'il y a globalement un respect de l'ordre ? Est-ce qu'il y a des déformations.

14. *page 64, partie 5.4.1 :* Nous ne comprenons pas très bien quelle est la différence entre une mise en correspondance dense et une mise en correspondance éparse.

Réponse : On parle de mise en correspondance dense lorsque tous les pixels de l'image sont appariés alors qu'une correspondance éparse correspond à l'appariement uniquement d'un sous-ensemble de pixels, comme par exemple, des points d'intérêt.

15. *page 66, partie 4.3 :* le principe des méthodes des régions n'est pas très clair pour nous. Dans quels cas pouvons nous appliquer cette méthode ? Il est indiqué que la zone de couleur homogène doit être suffisamment petite (selon l'hypothèse de départ) afin qu'elle puisse être projetée et approchée par un modèle dans l'espace des disparités, mais nous ne comprenons pas ce que cela veut dire.
16. Pourquoi B-spline est il plus complexe à configurer que le modèle plan ? et pourquoi peut il entraîner des oscillations ?

Réponse : Il est difficile de donner une règle générale des cas d'application mais voici des éléments de réponse : nous pouvons appliquer ce type de solution lorsque l'image peut être sur-segmentée sans générer trop d'erreurs, c'est-à-dire lorsque nous avons des contours bien contrastés et des régions avec des textures ou couleurs très différentes lorsqu'elles correspondent à des objets très différents. Je vais fournir un exemple en cours (que vous pourrez retrouver dans les transparents partagés) pour illustrer ces explications. En ce qui concerne la seconde partie de la question : dans ce type d'approches, on suppose que chaque région repecte un modèle de surface, comme par exemple, un plan. Afin d'être au plus prêt de cette hypothèse, il est important de travailler plutôt avec des régions petites, car localement, une surface peut être approximée par un plan. On peut également utiliser un modèle de surface plus complexe, les B-spline. Il y a plus de paramètres à estimer pour ajuster une B-spline. C'est la raison pour laquelle c'est plus complexe. Vous avez également vu en cours de modélisation géométrique l'année dernière que si on choisit (ici estime) mal les paramètres, la surface peut osciller au niveau des points de contrôle.

17. *page 72, partie 6.7.1 sur la mesure ISC :* Au vu de la définition de la fenêtre de corrélation $\mathbf{f}_l$, nous bloquons sur le parcours de cette dernière par l'indice k : la valeur $\mathbf{f}_l^0$ correspond au coin supérieur gauche d'après le tableau des notations, et donc ISC compare progressivement deux valeurs côte à côte ? Est-ce que la valeur de droite est plus grand que celle de gauche ? Comment la mesure évolue lors par exemple d'un passage à la ligne du dessous (les deux pixels étant dans ce cas quand même espacé) ? Potentiellement nous avons mal compris comment cette mesure parcours la fenêtre de corrélation et voulons donc plus d'explication à ce sujet, s'il vous plaît.

Réponse : Cette mesure va comparer l'intensité des pixels voisins deux à deux. Si le pixel courant est d'intensité plus faible que le suivant, alors la transformation appliquée vaut 0, sinon, elle vaut 1. Je reprend cette explication en cours.

18. *page 73, section 6.8 :* Que signifie être robuste aux occultations ? Ne pas utiliser ces données en les jugeant aberrantes n'est pas un problème ? Dans le cas de la reconstruction 3D, comment gère-t-on le fait qu'on voit l'objet sous des angles très différents et donc que les points d'intérêt ne se correspondent pas forcément entre chaque image ? Cela revient-il a un problème d'occultation ?

Réponse : Être robuste veut dire que le calcul n'est pas affecté par le fait que certaines parties sont occultées dans l'image. Ce ne sont pas les points occultées qui ne sont pas considérés, ce sont les parties qui diffèrent entre deux voisinages de points homologues qui ne sont pas utilisées. Cela peut être problématique de faire cela s'il n'y a pas d'occultations car nous pouvons éliminer des éléments significatifs pour réaliser la mise en correspondance.

19. *page 78, section 7.7 :* Pouvez-vous expliquer comment la contrainte d'unicité peut être affectée dans le cas de scènes où un plan est fortement incliné par rapport à l'une des caméras, et pourquoi cela peut remettre en question cette contrainte ?

Réponse : Lorsque la surface est fortement inclinée par rapport au plan image, alors de nombreux points 3D vont se projeter au même point de l'image. Je reprend cette notion avec un exemple en cours.

20. *page 82, chapitre 7, figure 7.4 :* plusieurs contraintes sont présentées voire comparées (figure 7.4). Est-il judicieux d'appliquer plusieurs d'entre elles à la fois où mieux vaut-il se cantonner à un seul lors de l'application ? Est-ce que trop de contraintes risque-t-il d'empêcher la mise en correspondances ou les contraintes sont trop légères pour que cela n'arrive ?
21. Vaut il mieux avoir peu de point d'intérêt mais qui sont précis, ou moyennement mais avec des incertitudes ?

Réponses : Oui, on peut combiner l'utilisation des contraintes comme contrainte de symétrie et contrainte d'ordre. Oui, plus on utilise de contraintes, plus le nombre de correspondances trouvées sera faible (mais plus il sera fiable), comme je l'ai montré pour SUSAN. Tout dépend de l'utilisation qu'on en fait ensuite :

- Si nous avons besoin de couvrir le maximum de zones dans l'image, on accepte les erreurs et on n'applique pas ou peu de contraintes.
- Si nous avons besoin de correspondances fiables, alors, on applique le plus de contraintes possibles.

Cette réponse s'applique également au cas des points d'intérêt : tout dépend de ce que l'on en fait après.

22. *page 85, chapitre 8 :* Quel est l'intérêt du calcul de la pyramide d'images dans la méthode SIFT ? Sert-il à extraire les features ?

23. Dans un détecteur multi-échelle (comme SIFT), une fois des points d'intérêt détectés à différentes échelles, comment sont réalisées les correspondances entre les différentes échelles ?

Réponse : Les pyramides d'images servent à extraire des points d'intérêt en utilisant différents niveaux d'analyse. Cela doit permettre d'extraire des points sur des objets ou des éléments de la scène avec des tailles variées. Pour chaque point détecté, on calcule 128 valeurs différentes qui correspondent à la description ou la signature du point. On effectue ensuite une recherche parmi tous les points d'intérêt détectés dans l'autre image en choisissant le point dont le descripteur de 128 aura la plus grande corrélation avec le descripteur du point étudié. Ils s'agit d'une mise en correspondance par corrélation croisée centrée.

24. *page 86, section 8.1.2 :* Au niveau du calcul du descripteur, comment la deuxième normalisation du vecteur-descripteur contribue-t-elle à assurer l'invariance aux changements affines d'illumination, en complément de la première normalisation ?

Réponse : La normalisation permet de recadrer les valeurs entre 0 et 1. Si une transformation affine est appliquée entre les deux images, sans normalisation, les calculs, pour des points qui se correspondent, seront différents. La normalisation permet de pallier cette difficulté.

25. *Question générale :* Est-il possible de déterminer à l'avance pour une image donnée, le type de méthode, contraintes, algorithme qui donnera les meilleurs résultats (aussi bien pour la détection que pour l'appariement), et si oui quels sont les éléments à prendre en compte ?

Réponse : Les choix sont faits en fonction des éléments manipulés et des résultats visés. On ne peut pas garantir à l'avance ce qui donnera les meilleurs résultats mais on peut a minima éliminer toute méthode qui ne respecte pas les hypothèses initiales. Par exemple, s'il y a des changements d'échelle dans les images traitées, je vais favoriser une approche robuste aux changements d'échelles. Un autre exemple : si les contours sont mal contrastés, je ne vais pas utiliser une approche qui s'appuie sur la détection de contours. Tous les critères pour faire les choix, en termes de suivi, sont donnés à la page 13.

26. *Question générale :* Mise à part un filtrage gaussien, existe-t-il d'autres filtrages afin de faire varier l'échelle d'une image ?

Réponse : Pour faire varier la taille/résolution de l'image, dans le but de construire une pyramide d'images, on peut utiliser un simple filtre moyenneur. Lorsqu'on parle de faire varier l'échelle d'analyse, on parle de faire varier la taille du filtre quelqu'il soit. Dans le cours, je vous montre plusieurs type de filtre qui peuvent être utilisés, il n'y a pas que des filtres gaussiens.

27. *Question générale :* Je n'ai pas compris quand il faut utiliser telle ou telle famille de mesure de similarité.

Réponse : Encore une fois, c'est le contenu de l'image qui va guider le choix : présence d'occultations, variation d'intensité, faible texture.

28. *Question générale :* Pouvez-vous, s'il vous plait, donner des exemples de transformations d'images qui ne devraient pas affecter la détection de points d'intérêt ?

Réponse : On aimerait que les transformations qui peuvent arriver n'affectent pas la détection. Ces transformations sont : les translations (lorsqu'on déplace le capteur ou que l'on a plusieurs capteurs), les rotations (même chose), les changements d'échelle (même et également lorsqu'on utilise le zoom du capteur), les changements d'illumination (qui arrivent lorsque le temps entre les différents acquisition est significatif mais également car l'éclairage diffère d'un point de vue à l'autre), les

déformations (certains objets se déforment comme par exemple un ballon). Tous les détecteurs n'ont pas été proposés pour considérer ces aspects, c'est pourquoi je donne un tableau qui récapitule toutes les propriétés suivant les détecteurs, cf. page 38.

29. *Question d'ouverture :* Comment les détecteurs de points d'intérêt et les descripteurs sont-ils adaptés pour la détection et le suivi de points d'intérêt dans des vidéos ?

Réponse : Tout ce que nous avons vu peut être utilisé dans le cadre des vidéos. Mais il va falloir se poser des questions sur les temps de calculs et les coûts en mémoire. Dans le cas de la vidéo, il faudra appliquer la sélection de vues clés, utiliser des astuces algorithmiques pour réduire les calculs.

30. *Question d'ouverture :* Dans quelle mesure les CNN sont-ils adaptés à la génération de descripteurs d'images par rapport aux méthodes vues en cours telles que SIFT ou SURF, et quelles sont les applications spécifiques où les CNN surpassent ces méthodes ?

Réponse : Toute approche par CNN est intéressante lorsque le problème est mal défini/difficile à expliciter.

31. *Question d'ouverture :* Vous évoquiez au dernier cours l'apprentissage profond, comment les détecteurs HOG et SIFT par exemple peuvent ils être compatibles avec des méthodes de construction de features apprises à partir des données ? Et comment pourraient-ils être intégré dans un pipeline de détection d'object basé sur du deep learning ?

Réponse : C'est une question de recherche ouverte. De mon point de vue, le but est identique mais la méthode diffère. Cependant, les détecteurs HoG et SIFT s'appuient sur un filtrage de l'image pour extraire des gradients et des orientations de gradients. Par définition, un CNN applique des convolutions en cascade. Là où HoG et SIFT ont de paramètres figés, un réseau de neurones apprend les paramètres idéaux pour extraire les caractéristiques les plus significatives. Je pense qu'il faut toujours regarder ce que nous apporte un détecteur sans apprentissage avant d'utiliser une approche par apprentissage. Si ce détecteur apporte une solution exploitable, je ne vois pas l'intérêt d'utiliser un CNN qui va demander la construction d'une base de données avec vérité terrain et un coût considérable d'entraînement ou de transfert de domaine.

32. *Question d'ouverture :* Est-ce que ces méthodes arrivent à détecter des points d'intérêt sur ce qui est camouflé ? (Ex : uniforme militaire ou léopard des neiges dans la neige)

Réponse : On peut imaginer qu'on arrive à détecter les motifs (de l'uniforme militaire ou de l'environnement) et les contours internes (ceux du léopard) mais effectivement, si on cherche des points de contours plutôt à la frontière des objets, cela risque de ne pas être concluant.

33. *Question d'ouverture :* Pouvez-vous, s'il vous plait, donner des exemples de transformations d'images qui ne devraient pas affecter la détection de points d'intérêt ?

Réponse : On aimerait que les transformations qui peuvent arriver n'affectent pas la détection. Ces transformations sont : les translations, les rotations, les changements d'échelle, les changements d'illumination, les déformations. Tous les détecteurs n'ont pas été proposés pour considérer ces aspects, c'est pourquoi je donne un tableau qui récapitule toutes les propriétés suivant les détecteurs, cf. page 38.