Vidéosurveillance intelligente : ré-identification de personnes par signature utilisant des descripteurs de points d'intérêt collectés sur des séquences

Omar HAMDOUN[1], Fabien MOUTARDE[1], Bogdan STANCIULESCU[1] et Bruno STEUX[1]

[1]Centre de Robotique (CAOR), Mines Paris Tech (ENSMP), 60 Bd St Michel, F-75006 Paris, FRANCE

Omar.Hamdoun@ensmp.fr, Fabien.Moutarde@ensmp.fr, Bogdan.Stanciulescu@ensmp.fr, Bruno.Steux@ensmp.fr

Résumé – Nous présentons et évaluons une méthode de ré-identification de personnes pour les systèmes de surveillance multi-caméras. Notre approche utilise la mise en correspondance de signatures fondées sur les descripteurs de points d'intérêt collectés sur de courtes séquences vidéos. Une des originalités de notre travail est d'accumuler les points d'intérêt à des instants suffisamment espacés durant le suivi de personne, de façon à capturer dans la signature la variabilité d'apparence des personnes. Une première évaluation expérimentale a été effectuée sur une base publique d'enregistrements à basse résolution dans un centre commercial, et les performances de re-identification sont très prometteuses (une précision de 82% pour un rappel de 78%). De plus, notre technique de ré-identification est particulièrement rapide : ~1/8 s pour une requête à comparer à 10 personnes vues précédemment, et surtout une dépendance logarithmique avec le nombre de modèles stockés, de sorte que la ré-identification parmi des milliers de personnes prendrait moins de ¼ s de calcul.

Abstract – We present and evaluate a person re-identification scheme for multi-camera surveillance system. Our approach uses matching of signatures based on interest-points descriptors collected on short video sequences. One of the originalities of our method is to accumulate interest points on several sufficiently time-spaced images during person tracking within each camera, in order to capture appearance variability. A first experimental evaluation conducted on a publicly available set of low-resolution videos in a commercial mall shows very promising inter-camera person re-identification performances (a precision of 82% for a recall of 78%). It should also be noted that our matching method is very fast: ~ 1/8s for re-identification of one target person among 10 previously seen persons, and a logarithmic dependence with the number of stored person models, making re-identification among thousands of persons computationally feasible in less than ~ 1/4s second.

1. Introduction

Dans de nombreuses applications de surveillance, il est souhaitable de déterminer si une personne a déjà été observée par un réseau de caméras. Cela définit le problème de la ré-identification des personnes (voir par exemple [1] pour une présentation générale). Les algorithmes de ré-identification doivent traiter plusieurs situations telles que : les différents angles de caméra, des conditions variées d'éclairage, les variations de pose des personnes, et l'évolution rapide de l'apparence des vêtements. Une première catégorie de méthodes utilise des techniques biométriques (par exemple reconnaissance de visage ou de démarche), mais on ne s'intéressera ici qu'à la seconde catégorie qui exploite l'apparence globale. Parmi celles-ci, diverses approches ont été proposées, par exemple utilisant des histogrammes de couleurs [2], ou des caractéristiques de texture [3], ou enfin la mise en correspondance de points d'intérêt [4].

Nous proposons ici une méthode d'identification de personne utilisant l'appariement des points d'intérêt trouvés dans plusieurs images. Le point central de notre algorithme réside dans l'exploitation de séquences d'images, qui permet de disposer d'informations supplémentaires (aspect 3D, dynamique, ..) par rapport à l'utilisation d'une image seule. Les informations extraites des séquences sont par la suite intégrées dans un modèle caractérisant l'objet. Pour réaliser la détection des points d'intérêt et pour calculer les descripteurs, nous utilisons des fonctions de la bibliothèque de traitement d'images Camellia, développée au sein du laboratoire (http://camellia.sourceforge.net). Ces fonctions, qui seront présentées ailleurs, implémentent une variante inspirée de SURF [5] et encore plus efficace. SURF lui-même est un algorithme très rapide, inspiré des plus classiques et plus couramment utilisés détecteur et descripteur de points d'intérêt SIFT [6].

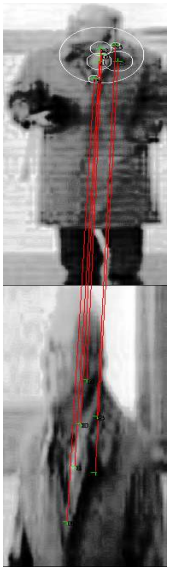

Figure 1 : Exemple d'appariement de points d'intérêt sur une même personne, vue sous 2 angles différents et à 2 échelles différentes.

2. Schéma algorithmique

Dans cette section nous détaillons les choix algorithmiques pour la construction du système de reconnaissance. Cet algorithme suit le schéma classique des algorithmes de DRI (Détection- Reconnaissance - Identification), et peut être séparé en 2 étapes : une étape d'apprentissage (figure 2a), et une étape de reconnaissance (figure 2b). L'étape d'apprentissage vise à détecter et suivre l'individu dans la séquence, pour extraire les points d'intérêt nécessaires à la construction du modèle, en utilisant une caméra. L'étape de reconnaissance exploite les modèles issus de l'étape d'apprentissage pour déterminer s'il s'agit du même individu dans une autre caméra.

1. Etape de la construction du modèle : Un modèle est construit pour chaque individu détecté et suivi dans la séquence. Pour augmenter la quantité d'informations discriminantes nous n'utilisons pas toutes les trames d'images successives, mais des trames espacées d'une demi-seconde (donc généralement une sur dix), et nous accumulons dans le modèle les points d'intérêt de ces différentes images de la même personne.

2. Étape de construction de la requête : la requête est construite par la même méthode que le modèle, mais à partir d'images issues d'une deuxième caméra.

3. Comparaison des descripteurs : la mesure de la ressemblance entre les différents descripteurs est la SAD (Sum of Absolute Differences).

4. Appariement robuste : pour effectuer un appariement robuste et surtout rapide, la fonction de Camellia que nous utilisons implémente l'algorithme de recherche BBF (Best Bin First) dans un K-D arbre [7] contenant l'ensemble des modèles.

5. Identification : la recherche de la requête dans l'ensemble des modèles s'effectue à l'aide de la technique de vote. Chaque point extrait sur la requête est comparé à l'ensemble de points des modèles enregistrés dans un K-D arbre. Un vote est ajouté à l'entrée d'une table de votes si la distance absolue entre le point d'un individu et un point de l'arbre est inférieur à un seuil donné (0.8*deuxième distance inférieure). On obtient une table de votes où les meilleurs scores correspondent aux modèles les plus similaires à l'individu.

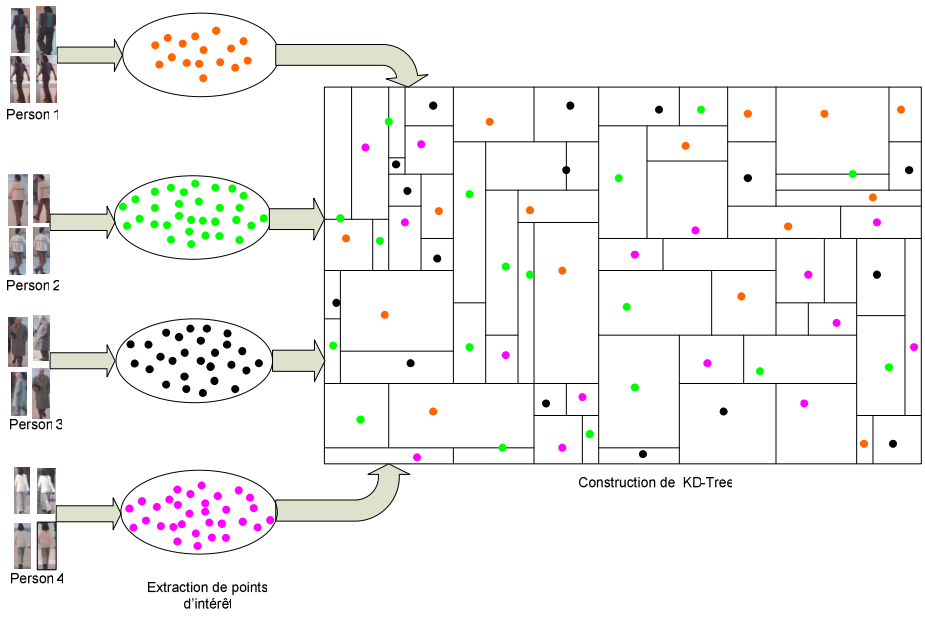

Figure 2a : Vue schématique de la construction du modèle

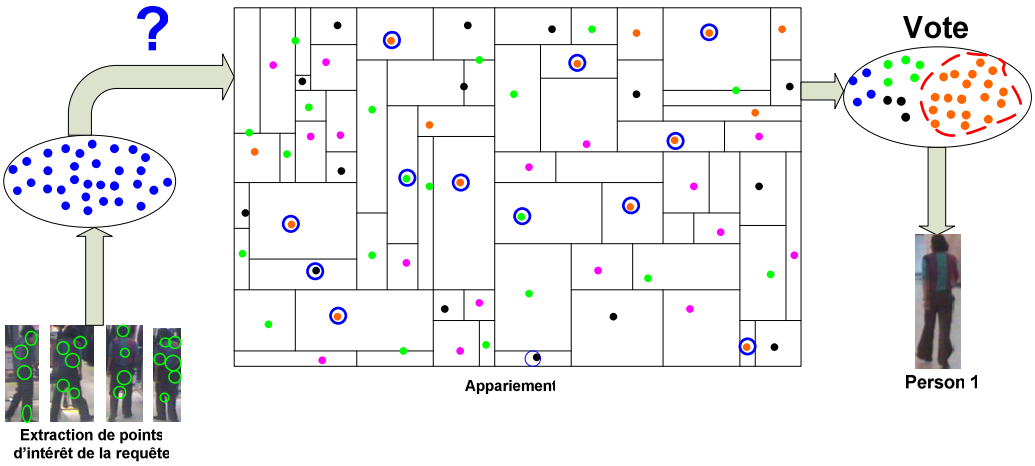

Figure 2b : Vue schématique de la ré-identification d'une requête

3. Validation expérimentale

L'algorithme proposé a fait l'objet de premiers tests sur des séquences d'assez faible résolution, avec des personnes évoluant dans un centre commercial (base publique issue du projet européen CAVIAR **[**IST 2001 37540], ***http://homepages.inf.ed.ac.uk/rbf/CAVIAR/***). La reconnaissance s'est effectuée pour deux ensembles d'images prises par deux caméras avec des angles de vue différents, présentés dans la figure 3. Ils correspondent aux mêmes 10 personnes, avec 21 imagettes extraites de chaque séquence-modèle, et 6 imagettes extraites de chaque séquence-requête. La variabilité potentielle des couleurs entre caméra est évitée en travaillant sur les images en niveau de gris. L'invariance aux conditions d'éclairage est résolue par l'égalisation d'histogramme de toutes les images. Les deux groupes de personnes correspondant aux deux caméras sont comparés afin d'évaluer la performance de l'algorithme proposé.

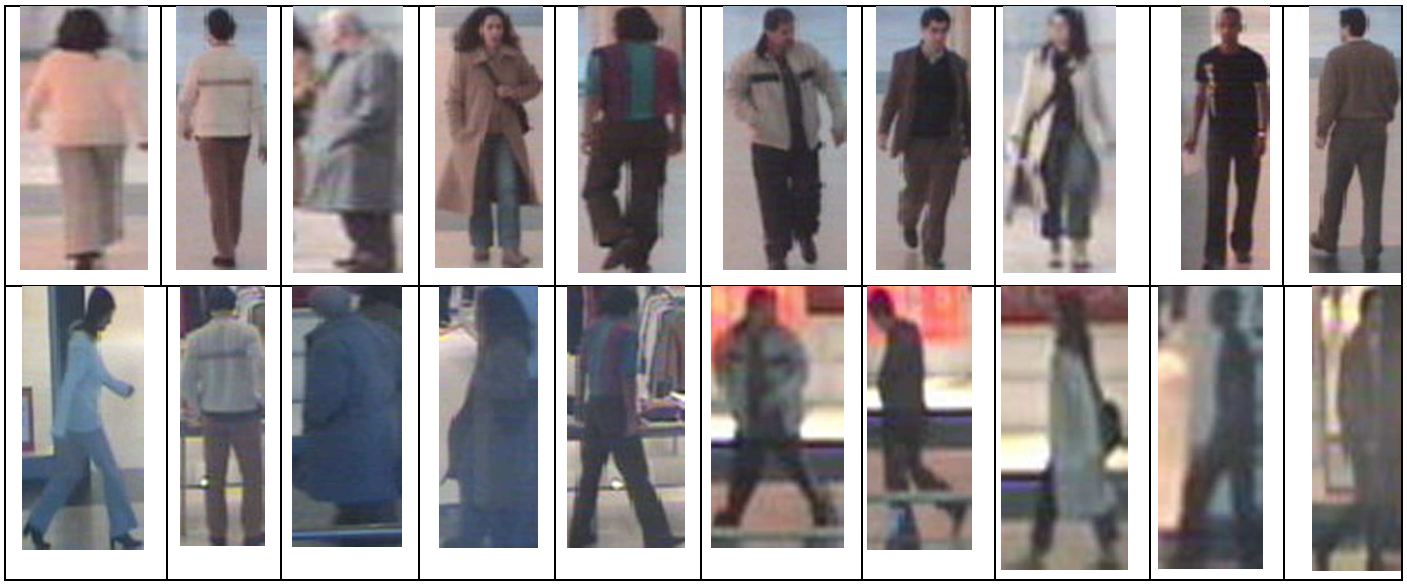

Figure 3: Exemples typiques de vues de personnes utilisées pour constituer les modèles (ligne du haut), et exemples correspondant de vues des mêmes personnes avec l'autre caméra à partir de laquelle on tente de ré-identifier.

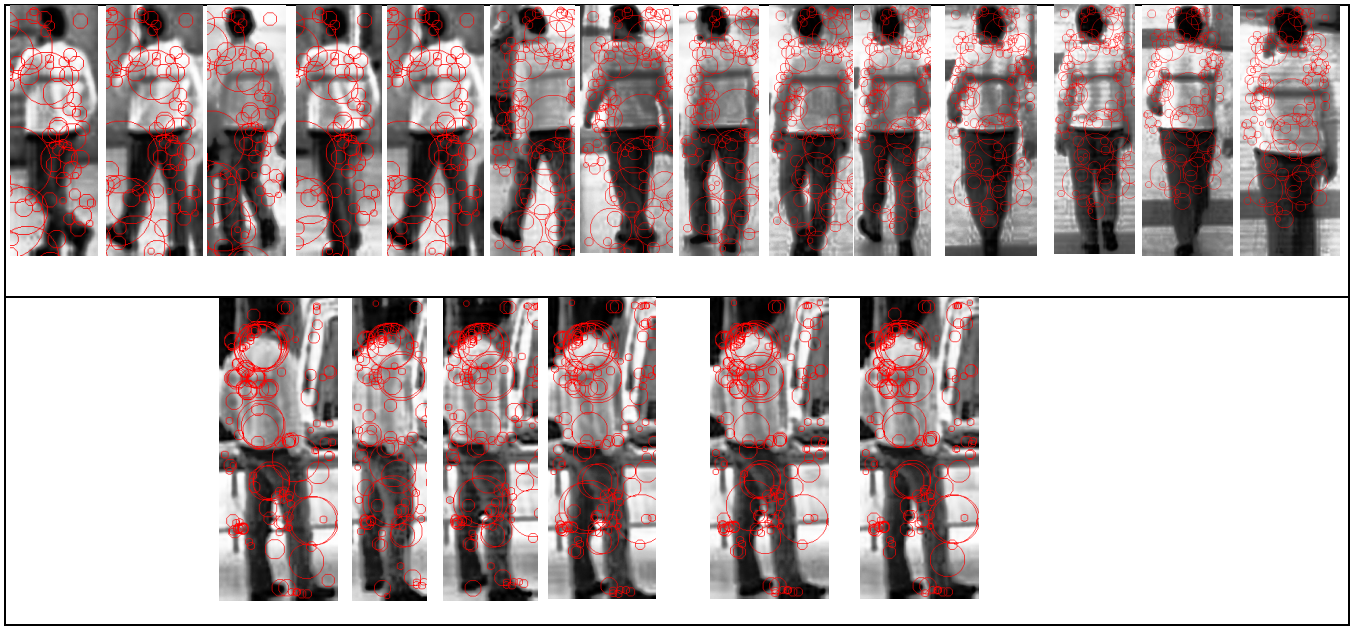

Figure 4 : Visualisation des points d'intérêt détectés sur 14 des 21 images utilisées pour le modèle d'une personnel (ligne du haut), et sur les 6 images d'une requête correctement appariée pour la même personne (ligne du bas).

Nous utilisons les métriques PR (précision et rappel) pour évaluer les performances :

$$\text{Pr}\,écision = \frac{TP}{TP + FP}$$

$$Rappel = \frac{TP}{Nombre\ de\ requêtes}$$

avec TP (True Positives) = nombre de bons appariements requête-modèle, et FP (False Positives) = nombre d'appariements erronés.

Les résultats de reconnaissance, calculés sur 760 requêtes, sont présentés dans la table 1 et sur la figure 5. Le principal paramètre de réglage est le « seuil du score d'appariement », i.e. le nombre minimum de points à apparier entre une requête et un modèle pour valider une ré-identification. Tout d'abord on constate, comme attendu, que plus on augmente le seuil du score d'appariement, plus la précision augmente, mais au détriment du rappel qui diminue.

Compte tenu de la faible résolution des images, les performances obtenues de ré-identification de personnes sont bonnes, avec par exemple 82% de précision et 78% de rappel quand on règle le seuil de score à 15 points.

Table 1 : Précision et rappel, en fonction du seuil adopté pour le score d'appariement entre requête et modèle (i.e. nombre minimum de points similaires).

Seuil du score d'appariement entre requête et modèle (nombre de points appariés)	Précision (%)	Rappel (%)
40	**99**	**49**
35	**97**	**56**
30	**95**	**64**
25	**90**	**71**
20	**85**	**75**
15	**82**	**78**
10	**80**	**79**
5	**80**	**80**

Il faut noter aussi la grande rapidité d'exécution de notre algorithme : moins de 1/8 s de calcul par requête, ce qui est négligeable devant les 3 secondes nécessaires pour collecter les 6 imagettes séparées de ½ s constituant la requête. Mieux encore, grâce à la complexité logarithmique, par rapport au nombre de descripteurs, de la recherche dans le KD-arbre, le temps de traitement d'une requête devrait rester très faible même si un très grand nombre de modèles de personnes était stocké. Pour vérifier cela, nous avons étudié le temps de ré-identification quand on fait varier le nombre d'imagettes utilisées dans chaque "séquence-modèle", comme le rapporte la table 2. Et effectivement, la figure 6 montre bien que le temps de ré-identification ne croît que logarithmiquement avec le nombre de descripteurs stockés.

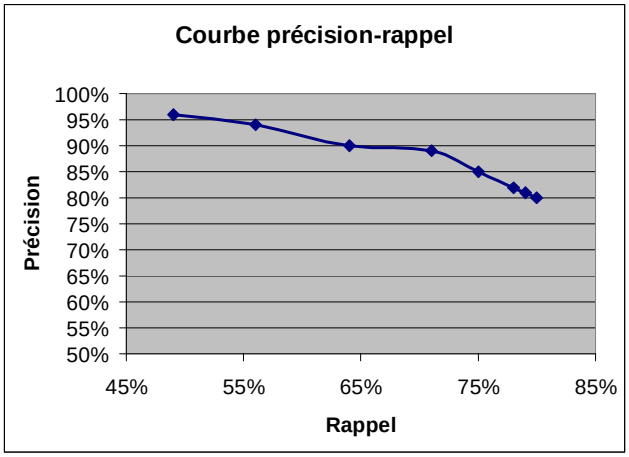

Figure 5 : Courbe précision-rappel pour la première évaluation expérimentale de notre méthode de ré-identification

Comme le nombre de descripteurs est approximativement proportionnel au nombre d'images utilisées dans les modèles, si les signatures de 1000 à 10000 personnes au lieu de 10 étaient stockées (avec ~20 images pour chacune), le KD-arbre contiendrait 100 à 1000 fois plus de points d'intérêt, soit ~ 2,5 à 25 millions de descripteurs. En extrapolant à partir de la figure 6, on s'attend donc à un temps de calcul de ~1/5 à ¼ s pour une requête de ré-identification parmi des milliers de personnes déjà vues et aux signatures pré-calculées et stockées.

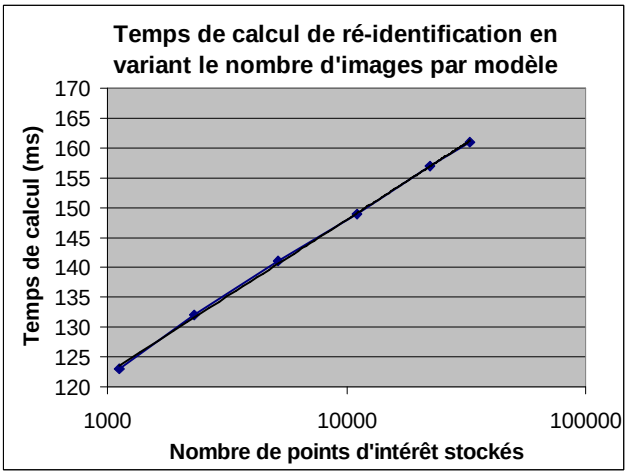

Figure 6 : Temps de calcul pour la ré-identification en fonction du nombre de descripteurs stockés ; la dépendance est clairement logarithmique

Table 2 : Nombre total de points d'intérêt, et temps de calcul de la ré-identification, en fonction du nombre d'images utilisées pour le modèle de chaque personne.

Nombre d'images utilisées pour chaque sequence-modèle	Nombre total de points d'intérêt stockés	Temps de calcul pour la ré-identification (ms)
1	**1117**	**123**
2	**2315**	**132**
4	**5154**	**141**
8	**11100**	**149**
16	**22419**	**157**
24	**32854**	**161**

4. Conclusions et perspectives

Nous avons présenté un algorithme de ré-identification utilisant la mise en correspondance de points d'intérêt collectés dans des séquences-requêtes et des séquences-modèles.

Nos expérimentations sur des vidéos de faible résolution ont montré de premiers résultats très prometteurs de notre méthode pour la ré-identification de personnes entre différentes caméras : une précision de 82% pour un rappel de 78%. Il faut noter que notre méthode d'appariement est très rapide, avec un temps de calcul typique de 1/8s pour la ré-identification d'une personne-cible parmi 10 signatures de personnes vues précédemment. De plus, ce temps ne croît que logarithmiquement avec le nombre de modèles stockés, de sorte que le temps de calcul resterait inférieur à ¼ seconde pour un système de taille réelle, dans lequel il faudrait potentiellement pouvoir ré-identifier parmi des milliers de personnes suivies.

Des évaluations plus approfondies et sur d'autres corpus, notamment avec plus de personnes, doivent cependant encore être effectuées. Aussi, notre algorithme de ré-identification sera prochainement intégré dans la chaîne globale de video-surveillance, ce qui permettra de restreindre les points d'intérêt à l'intérieur des silhouettes des personnes, et donc d'exclure les points "parasites" correspondant au fond derrière les personnes, ce qui devrait améliorer significativement les performances de ré-identification de notre système.

Enfin, nous espérons pouvoir améliorer encore les performances en raffinant le modèle utilisé, soit en exploitant les positions relatives des points d'intérêt, soit en appliquant une technique d'apprentissage artificiel sur les modèles construits.

Références

[1] Tu, P.; Doretto, G.; Krahnstoever, N.; Perera, A.; Wheeler, F.; Liu, X.; Rittscher, J.; Sebastian, T.; Yu, T. & Harding, K. “An intelligent video framework for homeland protection”, *Proceedings of SPIE Defence and Security Symposium - Unattended Ground, Sea, and Air Sensor Technologies and Applications IX*, Orlando, FL, USA, April 9--13, 2007.

[2] Park, U.; Jain, A.; Kitahara, I.; Kogure, K. & Hagita, N., “ViSE: Visual Search Engine Using Multiple Networked Cameras”, *Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06)-Volume 03*, 1204-1207 (2006).

[3] Lantagne, M.; Parizeau, M. & Bergevin, R., “VIP : Vision tool for comparing Images of People”, *Proceedings of the 16th IEEE Conf. on Vision Interface*, pp. 35-42, 2003.

[4] Gheissari, N.; Sebastian, T. & Hartley, R., “Person Reidentification Using Spatiotemporal Appearance”, *Proceedings of 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’2006)*-Volume 2, IEEE Computer Society, pp. 1528-1535, New-York, USA, June 17-22, 2006.

[5] Herbert Bay, Tinnr Tuytelaars & Gool, L. V., “SURF:Speeded Up Robust Features”, *Proceedings of the 9th European Conference on Computer Vision (ECCV’2006)*, Springer LNCS volume 3951, part 1, pp 404--417, 2006

[6] Lowe, D., “Distinctive Image Features from Scale-Invariant Keypoints”, *International Journal of Computer Vision, Vol. 60*, pp. 91-110, *Springer*, 2004

[7] Beis, J. & Lowe, D., ” Shape indexing using approximate nearest-neighbour search in high-dimensional spaces”, In *Proc. 1997 IEEE Conf. on Computer Vision and Pattern Recognition*, pages 1000-1006, Puerto Rico, 1997