



HAL
open science

Extraction et comparaison d'information saillante : Pose favorable et image 2D révélatrice d'un objet 3D

Marie Pelissier-Combescure, Géraldine Morin, Sylvie Chambon

► To cite this version:

Marie Pelissier-Combescure, Géraldine Morin, Sylvie Chambon. Extraction et comparaison d'information saillante : Pose favorable et image 2D révélatrice d'un objet 3D. 18èmes journées franco-phones des jeunes chercheurs en vision par ordinateur (ORASIS 2021), Centre National de la Recherche Scientifique [CNRS]; Equipe REVA, IRIT : Institut de Recherche en Informatique de Toulouse., Sep 2021, Saint Ferréol, France. pp.1-9. hal-03339730v2

HAL Id: hal-03339730

<https://hal.science/hal-03339730v2>

Submitted on 9 May 2022

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.

Extraction et comparaison d'information saillante : Pose favorable et image 2D révélatrice d'un objet 3D

M. Pelissier-Combesure

G. Morin

S. Chambon

IRIT, Université de Toulouse, Toulouse, France

marie.pelissier-combesure@irit.fr

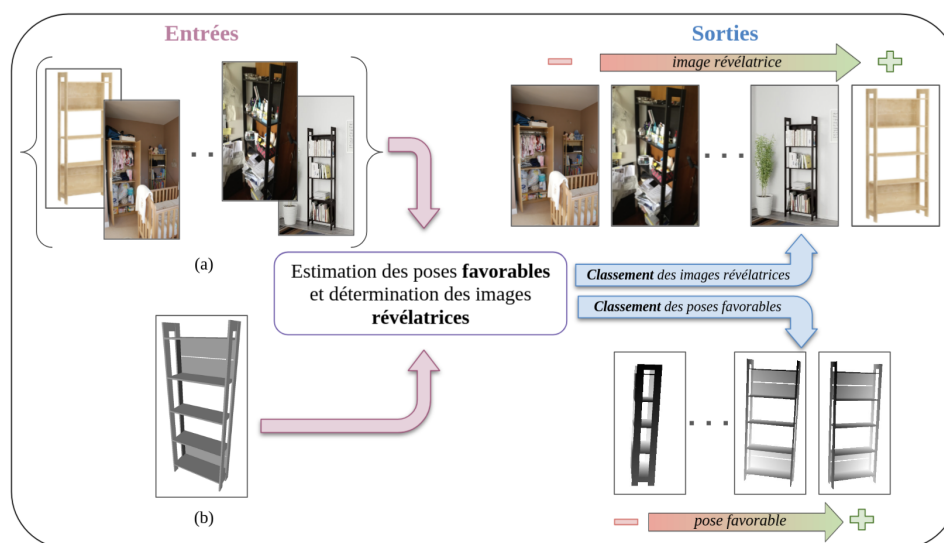


FIGURE 1 – Évaluation quantitative de poses et d'images 2D d'un objet 3D.

Résumé

L'objectif de ce papier est d'introduire une approche permettant de quantifier l'information d'un objet 3D présente dans une image 2D, en distinguant, d'une part, l'intérêt de sa pose et, d'autre part, l'intérêt de l'image 2D que nous observons. L'objectif est d'extraire, à partir de ces deux modalités différentes et complémentaires, l'information essentielle appartenant à l'objet dans l'image. Cette information s'obtient par extraction d'éléments saillants à partir de l'image 2D et du maillage 3D. Pour cette étape de détection, nous avons choisi d'utiliser un détecteur qui nous assure la meilleure répétabilité possible entre 2D et 3D. Ce détecteur s'appuie sur la saillance curviligne. Nous proposons alors de déterminer la pose mettant le plus en valeur l'objet 3D s'appuyant sur les points d'intérêt détectés. Nous introduisons également une méthode qui nous permet de déterminer si l'image 2D qui contient cet objet est révélatrice de cet objet. Plus précisément, connaissant un ensemble de poses et un ensemble d'images, nous sommes en capacité de proposer un classement de ces poses et de ces images en fonction de l'information saillante contenue. test bbb

Mots Clef

Modèles 3D, cartes de profondeur, images 2D, saillance curviligne, qualité d'une représentation, intérêt d'une pose.

Abstract

The objective of this paper is to quantify the information relative to a 3D object contained in a 2D image, by estimating, on the one hand, the interest of its pose and, on the other hand, the interest of the 2D observed representation. The goal is to extract, from these two different and complementary modalities, the essential information belonging to the object in the image. This information is computed by matching salient elements between the 2D image and the 3D mesh. For this detection step, we use a detector between 2D and 3D, based on the curvilinear saliency, that ensures the best possible repeatability for salient elements in different modalities. We are thus able to identify the pose that best highlights the 3D object based on the detected points of interest. Using a similar approach, we also quantify if the 2D image that contains this object highlights it. In particular, given a set of poses and images, we are able to

order these poses and images according to their saliency : the interest of the pose and the relevance of the 2D image.

Keywords

3D models, depth maps, 2D images, curvilinear saliency, essential information, quality of a representation, interest of a pose.

1 Introduction

Contexte. Aujourd’hui nous vivons dans une ère où obtenir des données massives et de natures différentes, que ce soit des images, des vidéos, des modèles 3D ou des cartes de profondeur, est assez aisé. En effet, il existe des bases de données avec des milliers d’images, comme la base Imagenet [3] ou avec des dizaines d’heures de vidéos, comme la base ToCaDa [10]. Certaines bases de données ont la particularité de combiner des données de natures différentes. À titre d’exemple, nous pouvons citer le jeu de données Pascal3D+ [18] qui contient un ensemble d’images avec de multiples annotations et des modèles 3D associés aux objets présents dans les images.

De nombreuses applications, allant de la médecine aux voitures autonomes, favorisent la manipulation de données de natures différentes car elles apportent des informations complémentaires. En effet, les modèles 3D sont intrinsèquement indépendants du point de vue et permettent d’extraire des informations géométriques alors que les images sont liées à un point de vue et apportent bien souvent une information complémentaire de texture. Dans le cas de la médecine, l’appariement de données de sources différentes permet une meilleure prise en charge de la pathologie par la détection plus précise des zones à soigner [11]. De plus, l’utilisation d’images 2D et de modèles 3D peut également servir à déterminer la localisation du lieu exact d’une photographie grâce à son appariement [19]. Enfin, dans le cas des voitures autonomes [2], il est possible d’acquérir des données LIDAR¹ ou RADAR², ainsi que des flux vidéo provenant de caméras embarquées.

Toutefois, ces volumes importants de données dynamiques ou non, temporelles ou non, contiennent des informations plus ou moins pertinentes et importantes. En particulier, une image contient de nombreux objets qui peuvent être au premier plan ou plutôt faire partie du fond. De plus, les objets peuvent être tronqués, occultés ou ne pas apporter une grande contribution à l’information contenue dans l’image. Cette information est ce que nous appelons une **information essentielle**. Par définition, l’information extraite est plus **compacte** que l’intégralité de la donnée initiale. En effet, l’opération d’extraction des éléments significatifs rend l’information plus condensée.

Problématique. L’objectif de cet article est d’identifier dans une image si un objet est bien illustré, c’est-à-dire, si nous retrouvons ses informations essentielles dans la

vue que nous observons. Plus précisément, dans un premier temps, nous nous sommes concentrées sur l’étude d’un couple de données formé d’une image 2D contenant un objet à étudier, cf. Figure 1(a), et d’un maillage 3D de cet objet, cf. Figure 1(b). **Nous connaissons la pose de la caméra et sa focale.** Nous sommes donc capables de placer les silhouettes 3D dans la même position que celles qu’elles occupent dans l’image en 2D.

Ainsi nous souhaitons déterminer si le point de vue fourni dans l’image 2D est avantageux pour l’objet, en quantifiant l’information essentielle de l’objet disponible dans la vue proposée. Ici, le terme **avantageux** décrit deux notions :

- La pose est la plus **favorable** pour l’objet 3D : celle qui caractérise le mieux l’objet, dans le sens où elle permet d’obtenir une grande majorité des informations essentielles de l’objet, c’est-à-dire les éléments caractéristiques de cet objet, cf. Figure 2(a) et 2(b).
- La vue de cet objet dans l’image proposée est satisfaisante, c’est-à-dire que l’objet au sein de l’image est correctement visible. En d’autres termes, l’image est **révélatrice** de l’objet, cf. Figure 2(c) et 2(d).

Dans la littérature, il existe des méthodes pour évaluer la qualité subjective d’une image basées sur la fidélité de la couleur acquise à la couleur perçue ou encore en évaluant la netteté d’une image [7]. Cette tâche est délicate car elle est subjective et donc difficile à quantifier et à évaluer. Dans ce travail, nous proposons d’aborder une problématique proche mais légèrement différente, dans le sens où nous souhaitons évaluer l’intérêt d’une représentation 2D par rapport à un objet 3D connu. De plus, nous apportons une solution différente en nous appuyant sur des outils classiques en vision : l’extraction de primitives d’intérêt.

La section suivante va nous permettre d’introduire brièvement les méthodes d’évaluation de la qualité d’une image avant de nous concentrer sur les travaux existants pour l’extraction de primitives saillantes dans des données 2D et 3D. Cette présentation nous permettra d’introduire notre proposition en section 3 qui s’appuie sur deux aspects : une méthode de caractérisation de l’intérêt d’une pose, cf. Section 3.1, et une méthode d’évaluation de la qualité d’une représentation 2D de cet objet, cf. Section 3.2. Afin d’évaluer cette qualité, nous étudions différentes mesures, cf. Section 3.3. Enfin, la section 4 nous permettra de présenter des résultats de classement de poses plus ou moins favorables mais également des images en fonction de leur capacité à mettre en valeur un objet, avant de conclure et donner nos perspectives en section 5.

2 Qualité d’une vue 2D à partir de primitives saillantes

L’évaluation de la qualité d’une image dans sa globalité est un concept très subjectif. En effet, il existe de nombreuses mesures de similarité ou d’évaluation de la qualité mais il est difficile de déterminer le lien avec la perception visuelle humaine. Une approche classique consiste à

1. « laser imaging detection and ranging »

2. « radio detection and ranging »

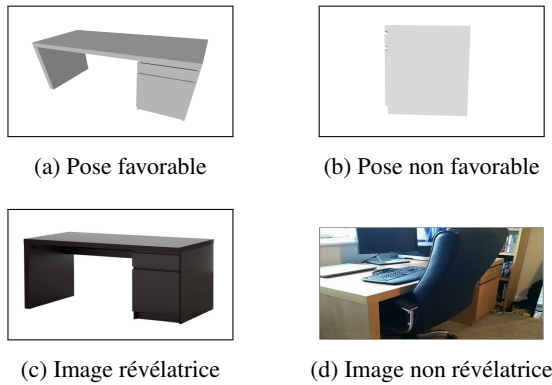


FIGURE 2 – Illustration des notions de pose favorable et d’image révélatrice. En (a), il s’agit d’une orientation favorable avec de nombreux détails alors qu’en (b), l’orientation est non favorable. En effet, le bureau est vu sur le côté. Sur la deuxième ligne, nous présentons deux images, contenant le même modèle de bureau. En (c), l’image est révélatrice du bureau alors qu’en (d) elle est non révélatrice. Le bureau est globalement occulté en (d).

évaluer la qualité de l’image en mesurant le niveau de flou ou la présence de bruit, comme dans [1]. Certains travaux s’appuient sur l’information colorimétrique, comme [7]. Une approche récente à base de réseaux de neurones est capable de trouver la combinaison de meubles 3D pour une décoration de pièce, la plus harmonieuse du point de vue du style [12]. Ces derniers mesurent à la fois la compatibilité et les similarités entre les différents styles des meubles. Le travail que nous proposons est complémentaire aux travaux que nous citons puisque nous souhaitons qualifier l’intérêt d’une vue d’un objet 2D en utilisant une information différente : la présence de points caractéristiques.

En 2D, il existe de nombreux détecteurs et il est classique d’utiliser des détecteurs du premier ordre qui s’appuient sur les dérivées premières des images. Dans cette catégorie, nous retrouvons des détecteurs classiques comme celui de [6]. Une autre catégorie de détecteurs s’appuie sur la notion de région. Plus précisément, nous considérons qu’il existe une région caractéristique autour du point étudié. L’étude de cette région permet de déterminer le caractère unique du point étudié. Nous pouvons citer les détecteurs connus comme [14, 15, 17]. Enfin, une dernière famille de détecteurs sont ceux utilisant les dérivées secondes et plus précisément la notion de courbure [4]. Plus récemment, la notion d’analyse multi-échelle a été introduite pour chacun de ces types de détecteur. En effet, suivant l’échelle d’analyse choisie, nous détectons des détails caractéristiques différents. De plus, avec une échelle trop faible, des détails non significatifs peuvent être pris en compte et produire des détections non pertinentes alors qu’avec une échelle trop importante, des détails pertinents de l’image peuvent être ignorés. Pour palier à tous ces problèmes, l’approche célèbre de SIFT, *Scale Invariant Features Transform*, a été introduite [9]. En 3D, nous retrouvons les primitives du

premier ordre, du second ordre et s’appuyant sur la courbure. La plupart des techniques s’appuient sur la généralisation des détecteurs du 2D au 3D, comme le montre de nombreuses publications qui visent à généraliser le détecteur SIFT à la 3D [5].

Dans le travail que nous initions, il est nécessaire d’avoir un détecteur permettant de récupérer des éléments saillants répétables, aussi bien en 2D qu’en 3D. La difficulté principale réside dans la nature des données manipulées. En effet, une image 2D présente une texture, un fond et un éclairage que nous ne retrouvons pas dans le modèle 3D non texturé. Le détecteur [13] s’appuie sur la notion de courbure tout en diminuant l’effet de la texture (grâce à une analyse multi-échelle et la prise en compte du concept de carte de *focus*). Les résultats présentés dans [13] montrent que ce détecteur permet la meilleure répétabilité des points extraits sur les différentes modalités en comparaison avec les détecteurs de l’état de l’art comme SIFT.

Pour conclure, dans ce papier, pour déterminer la qualité du point de vue, nous proposons d’utiliser les informations saillantes extraites dans chaque type de données, à savoir les images 2D et les modèles 3D. Pour comparer ces données, il est d’abord nécessaire de déterminer une représentation commune. Dans la littérature, un choix classique consiste à projeter le modèle 3D suivant différents points de vue pour obtenir une collection de cartes de profondeur et nous suivons ce choix. Les cartes de profondeur utilisées sont simplement générées en effectuant différentes projections, suivant différents points de vue, du maillage.

Enfin, pour détecter les **points saillants**, il nous semble important de favoriser un détecteur qui soit le plus répétable possible. En effet, la comparaison n’aura de sens que si nous observons, c’est-à-dire, nous comparons les mêmes points de l’objet. Nous avons cité de nombreux détecteurs en 2D et en 3D mais nous faisons le choix de nous appuyer sur la saillance curviligne.

3 Proposition pour évaluer la qualité du point de vue d’un objet 3D

3.1 Intérêt de la pose d’un objet 3D

Orientation favorable. L’objectif est de déterminer, pour un modèle 3D donné, les orientations qui lui sont les plus favorables, c’est-à-dire les poses qui permettront d’extraire le plus d’information possible. Par exemple, si nous observons une bibliothèque, nous voulons voir les étagères et non le dos de la bibliothèque. Nous supposons que plus il y a de points saillants visibles, plus la pose, dans laquelle nous observons, est avantageuse, c’est-à-dire favorable à l’objet. Plus précisément, comme illustré dans la Figure 3, à partir d’un modèle 3D, (a), et des cartes de profondeur générées, (b), nous estimons les cartes de saillance curviligne, (c). Puisque nous traitons des objets correspondant à des meubles, pour rester réaliste, nous appliquons successivement des rotations autour de l’axe vertical pour créer les

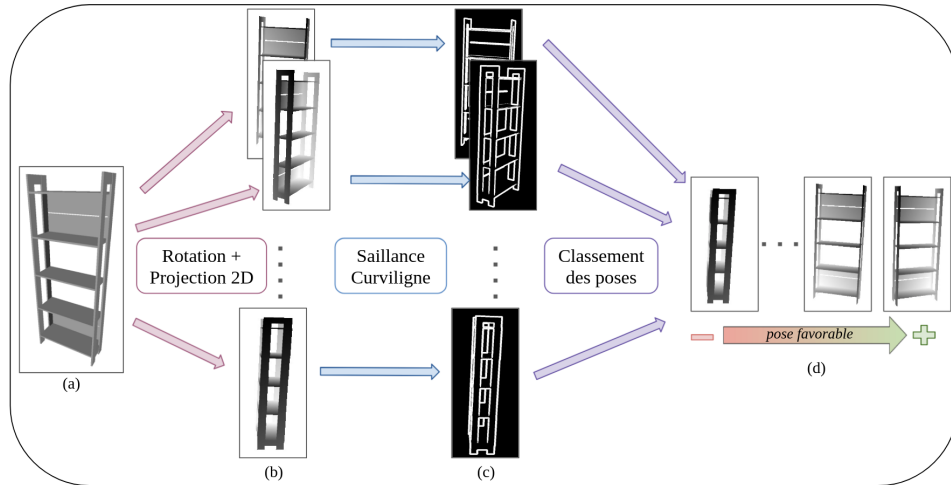


FIGURE 3 – Chaîne de traitement pour classer les poses. À partir du modèle 3D étudié, en (a), nous générons un ensemble de cartes de profondeur de n poses différentes, en (b). Pour chacune de ces cartes, nous calculons les cartes de saillance curviligne, en (c), qui nous permettent de proposer le classement des poses de la plus favorable à la moins favorable, en (d).

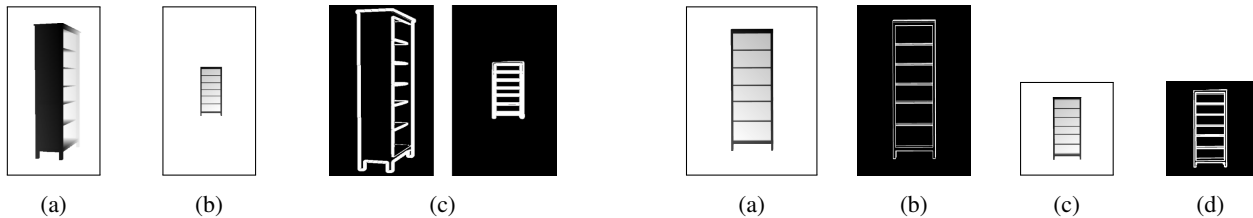


FIGURE 4 – Illustration de différentes orientations. En (a), l'orientation est peu favorable alors que la position est favorable. En (b), l'orientation est favorable mais la position est peu favorable. Les cartes de saillance curviligne associées en (c) possèdent 23K (à gauche) et 13K (à droite) points saillants.

cartes de profondeur. Le nombre de points d'intérêt obtenu nous permet de proposer la classement en (d).

Pose favorable. Dans le paragraphe précédent, nous avons uniquement considéré l'orientation de l'objet qui correspond seulement à une partie de la pose. Pour être complet, nous devrions prendre en considération la position de l'objet dans l'image ainsi que sa résolution. En effet, un objet, vu dans une image, peut avoir une orientation favorable, mais être éloigné, cf. Figure 4(b). De même, la résolution de l'image a une influence sur le nombre de points saillants détectés et cela peut avoir un impact sur la qualité de la vue observée. En effet, pour deux images représentant exactement la même vue d'un objet, celle de plus grande résolution contient plus de points saillants que celle de plus petite résolution, cf. Figure 5.

3.2 Qualité d'une image d'un objet

Nous souhaitons quantifier la qualité d'une image réelle d'un objet 3D, c'est-à-dire, à partir d'un ensemble

FIGURE 5 – Influence de la résolution. Le nombre de points saillants détectés est plus important dans une image de grande résolution (a) que dans une de plus petite résolution (c). Nous avons respectivement 177K et 42K points saillants dans les cartes de saillance curviligne (b) et (d).

d'images contenant le même objet 3D, identifier les images les plus révélatrices de ce dernier, cf. Figure 6. C'est à dire celles qui mettent en valeur les éléments caractéristiques de l'objet. Intuitivement, l'image sera révélatrice si l'objet n'est ni occulté, ni tronqué et s'il appartient au premier plan. De plus, il est préférable qu'il soit placé dans une scène peu complexe pour réduire le risque que d'autres objets de la scène ne soit plus mis en avant que lui.

Pour chacune des images de l'objet, nous calculons la carte de profondeur associée (la pose et le calibrage étant connus) et sa carte de saillance curviligne CS , cf. Figure 6(e), comme expliqué en Section 3.1.

En ce qui concerne l'image, contrairement à la carte de profondeur, elle est composée de formes et de textures. Par conséquent, comme préconisé dans [13], nous appliquons la saillance curviligne multi-échelle MCS . Une pyramide gaussienne multi-échelle est donc calculée et les cartes de saillance curviligne associée également. Les points d'intérêt retenus sont ceux qui apparaissent dans plusieurs échelles consécutives, cf. Figure 6(d).

Ainsi, nous avons une carte de saillance curviligne prove-

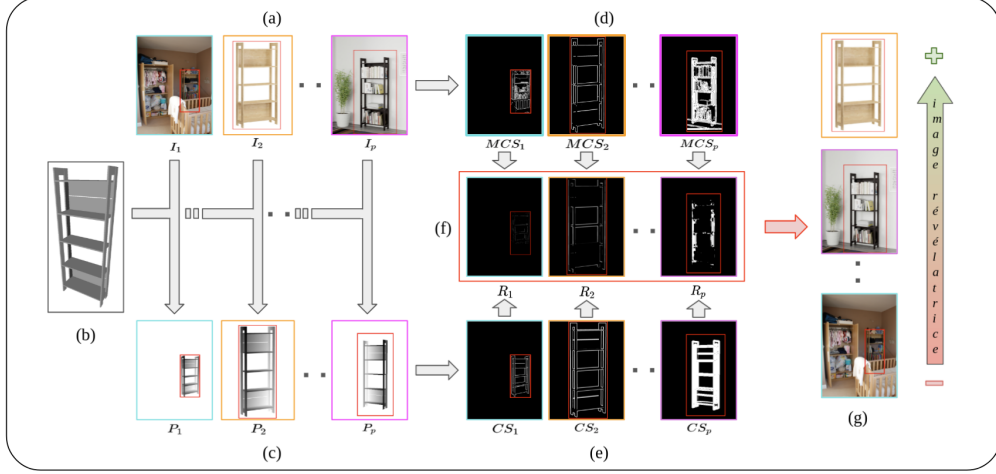


FIGURE 6 – Chaîne de traitement pour classer les images. En (a), il s’agit de l’ensemble des images notées I_i contenant le même objet et en (b) le modèle 3D associé. Pour chaque image I_i , nous sommes capables de calculer sa carte de profondeur P_i , (c). Ensuite, nous estimons les cartes de saillance curviligne multi-échelle MCS_i associées aux images I_i , en (d), ainsi que les cartes de saillance curviligne CS_i associées aux cartes de profondeur P_i , en (e). En (f), l’extraction des points saillants appartenant à l’objet dans les cartes réponses R_i , nous permet de proposer le classement des images de la moins révélatrice (en bas) à la plus révélatrice (en haut), en (g).

nant de la carte de profondeur de l’objet, et une carte de saillance curviligne multi-échelle, provenant de l’image. À partir de ces cartes seuillées, nous obtenons un ensemble de points d’intérêt pour chaque modalité. L’objectif suivant est de conserver uniquement les points saillants appartenant à l’objet, c’est-à-dire redondant entre l’image et la carte de profondeur. Nous supposons que la carte de profondeur nous fournit, via la carte de saillance curviligne, l’intégralité de l’information caractéristique de l’objet alors que les éléments saillants de l’image ne sont pas forcément dus à la présence de l’objet d’intérêt. Il est donc nécessaire de faire un tri pour récupérer les points saillants appartenant à l’objet, soit ceux appartenant à l’intersection entre la carte de saillance curviligne et la carte de saillance curviligne multi-échelle. La section suivante détaille les méthodes de mise en correspondance de ces deux cartes (CS et MCS), afin de déterminer la réponse des points saillants de l’objet 3D présents dans l’image (la carte réponse est illustrée en, cf. Figure 6(f)).

3.3 Métriques utilisées

Nous faisons l’hypothèse que si nous sommes en présence de deux points homologues alors leur voisinage doivent être similaires, c’est-à-dire contenir la même distribution de points d’intérêt. Ainsi, nous comparons les deux voisinages des deux points correspondants entre l’image et la carte de profondeur. Dans notre étude, deux points sont dits correspondants lorsque les deux sont détectés à la même position, respectivement dans la carte de saillance curviligne et la carte de saillance curviligne multi-échelle. Nous pouvons supposer qu’il y a parfois un décalage entre les positions des deux points. C’est pourquoi nous avons testés l’utilisation d’une fenêtre glissante. Nous ne donnons

pas les résultats dans les expérimentations, par manque de place. Toutefois l’ajout de cette fenêtre glissante n’améliore pas la qualité des résultats obtenus. Il existe dans la littérature plusieurs métriques capables de mesurer le degré de similarité entre deux voisinages et nous présentons celles que nous avons considérées.

Nous notons p_i et p_c deux points détectés respectivement dans la carte de saillance curviligne multi-échelle (relative à l’image) et dans la carte de saillance curviligne (liée à la carte de profondeur associée).

Nous notons $N \times N$ le voisinage carré pris en compte et celui-ci sera noté \mathbf{V}_i pour le voisinage de p_i dans la carte de saillance curviligne multi-échelle et \mathbf{V}_c pour le voisinage de p_c dans la carte de saillance curviligne. Nous avons considéré deux types de mesures de similarité : s’appuyant sur une distance classique ou utilisant des outils de statistiques classiques.

Métriques utilisant une distance. Il s’agit des mesures faisant intervenir une distance. Les plus connues sont :

- ▷ **Distance de Hausdorff** : Les deux voisinages sont préalablement binarisés : $\forall k \in \{i, c\}, \forall lig, col \in [1, N]^2$,

$$\mathbf{V}_k[lig, col] = \begin{cases} 1 & \text{si } \mathbf{V}_k[lig, col] \neq 0 \text{ (point saillant)} \\ 0 & \text{sinon.} \end{cases} \quad (1)$$

Et la distance est donnée par :

$$d(\mathbf{V}_c, \mathbf{V}_i) = \max \left\{ \sup_{p_c \in \mathbf{V}_c} \inf_{q_i \in \mathbf{V}_i} \delta(q_i, p_c), \sup_{q_i \in \mathbf{V}_i} \inf_{p_c \in \mathbf{V}_c} \delta(q_i, p_c) \right\}. \quad (2)$$

- ▷ **Sum of absolute differences** : elle correspond à la somme des différences en valeur absolue entre les deux voisinages \mathbf{V}_c et \mathbf{V}_i . Cette métrique correspond à la norme L_1 des différences. Les voisinages utilisés ne sont pas binarisés et contiennent les valeurs de saillance curviligne. Cependant pour s’assurer d’être dans le même intervalle de valeur, nous recalons les valeurs de \mathbf{V}_i autour de la même moyenne que \mathbf{V}_c et nous obtenons :

$$\text{SAD}(\mathbf{V}_c, \mathbf{V}_i) = \|\mathbf{V}_c - \frac{\overline{\mathbf{V}_c}}{\overline{\mathbf{V}_i}} \cdot \mathbf{V}_i\|_1 \quad (3)$$

avec $\forall k \in \{i, c\}$, $\overline{\mathbf{V}_k}$ la moyenne des intensités de \mathbf{V}_k .

Métriques utilisant les statistiques. Pour les métriques suivantes, les deux voisinages sont binarisés. Voici les mesures que nous considérons :

- ▷ **Indice de Jaccard** : Il est défini par :

$$J(\mathbf{V}_c, \mathbf{V}_i) = \frac{|\mathbf{V}_c \cap \mathbf{V}_i|}{|\mathbf{V}_c \cup \mathbf{V}_i|} \quad (4)$$

- ▷ **Précision-Rappel** : La précision est la proportion des éléments pertinents parmi l’ensemble des éléments sélectionnés. Tandis que le rappel est la proportion des éléments pertinents sélectionnés parmi l’ensemble des éléments pertinents.

$$\text{Précision} = \frac{VP}{VP + FP} \quad (5)$$

$$\text{Rappel} = \frac{VP}{VP + FN} \quad (6)$$

avec VP les points saillants dans les deux cartes, FP les points saillants détectés uniquement dans la carte de saillance curviligne multi-échelle et FN les points saillants détectés uniquement dans la carte de saillance curviligne.

4 Résultats de la qualité d’une vue

Dans un premier temps, nous présentons les bases de données utilisées et dans un second temps, nous présentons et commentons les résultats que nous avons obtenus.

4.1 Bases de données utilisées

La base Pascal3D+ [18] regroupe une dizaine, voire une centaine, de catégories d’objets. Chacune de ces catégories possède des dizaines de modèles 3D différents avec des milliers d’images 2D associées. Il existe également la base de données [8] qui contient des couples d’images avec leur carte de profondeur. Cependant, ces ensembles de données existants ont certaines limites : les modèles 3D peuvent ne pas correspondre aux objets des images ; les annotations de pose peuvent être imprécises ; l’ensemble de données peut être relativement petit ; les relations entre les diverses

données ne sont pas fournies. En conséquence, nous avons choisi la base de données Pix3D [16] qui contient 9 catégories de modèles 3D : {lit, bibliothèque, chaise, bureau, canapé table, outils, penderie, objets divers}. Chaque catégorie d’objets possède plusieurs dizaines de modèles différents et entre des centaines et des milliers d’images, plus ou moins révélatrices de l’objet qu’elles contiennent. Un exemple est fourni dans la Figure 2. Chacune image possède une annotation précise de la pose 3D pour assurer un alignement 2D-3D précis. Ainsi, nous avons à notre disposition les poses des caméras ainsi que d’autres annotations telles que des points d’intérêt 2D et 3D appartenant respectivement aux images et aux modèles 3D, les masques des objets dans les images et les boîtes englobantes.

4.2 Expérimentations

Classement des orientations et poses favorables. Nous générons $n = 10$ poses pour chaque objet 3D en effectuant une rotation autour d’un axe vertical. La figure 7(a) illustre les cartes de profondeur et (b) les cartes de saillance curviligne correspondantes. Pour le filtre gaussien utilisé pour le calcul de la saillance curviligne, nous avons fixé $\sigma = 1.4$. Dans la figure 7, nous présentons les résultats obtenus en supposant une position et un point de vue fixe. Seule l’orientation de l’objet est variable. Les résultats obtenus montrent que les poses les plus favorables que nous obtenons sont conformes à ce qui était attendu, c’est-à-dire que notre algorithme conclut que celles qui contiennent le plus d’information sur l’objet, sont celles avec l’objet vu de face ou éventuellement de biais. La Figure 8 montre le classement obtenu en faisant varier l’orientation mais également la position relative de l’objet et son point de vue. Nous observons que le classement obtenu est cohérent par rapport à une qualité subjective. Néanmoins, une évaluation basée sur une étude utilisateur serait nécessaire pour valider plus objectivement notre mesure de qualité.

Classement des images révélatrices. Pour réaliser le classement des images de la moins révélatrice à la plus révélatrice de l’objet, nous avons utilisé la saillance curviligne multi-échelle pour l’image 2D. Plus précisément, nous construisons, pour chaque image, une pyramide gaussienne constituée de $nb_echelle = 4$ échelles, respectivement filtrées avec :

$$\sigma_i = \sigma k^{i-1}, \forall i \in [1, nb_echelle]$$

avec $k = 2^{\frac{1}{nb_echelle}}$ et $\sigma = 1.4$. De plus, nous conservons, pour chaque échelle, uniquement 90% des valeurs de saillance curviligne les plus élevées. En effet, les 10% plus basses sont considérées comme du bruit. De plus, un point saillant est conservé dans ce processus, s’il est présent dans au moins $N = 3$ échelles consécutives.

Dans nos expérimentations, nous avons choisi des ensembles d’images qui minimisent les biais induits par l’orientation, la position de l’objet et la taille des images. Ces choix nous permettent d’obtenir un nombre relativement homogène de points saillants dans les différentes

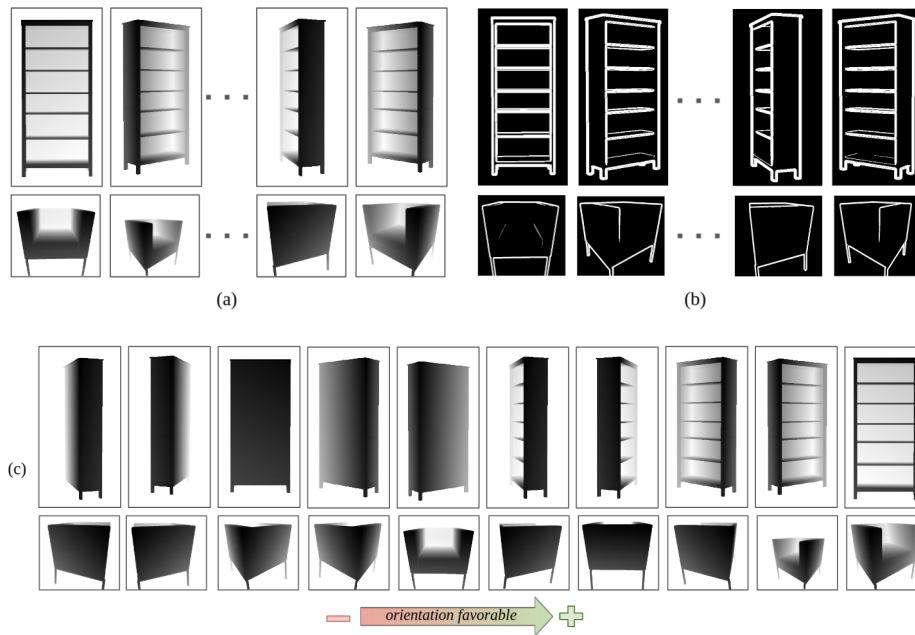


FIGURE 7 – Cartes de profondeur avec des poses différentes, (a) cartes de profondeur après rotation, (b) cartes de saillance, (c) classement des poses de la moins favorable (à gauche) à la plus favorable (à droite).

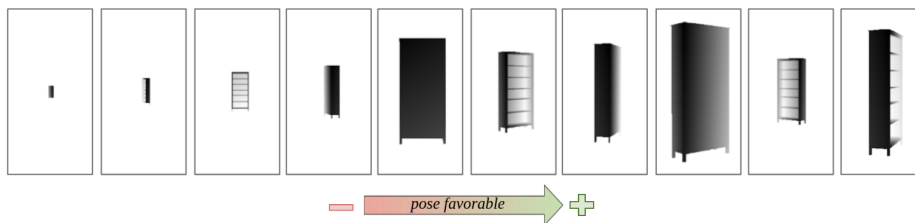


FIGURE 8 – Classement des poses.

cartes de saillance curviligne. Ceci nous permet de nous appuyer sur le rapport entre le nombre de points saillants extraits par nos différentes méthodes, présents dans la carte réponse de chaque couple (image, carte de profondeur) et le nombre de points saillants présents dans chacune des cartes de saillance, pour indiquer à quel point l'image est révélatrice de l'objet.

La Figure 10 montre que chaque métrique utilisée permet d'obtenir la même image en première et en dernière place. Nous avons établi un classement subjectif de ces images et nous espérons l'obtenir. Plus précisément, la première image est naturellement considérée comme l'image la plus révélatrice (aucun objet occultant ou également saillant dans son voisinage). Les trois images intermédiaires sont équivalentes. Enfin, la dernière image est occultée mais sa taille est légèrement plus grande que les autres, cf. Figure 5. Malgré cet aspect avantageux, il est intuitif de la classer en dernière position. Cette méthode de classement pourrait être utilisée pour diverses applications qui nécessitent une prise de décision, telles que sélectionner les

images les plus pertinentes pour mettre dans un catalogue dédié à la vente.

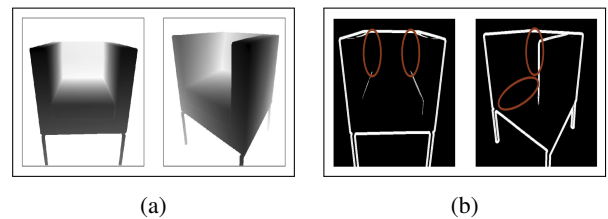


FIGURE 9 – Cartes de profondeur en (a), cartes de saillance curviligne en (b) avec les zones où il manque de l'information entourées en rouge.

5 Conclusion et perspectives

Nous avons proposé deux méthodes afin d'évaluer, d'une part, l'intérêt d'une pose d'un objet, et d'autre part, la pertinence d'une image 2D. Les résultats préliminaires ob-

tenus sont encourageants mais nous souhaitons améliorer certains aspects.

Tout d'abord, nous avons constaté que certaines cartes de profondeur ne contiennent pas toute l'information souhaitée. En effet, comme illustré dans la Figure 9, il manque des ensembles de points saillants qui devraient correspondre aux discontinuités situées au fond du fauteuil. Seuls les contours externes sont détectés. Ce manque d'information caractéristique est un biais lorsque nous déterminons le classement des poses favorables. L'ordre espéré n'est pas retrouvé. Un exemple d'un tel classement est visible dans la Figure 7(c) pour le fauteuil. Il est donc nécessaire de trouver un moyen pour détecter la totalité des points saillants d'une carte de profondeur.

De plus, notre approche n'est pas suffisamment robuste aux occultations. En effet, dans l'exemple de la carte de réponse de la dernière image du classement obtenu dans la Figure 10, il y a beaucoup de points saillants qui appartiennent à l'objet occultant (la nappe). Ces points auraient du être rejetés.

Enfin, pour réaliser l'évaluation de la qualité d'une image 2D, trois aspects peuvent être pris en compte : l'orientation, la position de l'objet ainsi que la résolution de l'image. Dans ce papier, nous avons supposé que ces trois aspects étaient fixes. La prochaine étape est d'étudier des images avec un objet dans différentes orientations, et/ou différentes positions, et/ou une résolution différente, en utilisant à la fois l'information liée à l'orientation de l'objet et sa position, pour construire le classement des images. Ainsi nous souhaitons élaborer une évaluation de chaque image permettant de prendre en compte ces trois critères en les combinant.

Références

- [1] M. BEN AMOR, A. SAMET, F. KAMMOUN, et N. MASMOUDI. « exploitation des caractéristiques du système visuel humain dans les métriques de qualité ». Dans *AMINA*, 2010.
- [2] A. BIGLIA et P. BELLEFLAMME. « Analyse prospective sur l'implémentation de la voiture autonome : impact sur l'industrie automobile et le citoyen ». Mémoire de master, Université Catholique de Louvain, Belgique, 2015.
- [3] J. DENG, W. DONG, R. SOCHER, L.-J. LI, K. LI, et L. FEI-FEI. « Imagenet : A large-scale hierarchical image database ». Dans *IEEE CVPR*, 2009.
- [4] P. FISCHER et T. BROX. « Image Descriptors Based on Curvature Histograms ». Dans *GCPR*, 2014.
- [5] G. FLITTON, T. BRECKON, et N. MEGHERBI BOUALLAGU. « Object Recognition using 3D SIFT in Complex CT OPTvolumes ». Dans *BMVC*, 2010.
- [6] C. HARRIS et M. STEPHENS. « A combined corner and edge detector ». Dans *Alvey Vision Conference*, 1988.
- [7] M. HERBIN, M. CHAMBAH, E. ZAGROUBA, et S. OUNI. « Vers une métrique de description objective d'une sensation subjective ». *Revue Africaine de la Recherche en Informatique et Mathématiques Appliquées*, 2009.
- [8] A. JANOCH, S. KARAYEV, Y. JIA, J. T. BARRON, M. FRITZ, K. SAENKO, et T. DARRELL. A category-level 3d object dataset : Putting the kinect to work. Dans *Consumer depth cameras for computer vision*. Springer, 2013.
- [9] D. G. LOWE. « Distinctive image features from scale-invariant keypoints ». *IJCV*, 2004.
- [10] T. MALON, P. GUYOT, G. ROMAN-JIMENEZ, S. CHAMBON, V. CHARVILLAT, A. CROUZIL, A. PÉNINOU, J. PINQUIER, F. SÈDES, et C. SÉNAC. « Toulouse campus surveillance dataset : scenarios, soundtracks, synchronized videos with overlapping and disjoint views ». Dans *MMSys*, 2018.
- [11] W. H. NAM, D.-G. KANG, D. LEE, J. Y. LEE, et J. B. RA. « Automatic registration between 3D intra-operative ultrasound and pre-operative CT images of the liver based on robust edge matching ». *Physics in Medicine & Biology*, 2011.
- [12] T.-Y. PAN, Y.-Z. DAI, W.-L. TSAI, et M.-C. HU. « Deep model style : Cross-class style compatibility for 3d furniture within a scene ». Dans *IEEE ICBD*, 2017.
- [13] H. A. RASHWAN, S. CHAMBON, P. GURDJOS, G. MORIN, et V. CHARVILLAT. « Using Curvilinear Features in Focus for Registering a Single Image to a 3D Object ». *IEEE Transactions on IP*, 2019.
- [14] E. ROSTEN et T. DRUMMOND. « Machine Learning for High-Speed Corner Detection ». Dans *ECCV*, 2006.
- [15] S. SMITH et J. BRADY. « SUSAN - A New Approach to Low Level Image Processing ». *IJCV*, 1997.
- [16] X. SUN, J. WU, X. ZHANG, Z. ZHANG, C. ZHANG, T. XUE, J. B. TENENBAUM, et W. T. FREEMAN. « Pix3D : Dataset and methods for single-image 3D shape modeling ». Dans *IEEE CVPR*, 2018.
- [17] T. TUYTELAARS et L. VAN GOOL. « Matching Widely Separated Views Based on Affine Invariant Regions ». *IJCV*, 2004.
- [18] Y. XIANG, R. MOTTAGHI, et S. SAVARESE. « Beyond PASCAL : A Benchmark for 3D Object Detection in the Wild ». Dans *WACV*.
- [19] W. YANG, D. YI, Z. LEI, J. SANG, et S. Z. LI. « 2D-3D face matching using CCA ». Dans *IEEE AFGR*, 2008.

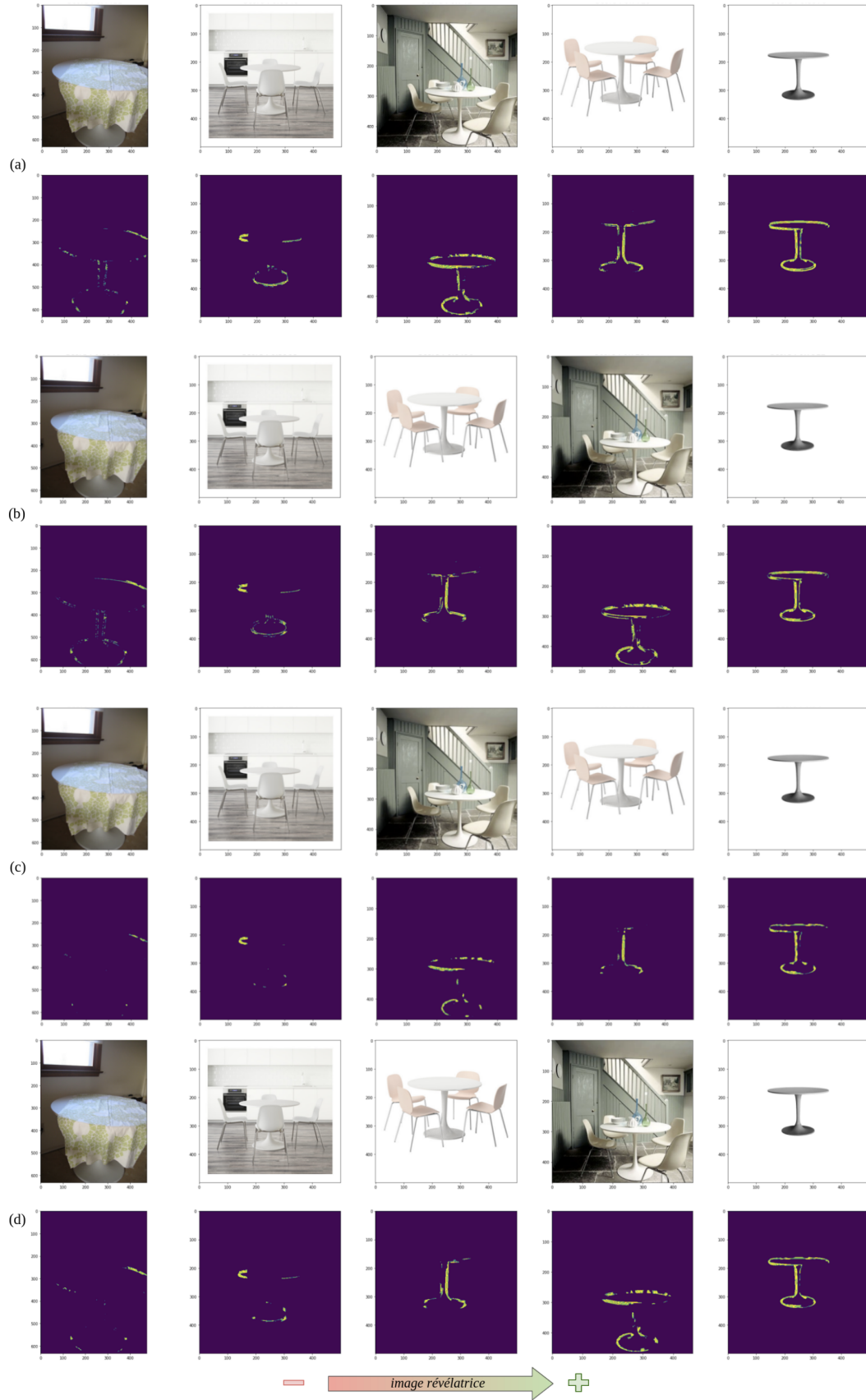


FIGURE 10 – Classement des images de la moins révélatrice (à gauche) à la plus révélatrice (à droite) pour chaque métrique en utilisant la distance de Hausdorff, (a), SAD, (b), l'indice de Jaccard, (c), la précision-rappel, (d).