

Détection robuste d'objets : une approche par modèle d'apparence

Rozenn DAHYOT¹, Pierre CHARBONNIER¹, Fabrice HEITZ²

¹Laboratoire Régional des Ponts-et-Chaussées de Strasbourg
11 Rue Jean Mentelin, BP 9, F-67035 STRASBOURG cedex

²LSIIT UPRESA CNRS 7005 Université Strasbourg I
Boulevard Sébastien Brandt F-67400 ILLKIRCH

Rozenn.Dahyot@equipement.gouv.fr, Pierre.Charbonnier@equipement.gouv.fr
heitz@lsiit.u-strasbg.fr

Résumé – Les méthodes classiques de détection basées sur la représentation de l'apparence par espace propre sont sensibles à la présence d'erreurs grossières dans les observations, induites, par exemple, par des occultations. Récemment, l'utilisation de techniques issues des statistiques robustes, les M-estimateurs, ont permis de gérer la présence de ces données erronées dans le cadre de la reconnaissance d'objets. Nous proposons dans cet article d'étendre cette approche robuste pour définir deux nouveaux *détecteurs*, capables de localiser les occurrences dégradées ou occultées d'objets d'intérêt dans des scènes texturées.

Abstract – Standard detection methods using eigenspace representation of an object class, are sensitive to gross errors due to, for example, partial occultations. Recently, robust statistical technics able to deal with those *outliers*, the so called M-estimators, were successfully used for recognition. We propose to extend their use to define robust *detectors* able to localize partially occluded objects in cluttered scenes.

1 Introduction

La représentation des objets par leur apparence a récemment donné lieu à de nombreux travaux en détection et reconnaissance [5, 7, 6]. Une des approches les plus employées est celle utilisant la représentation par espace propre. Celle-ci permet une réduction substantielle de la dimension du problème en capturant les principales variabilités d'une population d'images d'apprentissage. Les méthodes classiques de détection et de reconnaissance, basées sur cette représentation [5, 6], sont sensibles à la présence de données erronées ou *outliers*, induites par exemple, par des occultations partielles. Les M-estimateurs, proposés par Huber [4] dans le cadre de l'estimation statistique, ont récemment permis de définir des méthodes robustes de reconnaissance [1, 3]. Cet article propose d'étendre leur utilisation au problème de la détection d'une classe d'objets, éventuellement occultés, dans des scènes texturées.

Après un rappel sur la représentation par espace propre à la section 2, nous proposons, au paragraphe 3, une formulation probabiliste prenant en compte les occurrences dégradées, occultées et apparaissant sur des fonds texturés. Les vraisemblances proposées sont maximisées grâce à deux algorithmes issus de la théorie semi-quadratique, présentés à la section 4 [2]. Le premier correspondant à des moindres carrés pondérés itératifs, est classiquement utilisé pour la reconnaissance robuste [1, 3]. Le second algorithme est moins coûteux car il ne nécessite pas d'inversion de matrice.

À la section 5, nous proposons d'utiliser cette vraisemblance robuste, comme détecteur de nos objets d'intérêt.

Un deuxième détecteur est présenté, utilisant une des variables auxiliaires introduite par la théorie semi-quadratique. Les détecteurs classiques et robustes sont comparés au paragraphe 6, sur des images en niveaux de gris.

2 Représentation par espace propre et détection d'objets d'intérêt

On considère une base de K images d'apprentissage centrées sur leur moyenne μ : $\mathcal{B} = \{\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_K\}$. Chaque image $\tilde{\mathbf{x}}$, composée de N pixels, peut se reconstruire sur une base de J vecteurs orthogonaux \mathbf{u}_j , calculés grâce à l'Analyse en Composantes Principales :

$$\tilde{\mathbf{x}} = \sum_{j=1}^J c_j \mathbf{u}_j + \mathbf{w}^r \quad (1)$$

où \mathbf{w}^r représente l'erreur entre $\tilde{\mathbf{x}}$ et sa reconstruction sur l'espace propre F défini par les vecteurs propres \mathbf{u}_j . Les coordonnées \mathbf{c} sont estimées au sens des moindres carrés, par minimisation de la norme de \mathbf{w}^r , la solution se ramenant à une projection orthogonale: $\mathbf{c} = \mathbf{U}^T \tilde{\mathbf{x}}$. \mathbf{U} est la matrice collectant sur chacune de ses colonnes, les J vecteurs propres \mathbf{u}_j .

Pour détecter un objet d'intérêt dans une image I , on extrait en chaque position (i, j) une sous-image \mathbf{y} . On calcule la ressemblance de \mathbf{y} avec la classe des images d'apprentissage. Une carte collecte l'ensemble des valeurs obtenues, et les positions présentant les plus fortes valeurs de similarité sont les positions des objets d'intérêt dans l'image I .

Un critère classique de ressemblance est la distance euclidienne à l'espace propre, ou *Distance From Feature Space* (DFFS), définie par [5] :

$$DFFS(\mathbf{y}) = \|\tilde{\mathbf{y}} - \mathbf{U}\mathbf{U}^T\tilde{\mathbf{y}}\|^2 \quad (2)$$

Un autre critère de ressemblance consiste à estimer la probabilité que l'observation \mathbf{y} soit une instance d'un des objets de la base \mathcal{B} , notée $\mathcal{P}(\mathbf{y}|\mathcal{B})$. En la supposant gaussienne, Moghaddam et Pentland l'approchent par le produit de deux probabilités marginales : la distribution de \mathbf{y} restreint à F , $\mathcal{P}_F(\mathbf{y}|\mathcal{B})$, dépendant des coordonnées $\mathbf{c} = \mathbf{U}^T\tilde{\mathbf{y}}$, et la distribution de \mathbf{y} restreint à F^\perp , $\mathcal{P}_{F^\perp}(\mathbf{y}|\mathcal{B})$, exprimée en fonction de la DFBS de \mathbf{y} , soit :

$$\begin{cases} \mathcal{P}_F(\mathbf{y}|\mathcal{B}) = \prod_{j=1}^J \frac{1}{\sqrt{2\pi\lambda_j}} \exp\left[-\frac{c_j^2}{2\lambda_j}\right] \\ \mathcal{P}_{F^\perp}(\mathbf{y}|\mathcal{B}) = \frac{1}{\sqrt{2\pi} \sigma^{N-J}} \exp\left[-\frac{DFFS(\mathbf{y})}{2\sigma}\right] \end{cases} \quad (3)$$

λ_j est la valeur propre associée au vecteur propre \mathbf{u}_j , et la variance résiduelle σ est estimée par la moyenne des valeurs propres non retenues $\{\lambda_j\}_{j \in \{J+1, \dots, N\}}$ [6].

3 Vraisemblance robuste

L'hypothèse gaussienne sur le résiduel n'est pas vérifiée lorsqu'on considère des occurrences dégradées, par exemple, par des occultations. C'est pourquoi, nous considérons que l'observation \mathbf{y} représente une réalisation bruitée d'une image de \mathcal{B} :

$$\tilde{\mathbf{y}} = \tilde{\mathbf{x}} + \mathbf{w}^o = \sum_{j=1}^J c_j \mathbf{u}_j + \mathbf{w} \quad (4)$$

où le bruit \mathbf{w} correspond cette fois ci, à l'addition de l'erreur de reconstruction \mathbf{w}^r et de l'erreur produite par d'éventuelles occultations \mathbf{w}^o . Pour modéliser cette distribution non gaussienne, nous proposons d'utiliser un modèle robuste qui prend en compte l'apparition de données erronées (M-estimation) :

$$P(\mathbf{w}|\mathcal{B}) \propto \prod_{n=1}^N \exp -\rho\left(\frac{w_n}{\sigma_\rho}\right) \quad (5)$$

où ρ est une fonction non quadratique, voire non convexe [4, 1, 3].

La vraisemblance conjointe de l'observation \mathbf{y} et de ses coordonnées \mathbf{c} dans l'espace propre s'écrit :

$$\begin{aligned} \mathcal{P}(\mathbf{y}, \mathbf{c}|\mathcal{B}) &= \mathcal{P}(\mathbf{y}|\mathbf{c}, \mathcal{B}) \mathcal{P}(\mathbf{c}|\mathcal{B}) \\ &= \mathcal{P}(\mathbf{w}|\mathcal{B}) \mathcal{P}(\mathbf{c}|\mathcal{B}) \end{aligned} \quad (6)$$

La vraisemblance $\mathcal{P}(\mathbf{y}|\mathcal{B})$, peut être calculée en intégrant $\mathcal{P}(\mathbf{y}, \mathbf{c}|\mathcal{B})$ par rapport à \mathbf{c} , pour servir de détecteur. Son calcul peut être complexe. Aussi, avons-nous choisi d'utiliser comme détecteur : $\max_{\mathbf{c}} \mathcal{P}(\mathbf{y}, \mathbf{c}|\mathcal{B})$. L'estimateur du Maximum a Posteriori $\hat{\mathbf{c}}_{MAP}$ qui maximise $\mathcal{P}(\mathbf{y}, \mathbf{c}|\mathcal{B})$, minimise l'énergie :

$$\mathcal{J}(\mathbf{c}) = -\log \mathcal{P}(\mathbf{w}|\mathcal{B}) - \log \mathcal{P}(\mathbf{c}|\mathcal{B})$$

Cette énergie est non quadratique lorsque l'on considère la distribution robuste du bruit. Deux algorithmes de minimisation de \mathcal{J} sont présentés à la section suivante.

4 Algorithmes d'estimation

Pour simplifier, on considère le cas $\mathcal{P}(\mathbf{c}|\mathcal{B}) = cste$. L'énergie à minimiser s'écrit à une constante additive près :

$$\mathcal{J}(\mathbf{c}) = \sum_{n=1}^N \rho\left(\frac{w_n}{\sigma_\rho}\right) \quad (7)$$

Cette énergie est minimisée grâce à la théorie semi-quadratique, qui introduit une nouvelle fonction d'énergie augmentée \mathcal{J}^* , ayant le même minimum global :

$$\mathcal{J}(\mathbf{c}) = \min_{\mathbf{b}} \left\{ \mathcal{J}^*(\mathbf{c}, \mathbf{b}) = \sum_{n=1}^N \rho^*\left(\frac{w_n}{\sigma_\rho}, b_n\right) \right\}$$

L'énergie augmentée, \mathcal{J}^* , est minimisée itérativement en \mathbf{c} et en \mathbf{b} par des techniques linéaires [2]. Deux formes d'énergie augmentée ont été proposées conduisant à deux algorithmes de minimisation différents.

4.1 ARTUR

La première forme d'énergie augmentée est :

$$\mathcal{J}^A(\mathbf{c}, \mathbf{b}^A) = \sum_{n=1}^N b_n^A \left(\frac{w_n}{\sigma_\rho}\right)^2 + \Psi(b_n^A)$$

Cette expression a déjà été mise en œuvre dans le cadre de la reconnaissance robuste [1, 3], et conduit à un algorithme de moindres carrés pondérés itératifs qui, à l'étape (m), s'écrit :

$$\begin{cases} \mathbf{w}^{(m)} = \tilde{\mathbf{y}} - \mathbf{U} \mathbf{c}^{(m)} \\ \forall n \in \{1 \dots N\} \quad b_n^{A(m+1)} = \frac{\rho'\left(\frac{w_n^{(m)}}{\sigma_\rho}\right)}{2 \frac{w_n^{(m)}}{\sigma_\rho}} \\ \mathbf{c}^{(m+1)} = (\mathbf{U}^T \cdot \mathbf{B}^{A(m+1)} \cdot \mathbf{U})^{-1} \cdot \mathbf{U}^T \cdot \mathbf{B}^{A(m+1)} \cdot \tilde{\mathbf{y}} \end{cases} \quad (8)$$

où $\mathbf{B}^{A(m+1)}$ est la matrice diagonale qui collecte les pondérations $b_n^{A(m+1)}$. Cet algorithme nécessite le calcul et l'inversion d'une matrice à chaque étape et demeure donc relativement coûteux en temps de calcul.

4.2 LEGEND

La seconde forme d'énergie augmentée proposée par la théorie semi-quadratique est :

$$\mathcal{J}^L(\mathbf{c}, \mathbf{b}^L) = \sum_{n=1}^N \left(\frac{w_n}{\sigma_\rho} - b_n^L\right)^2 + \xi(b_n^L) \quad (9)$$

Cette expression conduit à un algorithme de moindres carrés avec des résidus modifiés :

$$\begin{cases} \mathbf{w}^{(m)} = \tilde{\mathbf{y}} - \mathbf{U} \mathbf{c}^{(m)} \\ \forall n \in \{1 \dots N\}, \quad b_n^{L(m+1)} = w_n \left(1 - \frac{\rho'\left(\frac{w_n^{(m)}}{\sigma_\rho}\right)}{2 \frac{w_n^{(m)}}{\sigma_\rho}}\right) \\ \mathbf{c}^{(m+1)} = \mathbf{U}^T \cdot (\tilde{\mathbf{y}} - \mathbf{b}^{L(m+1)}) \end{cases} \quad (10)$$

Ces deux algorithmes sont proches de ceux proposés par Huber [4]. On peut montrer que chaque étape de LEGEND fait moins décroître l'énergie qu'une étape d'ARTUR, ce qui le rend plus lent à converger. Par contre, chaque étape correspond à une simple projection, sans inversion de matrice. Globalement, l'algorithme LEGEND est beaucoup moins coûteux en temps de calcul.

4.3 Cas gaussien

On a supposé dans le paragraphe précédent, que la distribution $\mathcal{P}(\mathbf{c}|\mathcal{B})$ était uniforme. Dans ce cas, l'énergie minimisée à l'équation (7), peut être interprétée comme une extension robuste de la DFFS. Par analogie avec la vraisemblance gaussienne utilisée par Moghaddam et Pentland, on propose une extension des algorithmes précédents à l'hypothèse de distribution gaussienne dans le plan propre :

$$\mathcal{P}(c_1, \dots, c_J | \mathcal{B}) = \prod_{j=1}^J \frac{1}{\sqrt{2\pi\lambda_j}} \exp \left[-\frac{c_j^2}{2\lambda_j} \right]$$

L'expression (7) devient :

$$\mathcal{J}(\mathbf{c}) = \sum_{n=1}^N \rho \left(\frac{w_n}{\sigma_\rho} \right) + \sum_{j=1}^J \frac{c_j^2}{2\lambda_j} \quad (11)$$

Les algorithmes de minimisation de l'énergie \mathcal{J} restent similaires à (8) et (10), mise à part l'étape de calcul de $\mathbf{c}^{(m+1)}$, qui devient pour ARTUR :

$$\mathbf{c}^{(m+1)} = (\mathbf{U}^T \cdot \mathbf{B}^{A(m+1)} \cdot \mathbf{U} + \frac{\sigma_\rho^2}{2} \mathbf{\Lambda}^{-1})^{-1} \cdot \mathbf{U}^T \cdot \mathbf{B}^{A(m+1)} \cdot \tilde{\mathbf{y}}$$

et pour LEGEND :

$$\mathbf{c}^{(m+1)} = (\mathbf{I}_J + \frac{\sigma_\rho^2}{2} \mathbf{\Lambda}^{-1})^{-1} \mathbf{U}^T \cdot (\tilde{\mathbf{y}} - \mathbf{b}^{L(m+1)})$$

où \mathbf{I}_J désigne la matrice identité de dimension $J \times J$ et $\mathbf{\Lambda}$ est la matrice diagonale collectant les valeurs propres $\{\lambda_j\}_{j \in \{1, \dots, J\}}$.

5 Détecteurs

Au paragraphe 4, nous avons présenté les algorithmes qui permettent de calculer l'estimée $\hat{\mathbf{c}}_{MAP}$ qui maximise la vraisemblance conjointe $\mathcal{P}(\mathbf{y}, \mathbf{c}|\mathcal{B})$. Comme premier détecteur robuste, on propose d'utiliser la vraisemblance :

$$\max_{\mathbf{c}} \mathcal{P}(\mathbf{y}, \mathbf{c}|\mathcal{B}) = \mathcal{P}(\mathbf{y}, \hat{\mathbf{c}}_{MAP}|\mathcal{B})$$

La pondération b_n^A , introduite dans l'algorithme ARTUR, est proche de 1 lorsque l'erreur w_n sur le pixel n , est petite, et tend vers 0 pour les forts résidus. Ainsi, l'image \mathbf{b}^A peut être interprétée comme une carte de confiance aux données. Comme second détecteur, nous proposons d'utiliser sa norme comme une mesure de similarité :

$$\|\mathbf{b}^A\|^2 = \sum_{n=1}^N (b_n^A)^2$$

Ces détecteurs sont calculés pour la fonction GM (de Geman et McClure), présentée au tableau 1. Cette fonction n'étant pas convexe, l'estimée $\hat{\mathbf{c}}_{MAP}$ est calculée en

continuation, en utilisant successivement les fonctions HS (hyper-surfaces), HL (Hebert et Leahy), et GM [3]. Le paramètre d'échelle, σ_ρ , est estimé *a priori* grâce aux images d'apprentissage [3].

TAB. 1: Fonctions robustes

	$\rho(x)$	convexité
HS	$2\sqrt{1+x^2}-2$	convexe
HL	$\log(1+x^2)$	non convexe
GM	$\frac{x^2}{1+x^2}$	non convexe

6 Résultats expérimentaux

Afin de tester nos deux mesures robustes, nous avons comparé leurs performances avec les mesures classiques. Nos images d'apprentissage sont les 72 images (64×64 pixels) du premier objet (le canard) de la base COIL [8]. Quelques images sont présentées à la figure 1. L'objet d'intérêt apparaît sur un fond uniforme. De manière à réduire l'influence du fond lors de l'apprentissage, nous proposons de calculer la moyenne μ des images, en ne considérant que les pixels qui appartiennent aux objets.

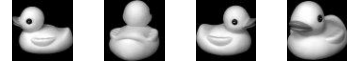
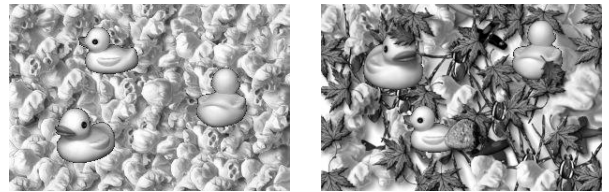


FIG. 1: Quelques images d'apprentissage.

Le test de détection a été réalisé sur 21 images présentant 57 occurrences de l'objet d'intérêt, apparaissant sur des fond texturés, et éventuellement partiellement occultées (cf. figure 2). Le nombre de vecteurs propres retenus est $J = 10$.



I_{12}

I_{15}

FIG. 2: Exemple d'images testées

Dans ce premier test, la distribution *a priori* de \mathbf{c} dans F est supposée uniforme : $\mathcal{P}(\mathbf{c}|\mathcal{B}) = cste$. Les figures 3, 4 et 5, présentent les cartes de détection calculées respectivement avec la DFFS, la vraisemblance robuste $\mathcal{P}(\mathbf{y}, \hat{\mathbf{c}}_{MAP}|\mathcal{B})$ et la norme des pondérations.

Le contour bleu des cartes indique les positions où l'on ne peut pas extraire un échantillon complet. Les positions en pixels sombres indiquent la présence éventuelle d'un objet. On remarque, notamment pour l'image I_{12} , que les détecteurs robustes permettent de mieux localiser les occurrences des objets d'intérêt.

Les performances des différents détecteurs sont évaluées à l'aide des courbes COR à la figure 6. Le pourcentage

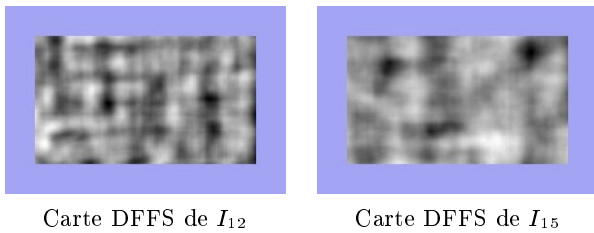


FIG. 3: Carte de détection obtenues avec la méthode classique DFFS de I_{12} et I_{15} .

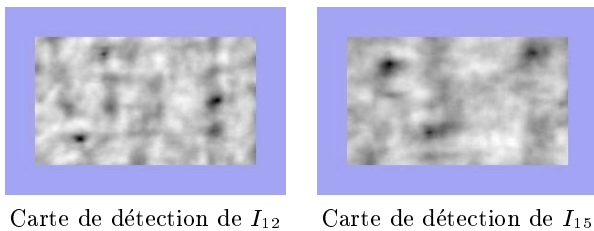


FIG. 4: Cartes de vraisemblances robustes $\mathcal{P}(\mathbf{y}, \hat{\mathbf{c}}_{MAP}|\mathcal{B})$.

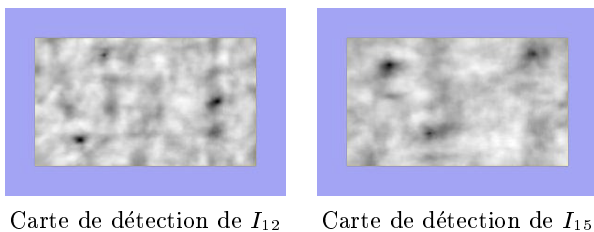


FIG. 5: Cartes de la norme des pondérations $\|\mathbf{b}^A\|^2$.

de fausses alarmes est reporté en abscisse et le pourcentage de bonnes détections en ordonnée. Elles montrent la nette supériorité des méthodes robustes de détection. On remarque également que la norme de la pondération surpasse la vraisemblance robuste.

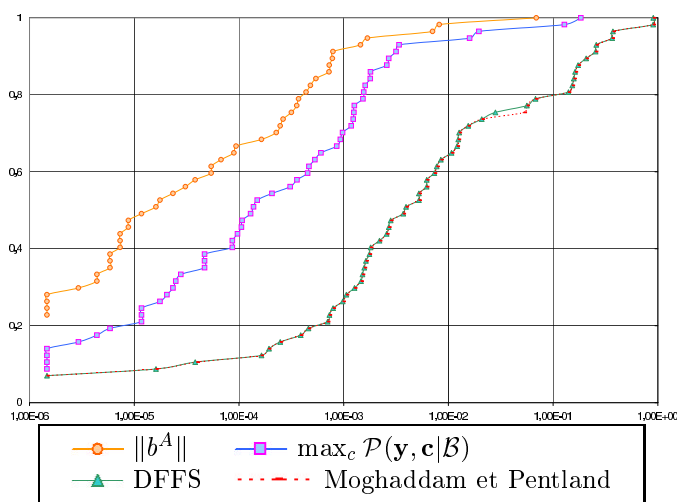


FIG. 6: Courbes COR obtenues pour les détecteurs classiques et robustes.

La vraisemblance de Moghaddam et Pentland obtient des résultats similaires à la DFFS. En effet, l'hypothèse de distribution gaussienne de \mathbf{c} dans F n'est pas vérifiée par

notre base d'apprentissage [7]. De même, cette hypothèse n'apporte pas d'améliorations dans le cas des détecteurs robustes. Il est possible de choisir d'autres hypothèses sur la distribution $\mathcal{P}(\mathbf{c}|\mathcal{B})$ mieux adaptées, afin d'améliorer encore ces résultats. Cependant, cela pourrait entraîner une augmentation de la complexité algorithmique pour l'estimation de $\hat{\mathbf{c}}_{MAP}$.

7 Conclusion

Deux méthodes robustes de détection basées sur une représentation de l'apparence par espace propre, ont été proposées. Elles se basent sur une prise en compte explicite de l'erreur grossière produite par d'éventuelles occultations, dans la définition de la fonction de vraisemblance utilisée. L'optimisation de cette fonction est réalisée par un algorithme semi-quadratique rapide. Les résultats obtenus montrent une amélioration de la détection par rapport aux méthodes classiques.

Références

- [1] M. J. Black and A. D. Jepson. Eigentracking: Robust matching and tracking of articulated objects using a view-based representation. *International Journal on Computer Vision*, 26(1):63–84, janvier 1998.
- [2] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud. Two deterministic half quadratic regularization algorithms for computed imaging. In *International Conference on Image Processing*, pages 168–172, Austin, USA, 1994.
- [3] R. Dahyot, P. Charbonnier, and F. Heitz. Robust visual recognition of colour images. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 685–690, Hilton Head Island, USA, juin 2000.
- [4] P.J. Huber. *Robust Statistics*. John Wiley and Sons, New York, 1981.
- [5] B. Moghaddam and A. Pentland. Probabilistic visual learning for object detection. In *The 5th International Conference on Computer Vision*, Cambridge, USA, juin 1995.
- [6] B. Moghaddam and A. Pentland. Probabilistic visual learning for object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):696–710, Juillet 1997.
- [7] H. Murase and S. K. Nayar. Visual learning and recognition of 3-d objects from appearance. *International Journal of Computer Vision*, 14(1):5–24, janvier 1995.
- [8] S. A. Nene, S. K. Nayar, and H. Murase. Columbia object image library (coil-20). Technical report, CUCS-005-96, 1996.