Indexation par sac-de-mots d'objets avec modèles déformables isométriques : apprentissage sur données synthétiques et vérification géométrique 3D

R. Rantoson A. Bartoli

ALCOV-ISIT, CNRS/Université d'Auvergne, Clermont-Ferrand, France
rindra.sanders@udamail.fr

Résumé

Les méthodes existantes de vérification spatiale lors de l'indexation d'objets déformables n'utilisent pas de modèle physique. Nous proposons d'utiliser SfT (Shape-from-Template), une technique qui reconstruit un objet isométrique à partir de son modèle 3D par estimation d'une déformation 3D. Le couplage indexation/SfT est bénéfique pour les deux méthodes, car l'indexation fournit à SfT le modèle 3D nécessaire à son fonctionnement. Nous avons également exploité le modèle physique pour la génération de données durant la phase d'apprentissage de l'indexation. A la différence de l'indexation classique par sac-de-mots utilisant un TDS (Template Descriptor Set), un ensemble de vecteurs de fréquence de mots des objets de la base de données, nous introduisons la MTDS, une combinaison de Multiples TDS. Les TDS de base sont générés à partir des critères de discriminance, de répétabilité et de stabilité des mots parus dans les données d'apprentissage et sélectionnés de manière à maximiser la performance du détecteur sur les données de validation. Des résultats d'indexation et de reconstruction d'objets réels concluants sont présentés à partir d'une base de données de 100 objets.

Mots Clef

Indexation, sac-de-mots, reconstruction 3D, SfT, objet, isométrie.

Abstract

Geometric verification in deformable objects retrieval is generally not handled with physical models. We propose to use SfT (Shape-from-Template), a technique which reconstructs an object from its template via the estimation of the 3D deformation. The retrieval/SfT combination is beneficial for both methods since retrieval provides to SfT with the required template. The physical deformation model also allows data generation. Thus, we have exploited it for the retrieval learning phase. Unlike standard BoVW (Bag-of-Visual-Words) based retrieval method using a TDS (Template Descriptor Set) which is a word frequency vectors set of the database objects, we introduce the notion of MTDS (Multiple Template Descriptor Sets), a combination of several TDS. These are generated by applying discriminancy, repeatability and stability criteria on learning data words. The TDS which maximise the detection performance on validation data are retained to define the MTDS. Performance evaluation on test data reveals the benefit of our retrieval approach using a database of 100 objects. Some real object detection and reconstruction are presented.

Keywords

Image retrieval, bag-of-visual-words, 3D reconstruction, SfT, object, isometry.

1 Introduction

L'indexation et la reconstruction 3D d'objets à partir d'une unique image sont deux thématiques abondamment traitées dans la littérature et qui ont connu d'importants progrès. D'une part, nous avons la démocratisation de l'approche par sac-de-mots avec l'utilisation de TF-IDF (Term Frequency-Inverse Document Frequency) [26] communément couplée avec une vérification spatiale [23, 13, 24, 8] pour l'indexation. D'autre part, la reconstruction 3D basée modèle d'objets isométriques rencontre un récent succès avec SfT [17, 4]. L'approche d'indexation que nous proposons permet de s'affranchir des limites actuelles de chacune d'elles : le manque d'un modèle physique valable pour la vérification géométrique des objets déformables et l'applicabilité de SfT restreint à un objet. Nous avons choisi la représentation par sac-de-mots pour notre base de données bien que d'autres méthodes d'encodage soient également efficaces telles que VLAD (Vector of Locally Aggregated Descriptor) [12, 10] et FV (Fisher Vector) [19]. Notre base de données contient l'apparence, la forme de référence et la loi de déformation des objets. SfT est utilisé pour vérifier physiquement chaque hypothèse d'objet durant le processus d'indexation alimentant la reconstruction à partir des connaissances a priori. Une tentative d'appliquer purement SfT sans indexation a été entreprise, limitant son exécution sur une base de données de 10 objets maximum [1]. En outre, bien que SfT fonctionne en temps réel pour un seul objet [17, 7], il n'est pas compatible avec une large base de données. Par le biais de notre approche, le coût de SfT demeure faible tout en s'adaptant à la taille de la base de données. Celle-ci pourrait être

implémentée dans un système de détection et de reconstruction rapide d'objets par un capteur visuel monoculaire où les objets à reconnaître seraient préalablement enregistrés en mémoire, pour des applications diverses telles que l'interaction homme-machine et la réalité augmentée. L'approche que nous proposons consiste à coupler la méthode standard par sac-de-mots [26] avec SfT pour une meilleure indexation des objets déformables. Dans l'approche classique, les descripteurs sont extraits des images et classifiés en mots visuels. Par la suite, chaque image est représentée par un vecteur de mots. L'ensemble des vecteurs de fréquence des mots des objets de la base de données forme un TDS. L'indexation se déroule en deux étapes successives. D'abord, les objets sont classés selon leur score de similarité obtenu avec l'image requête[1]. Ensuite, la vérification spatiale est appliquée sur les R premiers objets visant à améliorer le premier classement (reclassement), R étant défini en fonction de l'application considérée. De nombreuses pistes d'amélioration sont suggérées sur des points clés de la méthode [3], notamment la détection et la quantification des descripteurs d'image [16, 27], la pondération des mots visuels dans TDS [11, 29], la métrique de similarités utilisées [22], l'apprentissage de multiple images requête [2, 5], le reclassement par graphe [21], et les contraintes spatiales [28, 25]. Concernant ces dernières, la plupart des méthodes suggérées sont principalement dédiées aux objets rigides. Bien que certaines soient applicables aux objets déformables comme la contrainte de proximité spatiale [24], l'approche par pyramide spatiale [23, 13] ou encore le modèle DPM (Deformable Part-based Model) [8], aucun modèle physique n'a été auparavant proposé pour les objets isométriques. Nous proposons d'introduire le modèle physique SfT, une technique qui reconstruit un objet isométrique à partir de son modèle 3D par estimation d'une déformation 3D. Le modèle 3D est muni de sa carte de texture. Le couplage indexation/SfT est bénéfique pour les deux méthodes, car l'indexation fournit à SfT le modèle 3D nécessaire à son fonctionnement. Par ailleurs, le modèle SfT permet de générer des données de déformation 3D. C'est pour cette raison que nous l'avons également exploité en mettant en place l'idée de MTDS définie par apprentissage, dont le but est d'améliorer les résultats du premier classement. Pour ce faire, une base de TDS générateurs est définie initialement. Des critères de discriminance, de stabilité et de fréquence des mots parus dans les données d'apprentissage sont appliqués sur cette base afin d'engendrer de nouveaux TDS. Parmi les TDS générés, un TDS est appris pour chaque objet en maximisant la performance de son indexation et détection sur les données de validation. L'ensemble des TDS retenus définit MTDS manipulée lors de la procédure d'indexation.

[1]L'image soumise au système: "query image" ou "input image" en anglais.

2 Stratégie

Notre approche d'indexation, qui a pour finalité la détection d'objets pour la reconstruction SfT et qui est fondée sur l'exploitation du modèle physique SfT, s'effectue en trois étapes distinctes:
- **E1 :** Détection individuelle de chaque objet au moyen de MTDS et des seuils respectifs de détection appris selon le pourcentage de vrais positifs exigé. Il s'agit de vérifier l'hypothèse de présence ou non de chaque objet de la base dans l'image requête.
- **E2 :** Classement des objets détectés selon le score normalisé obtenu.
- **E3 :** Vérification spatiale des R premiers objets à l'aide du modèle physique SfT et des seuils de détection appris. La détection est effectuée sur une image requête contenant un ou quelques objets.

2.1 Indexation par/pour SfT

Dans l'approche d'indexation classique, chaque document de la base de données B est représenté par un vecteur de fréquence des mots pondérés. Nous avons opté pour l'approche de pondération classique TF-IDF [26] dans laquelle pour un vocabulaire donné de k mots, un document est représenté par un k-vecteur $V_d = (t_1, ..., t_i, ..., t_k)^\top$ pour lequel les composants t_i sont calculés par:

$$t_i = \frac{n_{id}}{n_d} \log \frac{N}{n_i} \tag{1}$$

où n_{id} est le nombre d'occurrences du mot i dans le document d, n_d est le nombre total de mots du document d, n_i est le nombre de documents contenant le mot i dans B et N est le nombre de documents de B. L'ensemble des vecteurs V_d de B définit un TDS, $T = \{V_1, ..., V_N\}$. Plusieurs TDS sont générés à partir des cartes de texture et des déformations simulées des objets durant une phase d'apprentissage, cf. section 3. Vient ensuite une phase de validation qui consiste à sélectionner pour chaque objet le TDS qui maximise la performance de sa détection, cf. section 4.1. L'ensemble des TDS retenus constitue la combinaison $M = \{T_1, T_2, ..., T_N\}$ manipulée lors de procédure d'indexation décrite ci-après. En premier lieu, les descripteurs SIFT [14] de l'image requête sont calculés puis associés aux mots visuels du vocabulaire en utilisant la méthode des plus proches voisins. L'image requête est ensuite représentée par le vecteur poids TF-IDF dénoté V_q. A la première étape d'indexation, la présence de chaque objet i dans l'image requête est vérifiée en utilisant le T_i appris et conjointement le seuil de détection $T1_i$. Ce dernier est déterminé à partir de la courbe ROC_i relatif à T_i en fonction du taux minimum de vrais positifs exigé au cours de la phase de validation. La deuxième étape consiste à classer les objets détectés dans l'ordre décroissant selon le score de similarité normalisé $\hat{s}_i$. Le score de similarité s_i entre une image requête et un objet de la base de données est obtenu en calculant le produit scalaire normalisé de V_q et V_d. Le score normalisé $\hat{s}_i$ est défini comme étant la différence en-

tre le score de similarité s_i et le seuil $T1_i$. L'usage de ce critère contribue à la performance du classement en assurant l'homogénéité des scores obtenus issus des différents TDS. D'autres pistes avec ou sans l'utilisation de cette normalisation voire une autre méthode de normalisation consistant à diviser $\hat{s}_i$ par la différence entre le maximum et le minimum des scores, ont été explorées. Néanmoins, les résultats sont sans équivoque, le critère basé sur la différence donne la meilleure performance. A la troisième étape, au maximum, R premiers objets sont choisis pour la vérification spatiale ($R \leq 10$). Pour chaque objet testé, les descripteurs SIFT complets[2] de l'objet et ceux de l'image requête sont appariés par la méthode des plus proches voisins, puis les coordonnées des points appariés sont soumis à l'estimateur robuste de déformation SfT qui rejette les points aberrants [20]. La présence de l'objet dans l'image requête est déterminée par le nombre d'appariemments résultants. Le nombre de inliers varie en fonction du nombre de descripteurs complets de l'objet. Un seuil de inliers pour chaque objet a été appris au préalable en se basant sur la courbe ROC selon un taux minimum de vrais positifs requis. Les objets détectés sont fournis à SfT pour la reconstruction 3D [6].

2.2 Génération de données

Notre base de données est constituée de 100 objets. De nombreuses déformations 3D ont été simulées en utilisant l'approche décrite dans [18] afin de générer trois types de données à partir des modèles 3D :
- Les données d'apprentissage ($D1$) sont définies par 1200 déformations sans fond par objet pour un total de 120K. Elles résultent de la combinaison de 100 déformations avec 15 projections caméra (1500 déformations) à partir desquelles sont tirées aléatoirement les 1200 déformations pour chaque objet.
- Les données de validation ($D2$) sont définies par 1200 déformations avec fond par objet pour un total de 120K. Elles résultent de la combinaison de 80 nouvelles déformations avec 15 projections caméra, l'ensemble combiné aléatoirement avec 20 fonds pour chaque objet.
- Les données test ($D3$) sont définies par 300 déformations avec fond par objet pour un total de 30K. Il s'agit de 300 déformations restantes des 1500 déformations générées à l'apprentissage combinées avec 20 nouveaux fonds.

Les données d'apprentissage sont utilisées pour la construction du vocabulaire et pour la génération de plusieurs TDS tandis que les données de validation ont servi à leur sélection pour la définition de MTDS. Ces dernières ont permis également l'apprentissage des différents seuils employés. Enfin, les données test sont destinées à l'évaluation de la performance de MTDS et des étapes de l'indexation.
Nous adoptons un vocabulaire de 1000 mots construit à partir des descripteurs SIFT des cartes de texture et des descripteurs spécifiques des déformations des données d'apprentissage en utilisant la méthode des k-moyennes.
Ces descripteurs spécifiques sont des descripteurs des points spécifiques aux images de déformation. Il s'agit des points engendrés par la déformation, qui n'existent pas initialement dans les cartes de texture. Afin de les identifier, nous cherchons la correspondance des points des images de déformation avec ceux des cartes de texture. Pour ce faire, nous utilisons la rétro projection de notre modèle physique SfT et adoptons le critère de Mikolajczyk et al. [15]. Il a été prouvé que cette fonction de rétro projection est approximée localement par une fonction affine [6], illustrée sur la figure 1. Le critère de Mikolajczyk et al. consiste à comparer les deux régions covariantes de deux points de correspondance présumés en les ramenant sur une même image par transformation affine voir la figure 2. La correspondance est validée si la zone de recouvrement des deux régions vérifie un seuil minimum fixé, qui est de $0,45$ dans notre cas.

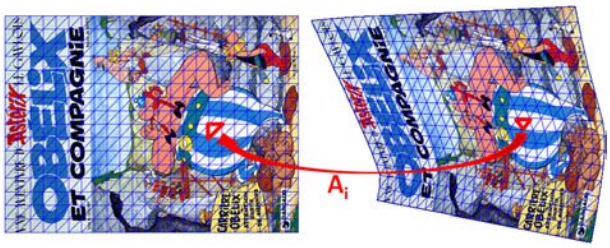


Figure 1: Illustration de l'approximation affine locale A_i, entre l'image de déformation (à droite) et la carte de texture (à gauche) d'un objet.

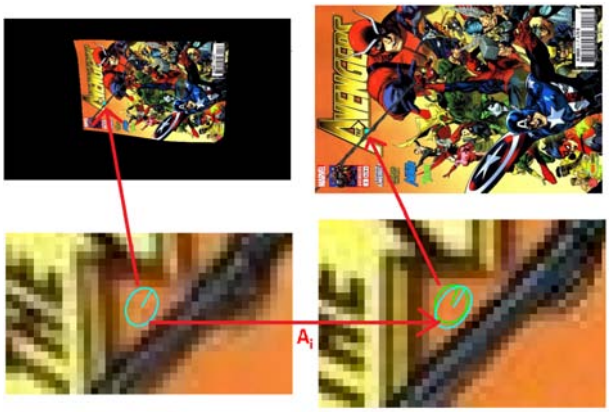


Figure 2: Illustration du critère de Mikolajczyk et al. [15].

Par la suite, nous caractériserons un objet :
- soit par les descripteurs initiaux de sa carte de texture ;
- soit par les descripteurs complets, qui comprennent les descripteurs initiaux de la carte de texture et additionnellement les descripteurs spécifiques des données d'apprentissage de déformation. Ceux-ci sont exploités notamment lors de la vérification spatiale.

[2]Il s'agit des descripteurs de la carte de texture et des images de déformation des données d'apprentissage de l'objet, cf. section 2.2.

3 Génération des TDS

Nous définissons au préalable trois TDS qui forment la base génératrice des TDS :

- T_1 relatif aux descripteurs initiaux des objets.
- T_2 relatif aux descripteurs des données d'apprentissage.
- T_3 relatif aux descripteurs complets des objets.

De nombreux TDS sont créés à partir de ces trois générateurs de base en appliquant trois critères comprenant la discriminance, la répétabilité et la stabilité des mots. Les critères sont appliqués séparément ou combinés deux par deux voire par trois, voir le tableau 1. Ce filtrage fournit une représentation différente des objets de la base de données en fonction du critère appliqué. Il a pour objectif d'exhiber les caractéristiques spécifiques des objets à travers les critères considérés.

	T_1	T_2	T_3
$D \geq T_{Di,i\in\{1,2,3\}}$	T_{11}	T_{12}	T_{13}
$R \geq T_R$	T_{21}	T_{22}	T_{23}
$S \geq T_S$	...	...	...
$D \cap R$	...	...	...
$D \cap S$	...	...	...
$R \cap S$	...	...	...
$D \cap R \cap S$	T_{A1}	T_{A2}	T_{A3}

Table 1: Génération des TDS (au nombre de $3 \times A$) en appliquant les critères de discriminance D, de répétabilité R et de stabilité S sur les TDS de la base $\{T_1, T_2, T_3\}$.

3.1 Discriminance

Le critère de la discriminance est appliqué sur les composants du TDS générateur, l'objectif étant de faire ressortir les objets dotés de mots discriminants. Rappelons qu'une finalité de la pondération TF-IDF consiste à attribuer un poids faible aux mots communs présents dans tous les objets de la base. Dans la mesure où plusieurs poids faibles contenus dans un objet peuvent perturber la reconnaissance de celui-ci, nous décidons de les filtrer par le critère de discriminance. Ce filtrage permet de définir un nouveau TDS par sélection des valeurs supérieures à un seuil prédéfini. Celui-ci est choisi à l'aide de l'histogramme des poids TF-IDF du TDS générateur. Une illustration de l'histogramme du nombre de mots de TDS1 en fonction des poids recensés est montrée sur la figure 3. Trois TDS en sont générés en considérant trois seuils, $T_{D_1} = 0,002, T_{D_1} = 0,005$ et $T_{D_1} = 0,01$. Divers seuils peuvent être appliqués[3] donnant ainsi lieu à la génération de plusieurs TDS en fonction du coût octroyé en terme de calcul et de mémoire. L'étape de validation assure ultérieurement la sélection finale des TDS retenus.

[3]De même pour les critères de répétabilité et de stabilité.

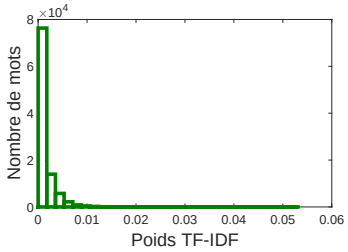


Figure 3: Histogramme du nombre de mots du T_1 en fonction de leurs poids.

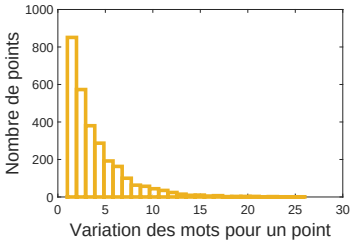


Figure 4: Histogramme du nombre de points d'un objet selon la variation de leurs mots dans les données d'apprentissage.

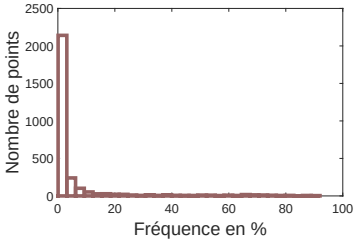


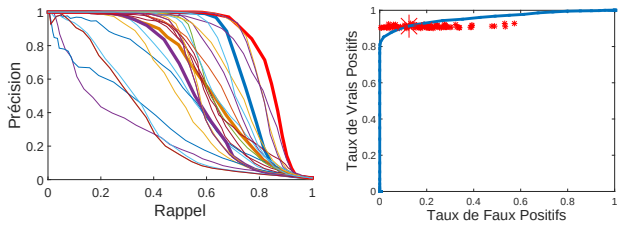
Figure 5: Histogramme du nombre de points d'un objet en fonction de leur fréquence dans les données d'apprentissage.

3.2 Répétabilité

Le critère de répétabilité (ou fréquence) vise à caractériser un objet par les mots des points qui résistent aux déformations, selon le niveau de résistance choisi. La répétabilité d'un point est quantifiée par le nombre de fois où le point apparaît dans les données d'apprentissage de déformation. Un seuil permet de sélectionner les points répétables et de filtrer en conséquence les poids des mots correspondants dans TDS, pour la génération d'un nouveau TDS. Trois seuils généraux ont été définis et appliqués aux trois TDS de base (10%, 30%, 60%). Un histogramme illustrant la répétabilité des points d'un objet est fourni sur la figure 5.

3.3 Stabilité

Le critère de stabilité permet de souligner les points d'un objet ayant de faible variation de mots dans les données de



(a) (b)

Figure 6: (a) Courbes de la précision moyenne/rappel issue de quelques TDS pour un objet donné : le TDS sélectionné en rouge, T_1 en bleu, T_2 en orange, T_3 en violet. L'aire sous une courbe définit moyPM. (b) La courbe ROC (en bleu) relative au TDS attribué à un objet et les points ROC retenus (en rouge) de tous les objets correspondants aux seuils appris selon le critère fixé $TP = 0,9$.

déformation et en conséquence d'exprimer le TDS à partir uniquement des poids des mots correspondants. Le mot associé au descripteur d'un point peut varier d'une donnée à une autre dû au clustering, à la méthode de correspondance utilisée ou encore aux déformations. Pour chaque point d'un objet, la variation des mots associés aux descripteurs de ses points de correspondance des déformations est relevée. Pour un seuil considéré, les points ayant de faible variation sont sélectionnés et les poids des mots correspondants sont retenus pour la génération d'un nouveau TDS. Deux seuils ont été testés $T_S = 3$, $T_S = 5$. Le critère est essentiellement bénéfique lorsqu'il est combiné avec les deux autres critères.

4 Définition de MTDS

4.1 Sélection des TDS

Cette étape a pour objectif d'identifier pour chaque objet le détecteur qui maximise la précision de son indexation basée sur sa moyenne des précisions moyennes (moyPM), figure 6(a). Cette dernière est calculée de la manière suivante : une courbe de précision rappel est créée pour chaque image requête (120K d'images de validation dont 1200 images positives et 1200 $\times$ 99 images négatives par objet) pour calculer la précision moyenne. La moyenne des précisions moyennes par objet est ensuite calculée sur les images positives de l'objet. A chaque objet est ainsi associé le TDS retenu et l'ensemble des TDS retenus constitue la MTDS. A chaque TDS sélectionné est appris également le seuil de détection qui satisfait le taux minimum de vrais positifs (TP) exigé en se basant sur la courbe ROC correspondante. La figure 6(b) donne une vue sur la répartition des points ROC des objets issus du critère fixé, $TP = 0,9$. Ce dernier dépend des exigences de l'application considérée. En outre, l'apprentissage des seuils peut être basé sur le taux de faux positifs minimum autorisé.

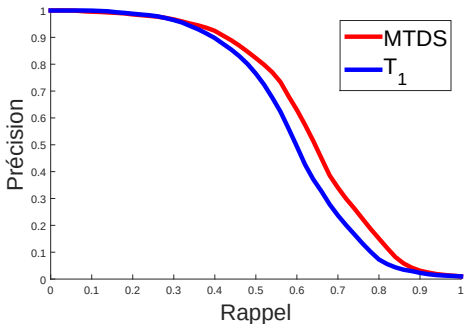


Figure 7: Courbes de la précision Moyenne/rappel relatives à l'utilisation de MTDS et de T_1 sur $D3$. L'aire sous une courbe définit MoyPM.

4.2 Evaluation de MTDS

A l'issue de cette étape de sélection, 18 TDS distincts sur 54 générés sont retenus. Ils découlent de l'application des différents critères élaborés. Cela prouve la nécessité de considérer tous les critères, laissant à l'algorithme d'optimisation la détermination empirique du seuil respectif à adopter. Nous avons évalué et comparé la performance du détecteur MTDS et celle des TDS uniques de la base sur l'ensemble des données. Pour ce faire, la moyenne et le médian des précisions moyennes sur toutes les images requêtes, MoyPM et respectivement MedPM, sont calculés aussi bien sur les données de validation $D2$ que sur les données test $D3$. La synthèse des résultats est rapportée dans le tableau 2. La performance de T_2 et T_3 étant faible sur $D2$, nous nous focaliserons désormais sur l'étude comparative des performances entre MTDS et de TDS1 sur $D3$. Une amélioration de $4,2\%$ et de $4,5\%$ avec une valeur p de 5.10^{-6} est constatée quant à la MoyPM et respectivement au MedPM de MTDS. La figure 7 illustre la différence des MoyPM et pour chaque détecteur, les histogrammes de la figure 8 donnent une vue sur la répartition des moyennes des précisions moyennes par objet moyPM sur $D2$ et sur $D3$.

	$MTDS$	T_1	T_2	T_3
MoyPM sur D2	$74,6$	$65,2$	$37,2$	$25,5$
MedPM sur D2	$76,6$	$70,1$	$33,8$	$19,9$
MoyPM sur D3	$63,1$	$58,9$	–	–
MedPM sur D3	$64,3$	$59,8$	–	–

Table 2: Synthèse statistique sur la performance des détecteurs en $\%$.

5 Résultats expérimentaux

Nous avons évalué la performance de l'indexation sur $D3$ par cas. Cas 1 (E1) : l'étape 1 ; Cas 2 (E1+SfT) : la combinaison de E1 avec l'étape 3 ; Cas 3 (E1+E2+SfT) : la

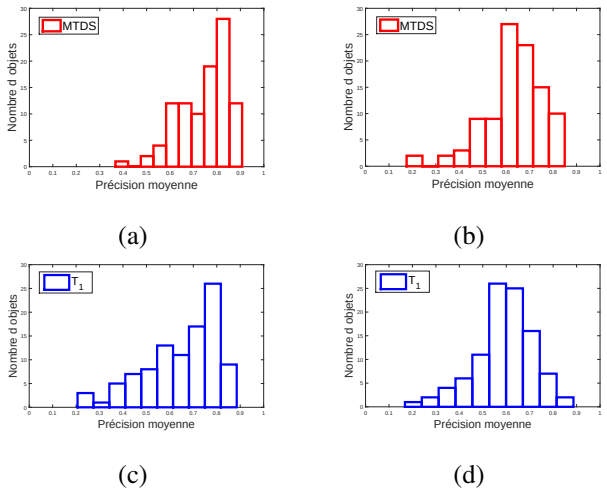


Figure 8: Histogramme du nombre d'objets en fonction de leur moyPM sur D2 (a)(c) et sur D3 (b)(d).

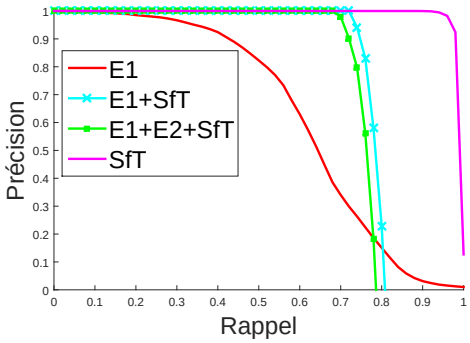


Figure 9: Courbes de la précision Moyenne/rappel relatives aux différentes étapes et combinaisons d'étapes sur $D3$, (cas 1 : E1 sans seuil).

procédure complète ; Cas 4 (SfT) : l'application directe de l'estimateur robuste de SfT i.e. E3 sur tous les objets de la base.

Pour chaque cas, les MoyPM sont représentées dans la figure 9 avec une synthèse sur la performance (MoyPM et MedPM), le coût respectif induit (temps moyen par image de 720×1280 sur une machine bi-processeur Intel Xeon E5-2670 v3) ainsi que le nombre d'objets (R) requis dans le tableau 3.

L'application directe de SfT assure la meilleure performance en terme de précision d'indexation avec néanmoins le coût le plus élevé. A l'issue de E1, le nombre d'objets individuellement détectés varie d'une image requête à une autre. Nous obtenons un nombre moyen de 15 objets sur 30K d'images requête avec une MoyPM de 81,9% à l'application de SfT et une MoyPM de 80,1% à sa combinaison avec E2 et SfT (cas 2 et cas 3). Ce qui est une bonne performance. Bien que SfT remonte la performance de E1, les rappels sont limités pour les cas 2 et 3 dû à la détection individuelle seuillée à l'étape 1. Par ailleurs,

	Moy PM (%)	**Med PM** (%)	**R**	**Temps (s)**
E1	$63,1$	$58,9$	100	$0,07$
E1+SfT	$81,9$	$81,6$	≈ 15	$0,28$
E1+E2+SfT	$80,1$	$79,8$	≤ 10	$0,20$
SfT	$99,0$	$99,3$	100	$4,92$
Modèle rigide	$87,7$	$87,7$	100	$5,04$

Table 3: Performance selon les étapes considérées.

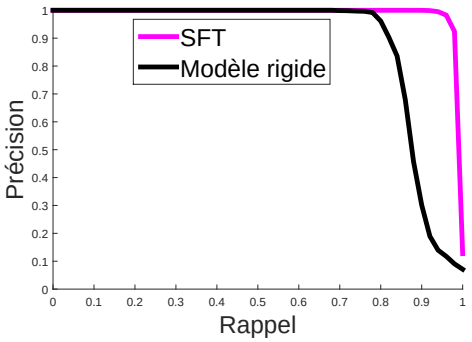


Figure 10: Courbes de la précision Moyenne/rappel obtenues par SfT et par le modèle rigide (directs) sur $D3$.

nous avons comparé SfT avec le modèle rigide classique basé sur la matrice fondamentale [9]. Une nette différence de performance est constatée visuellement dans la figure 10 que nous avons quantifiée en calculant la différence des MoyPM (tableau 3) qui est égale à $11,3\%$ avec une valeur $p = 7.10^{-30}$ en faveur du modèle SfT. En sus, nous avons couplé notre méthode d'indexation avec la reconstruction SfT en les appliquant sur une image requête réelle de deux objets, figure11(a). Les résultats de l'indexation sont présentés dans les figures 11(b)(c)(d) tandis que les résultats de la reconstruction SfT sont illustrés dans la figure 12.

6 Conclusion

Nous avons présenté une approche d'indexation pour les objets isométriques et pour l'applicabilité de la reconstruction 3D par SfT à partir d'une base de données importante. La connaissance a priori du modèle physique SfT constitue l'ossature de notre approche autour de laquelle s'articulent les méthodes de classement et de vérification spatiale mises en œuvre. D'une part, le modèle physique a été utilisé pour la vérification spatiale et s'avère être plus adapté aux objets isométriques que le modèle physique standard des objets rigides. D'autre part, le modèle physique permet de générer des données. Nous avons tiré avantage de cette connaissance a priori afin d'améliorer l'indexation classique par l'introduction de MTDS (combinaison de plusieurs sets de descripteurs de la base de données) déterminée durant une phase d'apprentissage et de validation.

(a) (b) (c) (d)

Figure 11: (a) Image requête réelle de taille 4800 × 3200. (b) Résultats issus de E1 et E2. (c)(d) Mise en correspondance au cours de E3. Temps total d'indexation $\approx 3s$.

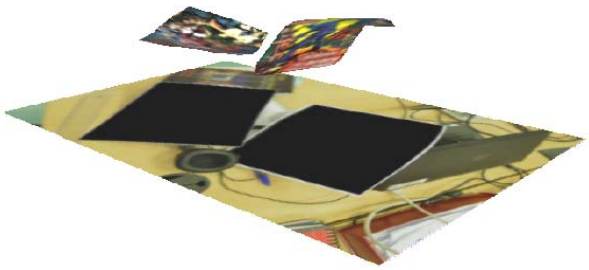


Figure 12: Reconstruction 3D par SfT des deux objets détectés par l'indexation, selon deux points de vue. Temps total de reconstruction $\approx 4s$.

Nous avons mis en évidence la contribution de MTDS dans l'amélioration des résultats en comparant notre approche avec la méthode classique n'utilisant qu'un set de descripteurs de la base de données. Notre approche d'indexation ayant pour objectif de servir SfT, a été également évaluée sur la performance de ses étapes constitutives. Ce qui permet d'avoir un aperçu des résultats attendus en terme de précision et du coût induit en fonction du nombre d'objets considérés de la base. En perspective, nous proposons d'évaluer la performance de notre approche en variant le dictionnaire puis de tester sa robustesse par rapport aux occultations et à la variation de l'éclairage.

References

[1] P. F. Alcantarilla, A. Bartoli, and A. J. Davison. KAZE features. In *ECCV*, 2012.

[2] R. Arandjelovic and A. Zisserman. Multiple queries for large scale specific object retrieval. In *BMVC*, pages 1–11, 2012.

[3] R. Arandjelović and A. Zisserman. Three things everyone should know to improve object retrieval. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2911–2918. IEEE, 2012.

[4] A. Bartoli, Y. Gérard, F. Chadebecq, T. Collins, and D. Pizarro. Shape-from-template. *IEEE Trans. Pattern Anal. Mach. Intell.*, 37(10):2099–2118, 2015.

[5] Y. Chen, X. Li, A. Dick, and A. Van den Hengel. Boosting object retrieval with group queries. *Signal Processing Letters, IEEE*, 19(11):765–768, 2012.

[6] A. Chhatkuli, D. Pizarro, and A. Bartoli. Stable template-based isometric 3d reconstruction in all imaging conditions by linear least-squares. In *Proceedings of the IEEE Con-*

ference on Computer Vision and Pattern Recognition, pages 708–715, 2014.

[7] T. Collins and A. Bartoli. Realtime shape-from-template: System and applications. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*, 2015.

[8] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(9):1627–1645, 2010.

[9] R. I. Hartley. In defense of the eight-point algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(6):580–593, Jun 1997.

[10] H. Jégou and O. Chum. Negative evidences and cooccurences in image retrieval: The benefit of pca and whitening. In *Computer Vision–ECCV 2012*, pages 774–787. Springer, 2012.

[11] H. Jégou, M. Douze, and C. Schmid. On the burstiness of visual elements. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1169–1176. IEEE, 2009.

[12] H. Jégou, M. Douze, C. Schmid, and P. Pérez. Aggregating local descriptors into a compact image representation. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 3304–3311. IEEE, 2010.

[13] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 2169–2178. IEEE, 2006.

[14] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[15] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool. A comparison of affine region detectors. *International journal of computer vision*, 65(1-2):43–72, 2005.

[16] A. Mikulik, M. Perdoch, O. Chum, and J. Matas. Learning vocabularies over a fine quantization. *International Journal of Computer Vision*, 103(1):163–175, 2013.

[17] J. O. M. Östlund, A. Varol, T. D. Ngo, and P. Fua. Laplacian Meshes for Monocular 3D Shape Recovery. In *ECCV*, 2012.

[18] M. Perriollat and A. Bartoli. A computational model of bounded developable surfaces with application to image-based three-dimensional reconstruction. *Computer Animation and Virtual Worlds*, 24(5):459–476, 2013.

[19] F. Perronnin and C. Dance. Fisher kernels on visual vocabularies for image categorization. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. IEEE, 2007.

[20] D. Pizarro and A. Bartoli. Feature-based deformable surface detection with self-occlusion reasoning. *International Journal of Computer Vision*, 97(1):54–70, 2012.

[21] S. Qi and Y. Luo. Object retrieval with image graph traversal-based re-ranking. *Signal Processing: Image Communication*, 41:101–114, 2016.

[22] D. Qin, C. Wengert, and L. Gool. Query adaptive similarity for large scale object retrieval. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1610–1617, 2013.

[23] Y. Ren, J. Benois-Pineau, and A. Bugeau. A comparative study of irregular pyramid matching in bag-of-bags of words model for image retrieval. In *Image and Signal Processing*, pages 539–548. Springer, 2014.

[24] C. Schmid and R. Mohr. Local grayvalue invariants for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(5):530–534, 1997.

[25] X. Shen, Z. Lin, J. Brandt, S. Avidan, and Y. Wu. Object retrieval and localization with spatially-constrained similarity measure and k-nn re-ranking. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 3013–3020. IEEE, 2012.

[26] J. Sivic and A. Zisserman. Video google: A text retrieval approach to object matching in videos. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 1470–1477. IEEE, 2003.

[27] X. Tian, L. Jiao, X. Liu, and X. Zhang. Feature integration of eodh and color-sift: Application to image retrieval based on codebook. *Signal Processing: Image Communication*, 29(4):530–545, 2014.

[28] G. Tolias and Y. Avrithis. Speeded-up, relaxed spatial matching. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 1653–1660. IEEE, 2011.

[29] L. Zheng, S. Wang, Z. Liu, and Q. Tian. Packing and padding: Coupled multi-index for accurate image retrieval. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1939–1946, 2014.