M1 Informatique

Épreuve écrite de l'UE optionnelle

Traitement d'image et vision par ordinateur (TIVO)

Partie Vision par ordinateur (Gilles Simon)

Mémento de 1 page au format A4 autorisé.

Durée : 1h

Il est préférable de traiter les questions dans l'ordre. Le début du sujet est inspiré d'un article scientifique publié en 2011 par des chercheurs japonais [1]. Les images infrarouge obtenues par une Kinect et les illustrations utilisées en seconde partie proviennent respectivement de [2] et de [3].

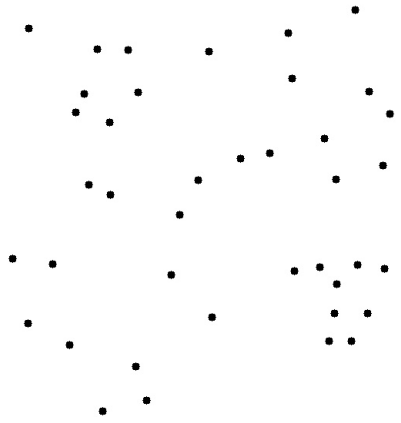
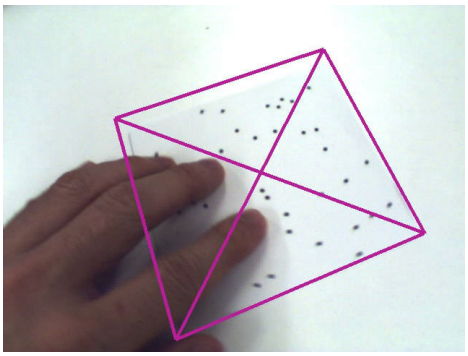
(a)  (b) 

FIGURE 1 – (a) : Exemple de RDM. (b) : Interaction avec un RDM.

1. Un *random dot marker* (RDM) est un marqueur artificiel composé de points (des petits disques noirs) disposés aléatoirement sur un fond blanc (fig.1(a)). On peut utiliser ce type de marqueur pour manipuler *physiquement* des objets virtuels en réalité augmentée (fig.1(b)). L'objectif est alors de calculer en temps réel la matrice de projection P permettant de dessiner dans l'image vidéo des objets virtuels décrits dans le repère du marqueur (par exemple, le carré avec ses diagonales visible en figure 1(b)). Rappeler comment utiliser P pour calculer les coordonnées image (u, v) d'un point 3D de coordonnées (x, y, z) exprimées dans le repère du marqueur.
2. Supposons que les paramètres intrinsèques de la caméra ne sont pas connus. Combien de paires de points $(u_i, v_i) \leftrightarrow (x_i, y_i, z_i)$ faut-il connaître pour calculer P (justifiez brièvement votre réponse) ? La difficulté est donc finalement d'apparier suffisamment de points $(u_i, v_i) \leftrightarrow (x_i, y_i, z_i)$ en temps réel, les coordonnées (u_i, v_i) des points de l'image vidéo changeant à chaque instant.
3. Proposez un algorithme pour détecter automatiquement les centres (u_i, v_i) des ronds noirs dans une image telle que l'image visible en figure 1(b). Donnez uniquement les grandes lignes de cet algorithme en 2 ou 3 lignes maximum (utilisez le fait que les ronds sont noirs, et le fond blanc). Pour le moment, ne vous préoccupez pas du problème des *faux positifs* (points "parasites" détectés par l'algorithme à des endroits de l'image où il n'y a pas de *random dots*).
4. On appelle *invariant projectif* une mesure liée à la position relative *dans une image* d'un groupe de points, et dont la valeur ne dépend pas du point de vue d'observation de ces points. Par exemple, si A,B,C,D,E sont cinq points coplanaires, le critère suivant est un invariant projectif :

$$cross_ratio(A, B, C, D, E) = \frac{S(A, B, C)S(A, D, E)}{S(A, B, D)S(A, C, E)},$$

où $S(A, B, C)$ désigne l'aire du triangle de sommets A,B,C mesurée dans l'image. En utilisant ce critère, proposez un algorithme pour apparier automatiquement les centres des points (u_i, v_i) obtenus grâce à l'algorithme décrit en question 3 avec les centres des points de référence $(x_i, y_i, 0)$ générés aléatoirement lors de la définition du marqueur (c'est-à-dire, par exemple, entre les points de la figure 1(b) et les points de la figure 1(a), considérant qu'il s'agit du même marqueur). Ne donner, encore une fois, que les grandes lignes de l'algorithme. Attention : il faut tenir compte du fait que le marqueur observé dans l'image vidéo peut n'apparaître que partiellement dans le champ de vision de la caméra ou être occulté (par exemple, par la main de l'utilisateur comme en figure 1(b)) et que des faux positifs peuvent être détectés dans l'image vidéo (cf. question 3). Autrement dit, tous les points de référence n'ont pas nécessairement un correspondant dans l'image vidéo et tous les points détectés dans l'image vidéo n'ont pas nécessairement un correspondant dans l'image de référence.

5. La Kinect de Microsoft (fig. 2) dispose d'un émetteur infrarouge (IR) permettant de projeter des *random dots* (RD) invisibles à l'œil nu sur toutes les surfaces de l'environnement rencontrées par la lumière IR. L'image des RD est obtenue à l'aide d'une caméra IR. La figure 3 montre deux images obtenue par une Kinect, et leur différence, avant et après avoir ajouté un livre dans la scène. Comment expliquer ces images ?

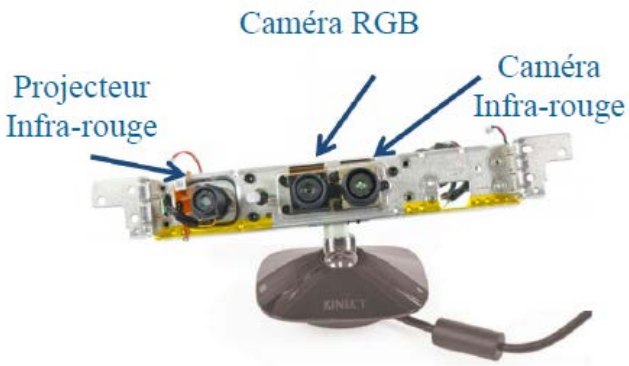


FIGURE 2 – Intérieur d'une Kinect de Microsoft.

6. (FACULTATIF) Sur le schéma de la figure 4, le point x correspond au centre d'un RD mesuré dans l'image émise par le projecteur IR, et le point x' au centre du même RD détecté dans l'image reçue par le capteur IR. En utilisant le théorème de Thalès deux fois (par exemple), montrer que la profondeur z du point X de la scène rencontré par le rayon IR passant par x est donné par la formule :

$$z = \frac{Bf}{d},$$

où $B = C - C'$ est la *baseline* du système projecteur/capteur et $d = x - x'$ est la *disparité* du point X entre l'image projetée et l'image capturée.

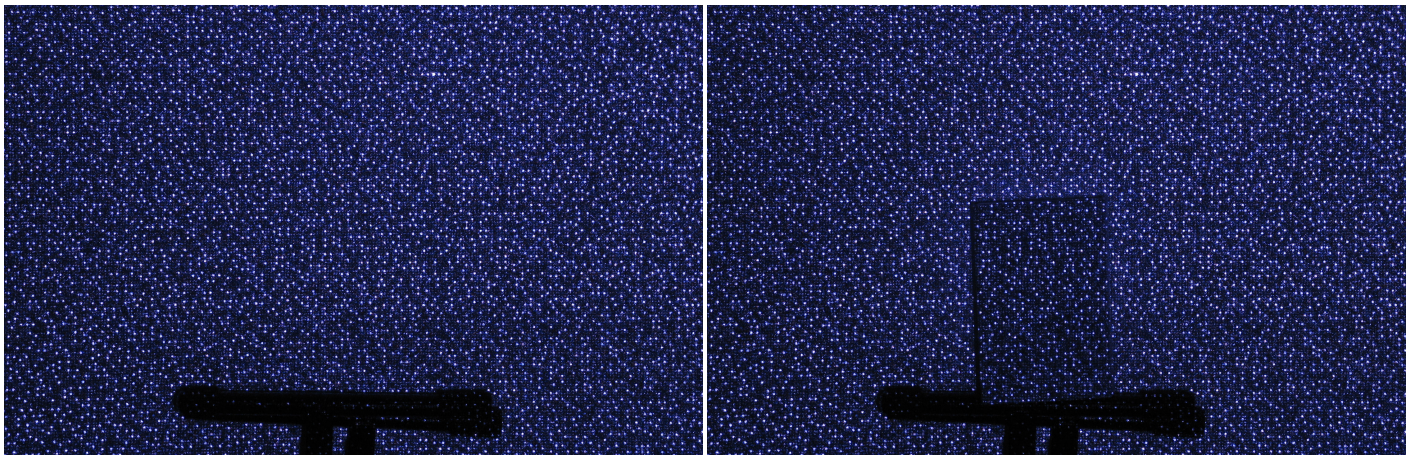
7. Pour apparier les RD entre l'image projetée et l'image capturée, on peut envisager d'utiliser l'algorithme que vous avez proposé pour répondre à la question 4, mais est-ce bien raisonnable ? Quelle contrainte pouvons-nous utiliser pour être plus efficace ? Est-il vraiment nécessaire ici d'utiliser un invariant projectif (justifiez votre réponse) ?

Sources

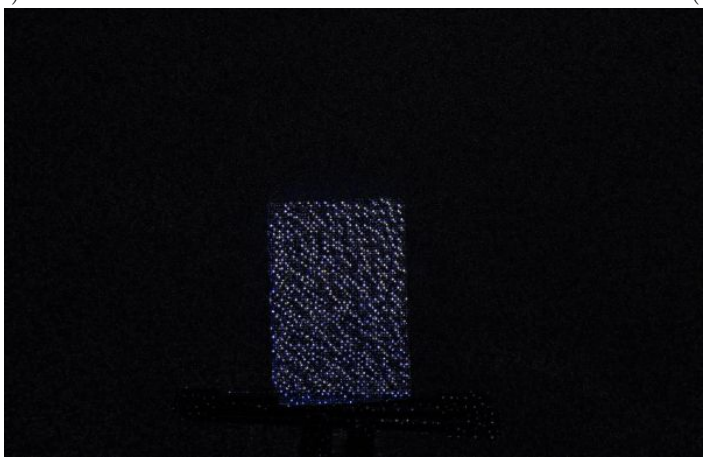
[1] H. Uchiyama & H. Saito, "Random dot markers", IEEE Virtual Reality Conference, pp.35-38, 19-23 March 2011.

[2] `http://www.futurepicture.org/`

[3] `https://courses.engr.illinois.edu/cs498dh/fa2011/lectures/Lecture 25 - How the Kinect Works - CP Fall 2011.pdf`



(a) (b)



(c)

FIGURE 3 – (a),(b) : Images obtenues à l'aide d'une Kinect avant et (resp.) après introduction d'un livre dans la scène. (c) : Différence entre les images (a) et (b).

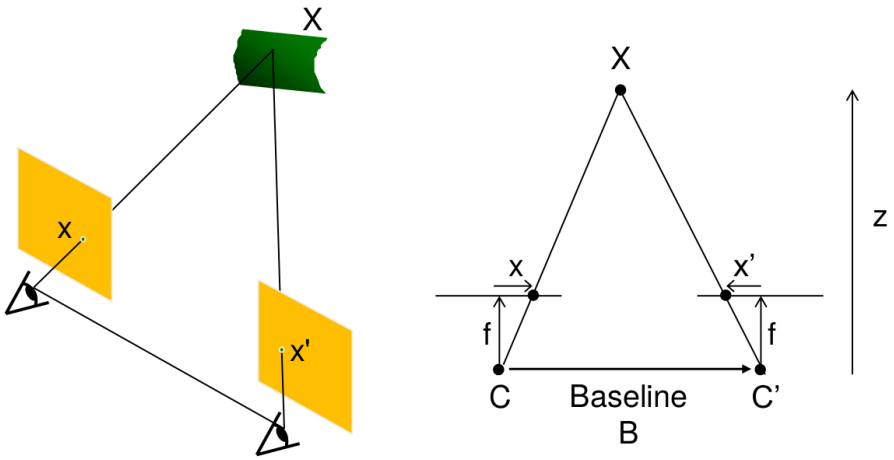


FIGURE 4 – Calcul de la profondeur d'un point en fonction de sa disparité.