

# UE Vision, Réalité Augmentée et Applications

Examen - Documents autorisés : photocopié de cours distribué en vision - Durée 1h

2 décembre 2021

## 1 Vision (5 points)

Nous souhaitons développer un algorithme de reconnaissance automatique d'éléments connus dans une scène de jeu. Plus précisément, dans la figure 1, nous souhaitons, à partir des imagerettes de fruits et légumes, retrouver tous les éléments qui leur correspondent dans l'image globale.



FIGURE 1 – La première image correspond à une scène de jeu où l'on cherche à reconnaître des éléments : les quatre fruits et légumes représentés en dessous.

1. Indiquez quelles sont les difficultés que vous identifiez pour réaliser cette reconnaissance. Appuyez votre réponse avec des éléments précis.



## 2 Réalité augmentée

1. La fonctionnalité Live View de l'application Google Maps pour les dispositifs mobiles est un système de navigation qui affiche les directions à suivre pour arriver à destination en surimpression sur le flux vidéo de l'appareil photo (voir Figure 2).

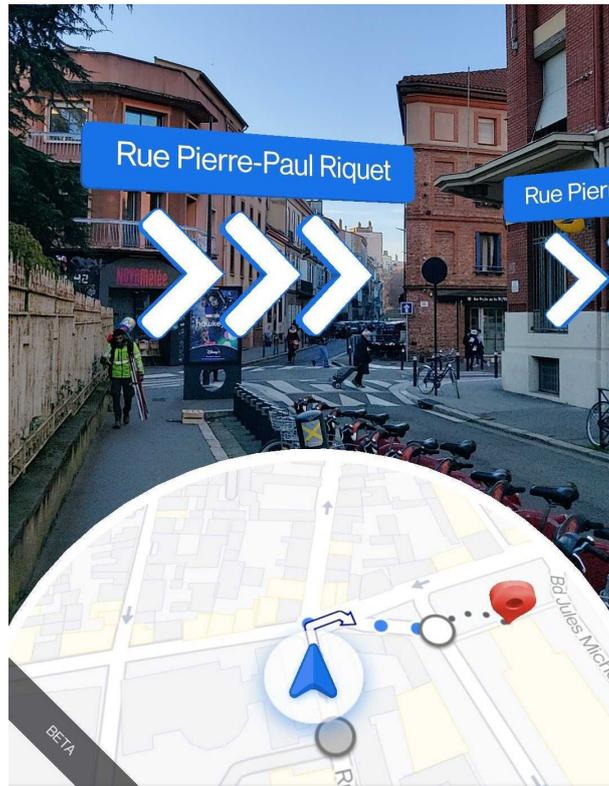


FIGURE 2 – La fonctionnalité Live View de l'application Google Maps.

- (a) Est-ce qu'on peut classer cette application comme une application de réalité augmentée au sens de la définition d'Azuma ? **Argumenter la réponse** (si nécessaire, faire des hypothèses sur les fonctionnalités de l'application).
  
  
  
  
  
  
  
  
  
  
- (b) Quel(s) type(s) de « tracker » (inertiel, optique, *etc.*) est (sont) plus adapté(s) pour réaliser cette application ? **Justifier la réponse.**

2. Lors des séances de TP vous avez réalisé deux versions d'un "camera tracker". À partir de cette expérience, illustrez brièvement, **pour chaque version**,
- le type d'algorithme de tracking réalisé ;
  - les algorithmes utilisés pour la détection et le suivi des points d'intérêt ;
  - les algorithmes utilisés pour l'estimation de la pose de la caméra ;
  - en général, les points forts et les points faibles de chaque approche.

3. On dispose d'une caméra panoramique qui peut tourner autour d'un axe passant par le centre optique de la caméra (rotation pure). On peut choisir une position de référence de la caméra pour laquelle un point 3D  $\hat{\mathbf{Q}} = \begin{bmatrix} \mathbf{Q} \\ 1 \end{bmatrix}$  est projeté sur le plan image dans un point  $\mathbf{q}_1$  selon l'équation

$$\mathbf{q}_1 \sim \mathbf{K} \begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix} \hat{\mathbf{Q}} = \mathbf{K} \mathbf{Q}. \quad (1)$$

Imaginons que la caméra tourne selon une rotation  $\mathbf{R}$  autour du centre optique et que le point  $\hat{\mathbf{Q}}$  reste visible par la caméra.

- (a) Écrire l'équation de projection de  $\hat{\mathbf{Q}}$  dans le point  $\mathbf{q}_2$  sur le plan image de la caméra après la rotation  $\mathbf{R}$ .

- (b) En utilisant cet équation et à l'aide de l'eq. (1) trouver la relation entre les deux points  $\mathbf{q}_1$  et  $\mathbf{q}_2$ .

- (c) Qu'est-ce qu'on peut dire de cette relation ? (hint : pensez à les dimensions des matrices/vecteurs...) Dans quel contexte vous l'avez déjà utilisé ?

- (d) Est-ce que cette relation est valable pour tous les autres points des 2 images ? Si oui, combien des points correspondants comme  $\mathbf{q}_1 \leftrightarrow \mathbf{q}_2$  sont nécessaires pour estimer les éléments de cette relation ?

- (e) Si on suppose que la caméra est calibrée (matrice  $K$  connue), est-ce qu'on peut estimer la rotation  $R$  ? Comment ?

4. Expliquez, en général, l'intérêt d'avoir des méthodes "Minimal solver" pour estimer les paramètres d'un modèle à partir d'un sous-ensemble minimal des données.

Par exemple, la pose d'une caméra peut être facilement estimée en résolvant un système linéaire aux moindres carrés (DLT) avec  $N \geq 6$  associations 2d-3d des points, mais l'algorithme P3P permet de résoudre le problème avec 3 associations seulement (avec un nombre fini de solutions). Au-delà du nombre des données, pourquoi on privilège ce type des méthodes même quand on a suffisamment de données ? Quel est l'intérêt ?