

Canaux projectifs intégraux pour la mise en correspondance de régions

Integral P-channels for region matching

Alain Pagani¹

Didier Stricker^{1,2}

¹ DFKI

² Université de Kaiserslautern

Trippstadterstr. 122, D-6766 Kaiserslautern, Allemagne
alain.pagani@dfki.de

Résumé

Nous présentons une nouvelle méthode de mise en correspondance de régions utilisant la représentation par canaux projectifs (P-canaux) et s'appuyant sur l'utilisation d'images intégrales. Nous proposons dans cet article une amélioration de la méthode des images intégrales qui permet d'utiliser les coordonnées spatiales dans le descripteur. Grâce aux images intégrales, la représentation par canaux projectifs d'une région donnée est calculée en quelques opérations arithmétiques, ce qui permet d'utiliser une recherche rapide du plus proche voisin parmi toutes les régions cibles possibles. Nous présentons nos résultats expérimentaux et montrons que notre approche produit de meilleurs résultats que les méthodes existantes pour la mise en correspondance de régions telles que les histogrammes ou la covariance de région.

Mots Clef

Mise en correspondance de régions, images intégrales

Abstract

We present a new method for matching a region between an input and a query image, based on the P-channel representation, and using integral images. In this paper, we propose an improvement of the integral image methods that allows using normalized spatial coordinates in the descriptor. Using integral images, the P-channel representation of a given region is extracted with a few arithmetic operations. This enables a fast nearest-neighbor search in all possible target regions. We present extensive experimental results and show that our approach compares favorably to existing methods for region matching such as histograms or region covariance.

Keywords

Region matching, integral images



FIGURE 1 – Mise en correspondance de régions par canaux projectifs. (haut) Région de référence. (bas) Image cible et région détectée.

1 Introduction

La mise en correspondance est une technique répandue en vision par ordinateur, car elle permet d'utiliser une image de référence connue comme source d'information pour analyser une image nouvelle. La mise en correspondance de points, par exemple, peut être utilisée pour calculer la pose d'un objet [9]. Dans cet article, nous considérons le problème de la mise en correspondance de régions rectangulaires, définie comme suit : étant données une image et une région rectangulaire de référence, trouver cette région dans une nouvelle image non connue. La solution de ce problème peut être appliquée à de nombreuses tâches, telles que la mise en correspondance d'objets, la reconnaissance d'objets, la classification de textures et le suivi 2D ou 3D. Notre motivation première est d'utiliser la mise en correspondance de régions pour l'initialisation de systèmes de suivi. Dans ce cas, aucune information sur la position de l'objet n'est disponible et il est nécessaire d'effectuer une recherche exhaustive de la région dans l'image. La figure 1 présente des exemples de mise en correspondance de régions. Une question centrale pour la mise en

correspondance est la définition du descripteur de la région et de la distance associée. Récemment, de nombreux descripteurs ont été proposés pour différentes applications. Par exemple, le descripteur SIFT [10] ainsi que les descripteurs invariants par transformation affine [11] sont très utilisés pour la mise en correspondance de points d'intérêt [13] [8], mais leur mise en oeuvre pour des régions plus grandes aux formes et tailles variables reste difficile. Par opposition, les descripteurs non paramétriques utilisant par exemple des histogrammes ou d'autres estimations de distribution sont mieux adaptés aux régions. Ces représentations ont connu un regain d'intérêt dans les dernières années grâce à une nouvelle technique permettant de calculer le descripteur pour n'importe quelle région à partir d'images intégrales. Récemment, une nouvelle méthode d'estimation de distribution appelée canal projectif, ou P-canal (P-Channel), a été présentée [3]. La représentation par canaux projectifs est un type particulier de représentation par canaux [7] qui combine les avantages des histogrammes et des modèles localement linéaires. Une autre caractéristique de la représentation par canaux projectifs est que chaque canal P contient une information sur la position spatiale des pixels qui lui contribuent. En cela ils peuvent être apparentés aux *spatiogrammes* [2]. La représentation par P-canaux a été appliquée avec succès dans le contexte de la reconnaissance d'objets par Felsberg [5] [4], où une seule représentation par P-canaux par image est calculée. Dans cet article, nous utilisons la représentation par P-canaux pour la mise en correspondance de régions. Pour cela, la représentation par P-canaux d'un grand nombre de régions de différentes échelles doit être calculée. Cependant, cette opération utilise trop de ressources si les P-canaux sont calculés de la manière standard. C'est pourquoi nous proposons l'utilisation d'images intégrales pour le calcul rapide de représentations par P-canaux. Une des difficultés majeures dans l'utilisation des images intégrales est que le descripteur considéré contient entre autres les coordonnées spatiales normalisées, qui constituent un élément capital pour la robustesse des P-canaux. Cependant, dans leur utilisation classique, les images intégrales ne peuvent pas tenir compte de la normalisation des coordonnées spatiales. Dans cet article, nous apportons une solution à ce problème, et montrons comment utiliser des représentations intermédiaires qui permettent d'utiliser les coordonnées spatiales dans le descripteur.

La première application des images intégrales a été le calcul efficace de filtres de type Haar pour la détection de visages [15]. Cette méthode a été ensuite étendue aux histogrammes intégraux [12] [1] et à la covariance de région [14]. Dans cet article, nous montrons que le concept d'images intégrales peut être étendu aux descripteurs qui incluent les coordonnées spatiales normalisées comme les canaux projectifs. L'extraction des P-canaux est alors assez rapide pour permettre une recherche exhaustive de la meilleure région dans l'image en temps réel. La contribution principale de cet article est notre nouvel algorithme

pour le calcul rapide de P-canaux en utilisant les images intégrales, ainsi qu'une méthode générale pour inclure les coordonnées spatiales normalisées dans les images intégrales. De plus, nous comparons les résultats obtenus par les P-canaux avec ceux des histogrammes intégraux et de la covariance de région et montrons l'avantage des premiers dans une série d'expériences.

Notre article est organisé comme suit : nous commençons par présenter les canaux projectifs dans la section 2. Nous montrons ensuite comment calculer la représentation par P-canaux en utilisant des images intégrales dans la section 3. Nos expériences sont détaillées dans la section 4, où nous comparons notre méthode avec les histogrammes intégraux et la covariance de région.

2 P-canaux

Dans cette section, nous présentons rapidement le concept de P-canaux. De plus amples détails pourront être trouvés dans l'article original de Felsberg [4].

2.1 La représentation par canaux

Dans la représentation par canaux [7], un ensemble de n opérateurs de projection B_i sont placés de manière arbitraire dans l'espace des caractéristiques. Les supports des opérateurs de projection peuvent se chevaucher (ce qui est souvent le cas). Un vecteur de caractéristiques \mathbf{f} , est codé en un vecteur de canal $\mathbf{p} = (p_1, \dots, p_n)$ en appliquant les opérateurs de projection un par un :

$$p_i(\mathbf{f}) = B_i(\mathbf{f}) \quad i = 1, \dots, n \quad (1)$$

Pour un ensemble de vecteurs de caractéristiques d'une distribution $\{\mathbf{f}_j\}, j = 1, \dots, J$, on calcule la moyenne des vecteurs de canal correspondants \mathbf{p}_j , ce qui équivaut à un système de vote flou :

$$\mathbf{p} = \frac{1}{J} \sum_{j=1}^J \mathbf{p}_j \quad (2)$$

En général, les opérateurs de projection sont infiniment dérivables et non négatifs. Ils peuvent être de forme diverses, comme par exemple les fonctions \cos^2 , les B-splines, ou les fonctions gaussiennes [7].

2.2 P-canaux

L'un des désavantages de la représentation par canaux est que le nombre de valeurs à calculer est exponentiellement proportionnel au nombre de dimensions, à cause du chevauchement des canaux. Cela peut être évité si des opérateurs à support disjoint sont utilisés. Le cas le plus commun des représentations par canaux sans chevauchement est celui des histogrammes, où les opérateurs sont les classes de l'histogramme. Cependant, la précision des histogrammes est limitée par le nombre de classes. Ce problème peut être résolu en utilisant une structure algébrique plus riche que les nombres réels positifs. L'idée des P-canaux [3], ou canaux *projectifs*, est de représenter les canaux en utilisant

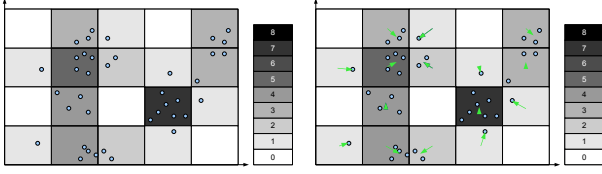


FIGURE 2 – Représentation par P-canaux et par histogrammes. Gauche : Histogramme 2D. Droite : Représentation par P-canaux en 2D (voir texte)

l'espace projectif \mathbb{P}^D (où D est la dimension de l'espace des caractéristiques), c'est-à-dire en utilisant des coordonnées homogènes pour les vecteurs. Les P-canaux utilisent $D+1$ fonctions de base pour chaque canal : une fonction de base rectangulaire (une classe ordinaire d'histogramme) et D composantes mémorisant la différence avec le centre du canal. Ainsi, la valeur d'un canal \mathbf{p}_i est un vecteur et la représentation par canaux $\mathbf{P} = (\mathbf{p}_1, \dots, \mathbf{p}_n)$ est une matrice (la matrice de P-canaux).

Montrons maintenant comment un ensemble de vecteurs de caractéristiques $\{\mathbf{f}_j\}$ est encodée en une matrice de P-canaux. Sans préjudice à la généralité, nous pouvons supposer que les canaux sont situés aux positions entières de l'espace de caractéristiques D -dimensionnel. Ainsi, la position d'un canal est représenté par un indice multiple \mathbf{i} , c'est-à-dire un vecteur d'indices (i_1, \dots, i_k) . Soit \mathbf{f}_j un vecteur de caractéristiques. Nous définissons $[\mathbf{f}_j]$ comme étant le plus proche entier composant par composant. Le vecteur \mathbf{f}_j est codé en une matrice de P-canaux de la manière suivante :

$$\begin{aligned} \mathbf{p}_i(\mathbf{f}_j) &= \begin{pmatrix} \mathbf{f}_j - [\mathbf{f}_j] \\ 1 \end{pmatrix} & si[\mathbf{f}_j] = \mathbf{i}, \\ \mathbf{p}_i(\mathbf{f}_j) &= \begin{pmatrix} \mathbf{0} \\ 0 \end{pmatrix} & si[\mathbf{f}_j] \neq \mathbf{i} \end{aligned} \quad (3)$$

Pour un ensemble de vecteurs d'une même distribution $\{\mathbf{f}_j, j = 1, \dots, J\}$, on prend la moyenne des matrices \mathbf{P}_j (la moyenne linéaire est valide dans l'espace \mathbb{P}^D) :

$$\mathbf{p}_i = \frac{1}{J} \sum_{j=1, [\mathbf{f}_j]=\mathbf{i}}^J \begin{pmatrix} \mathbf{f}_j - [\mathbf{f}_j] \\ 1 \end{pmatrix} \quad (4)$$

La figure 2 montre une visualisation schématique de la représentation par P-canaux. Dans les deux images, les points représentent un ensemble de vecteurs 2D provenant d'une certaine distribution. L'image de gauche est un histogramme classique avec 5×4 classes (la valeur de chaque classe est représentée par nuances de gris). L'image de droite est la représentation par P-canaux. Chaque canal correspond à une classe de l'histogramme et contient l'information de l'histogramme plus une information de décentrage représentée par une flèche. Le décentrage est la somme des décentrages de chaque vecteur contribuant au

canal. Dans cette représentation schématique, la flèche correspond au décentrage moyen, qui peut être calculé en divisant le décentrage par la composante de l'histogramme. En général, les vecteurs de caractéristiques \mathbf{f}_j contiennent des informations sur chaque pixel comme la couleur ou le gradient local. Dans les P-canaux, ce vecteur est complété explicitement par les coordonnées (x, y) du pixel. Les dimensions x et y du vecteur sont encodées en utilisant n_x , resp. n_y canaux disjoints. Cela induit une subdivision naturelle de la région en cellules spatiales.

2.3 Mise en correspondance de P-canaux

Afin de comparer deux représentations par P-canaux, une distance doit être définie pour les P-canaux. Dans les premières applications des P-canaux, deux distances distinctes ont été utilisées. Dans [5], les P-canaux sont d'abord convertis en canaux B -spline et la divergence de Kullback-Leibler est alors calculée. Cette distance a le désavantage de nécessiter une conversion supplémentaire vers les canaux de type B -spline. Dans [4], une combinaison linéaire des composants des P-canaux est utilisée pour un apprentissage par décomposition en valeurs singulières (SVD). Dans cet étude, nous utilisons la distance euclidienne entre deux représentations par P-canaux. Bien que les P-canaux ne sont pas des vecteurs, la distance euclidienne est suffisante pour une mise en correspondance robuste et a l'avantage d'être très rapide.

3 Mise en correspondance par canaux projectifs intégraux

Nous désirons effectuer une recherche exhaustive de la meilleure région parmi toutes les régions à toutes les échelles possibles dans l'image. Cependant, il est impossible de calculer les P-canaux individuellement pour toutes les régions possibles et pour un nombre raisonnable de changement d'échelle en temps réel. Une recherche exhaustive parmi 70000 régions sur 19 échelles nécessite plus de 15 secondes pour une image de 320 par 240 pixels.

Cependant, ce temps peut être réduit de manière considérable (moins d'une demi seconde pour la même taille d'image et le même nombre de régions) en utilisant des représentations intermédiaires appelées *images intégrales*.

3.1 Images intégrales

Le concept d'images intégrales a été proposé dans [15] pour le calcul rapide de sommes sur des régions rectangulaires. Chaque pixel d'une image intégrale contient la somme des valeurs de tous les pixels à l'intérieur d'un rectangle défini par le coin supérieur gauche de l'image et le pixel considéré :

$$ImageIntegrale(x_0, y_0) = \sum_{x < x_0, y < y_0} Image(x, y) \quad (5)$$

Les images intégrales ont été utilisées entre autres pour le calcul de la valeur moyenne d'une région rectangulaire

dans le but de calculer rapidement des filtres de Haar [15], et ont ensuite été étendus aux dimensions multiples pour le calcul rapide d'histogrammes [12] et de covariance de région [14]. Dans cet article, nous adoptons une idée similaire pour le calcul des P-canaux.

3.2 Formulation des P-canaux intégraux

L'idée des P-canaux intégraux est qu'il est possible de calculer globalement une représentation par P-canaux intégrale pour une image entière dans une étape préliminaire, et en déduire la représentation par P-canaux pour n'importe quelle région rectangulaire en quelques opérations arithmétiques. Soit $\mathbf{f}(x, y)$ un vecteur de dimension D défini pour chaque pixel. Par exemple, \mathbf{f} peut contenir des informations sur la couleur et la géométrie comme suit : $\mathbf{f} = (h, s, |\nabla\mathbf{f}|, \theta)^T$, où $D = 4$, h et s sont la teinte (*hue*) et la saturation du pixel, et $|\nabla\mathbf{f}|$ et θ sont l'amplitude et l'orientation du gradient. Avant d'être codé en P-canaux, ce vecteur est complété par l'information spatiale (x, y) , conduisant au vecteur de dimension $(D + 2)$: $\mathbf{f}' = (\mathbf{f}^T, x, y)^T$. Afin d'encoder ce vecteur en P-canaux en utilisant n canaux pour chaque dimension et resp. n_x et n_y canaux pour les dimensions spatiales x et y , nous avons besoin d'un total de $n^D n_x n_y$ P-canaux $\mathbf{p}_{i,j,k}$, où \mathbf{i} est un indice multiple de dimension D , et j et k sont les indices des dimension spatiales x et y . Chaque P-canal est de dimension $(D + 3)$: les premières D dimensions correspondent au décentrage dans l'espace des caractéristiques, les dimensions $D + 1$ et $D + 2$ sont le décentrage spatial, et la dimension $D + 3$ est la composante histogramme. La représentation finale est la matrice \mathbf{P} de dimension $(D + 3) \times n^D n_x n_y$.

Lorsque l'on utilise les dimensions spatiales, il est nécessaire de les normaliser par la taille effective de la région pour permettre une comparaison entre différentes échelles. Cependant, la taille de la région n'est pas connue lors de l'étape préliminaire (calcul de l'image intégrale). Nous ne pouvons donc pas utiliser directement l'encodage de l'équation (3), et utilisons une représentation intermédiaire \mathbf{q}_i , définie comme suit :

$$\mathbf{q}_i(\mathbf{f}_j) = \begin{pmatrix} \mathbf{f}_j - [\mathbf{f}_j] \\ x \\ y \\ 1 \end{pmatrix} \quad \text{if } [\mathbf{f}_j] = \mathbf{i}, \quad (6)$$

$$\mathbf{q}_i(\mathbf{f}_j) = \begin{pmatrix} \mathbf{0} \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad \text{if } [\mathbf{f}_j] \neq \mathbf{i}$$

Dans cette représentation intermédiaire, chaque canal mémorise la somme des coordonnées spatiales absolues des pixels qui lui contribuent. Il est important de noter que dans l'équation (6), x et y ne sont pas normalisés. Pour un ensemble de vecteurs, on ne forme pas la moyenne des contributions comme habituellement, mais simplement la

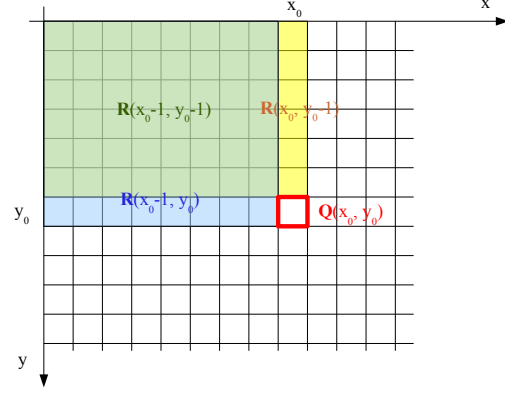


FIGURE 3 – Propagation de la représentation par P-canaux.

somme :

$$\mathbf{q}_i = \sum_{j=1, [\mathbf{f}_j]=\mathbf{i}}^J \begin{pmatrix} \mathbf{f}_j - [\mathbf{f}_j] \\ x \\ y \\ 1 \end{pmatrix} \quad (7)$$

Pour une région rectangulaire particulière définie par le coin supérieur gauche (x_0, y_0) et le coin inférieur droit (x_1, y_1) , la matrice de P-canaux correspondante dépend comme suit de la région :

$$\mathbf{Q}(x_0, y_0, x_1, y_1) = (\mathbf{q}_1(x_0, y_0, x_1, y_1), \dots, \mathbf{q}_{n^D}(x_0, y_0, x_1, y_1)) \quad (8)$$

La matrice de P-canaux *intégrale* $\mathbf{R}(x, y) = (\mathbf{r}_1(x, y), \dots, \mathbf{r}_{n^D}(x, y))$ est définie comme la matrice de P-canaux de la région définie par l'origine de l'image et le point (x, y) :

$$\mathbf{R}(x, y) = \mathbf{Q}(0, 0, x, y) \quad (9)$$

Propagation des P-canaux intégraux. Une propriété intéressante de la représentation par P-canaux est qu'elle peut être calculée par incréments. La représentation par P-canaux de l'union de deux régions disjointes A et B est donnée par :

$$\mathbf{P}(A \cup B) = \frac{1}{a+b} (a\mathbf{P}(A) + b\mathbf{P}(B)) \quad (10)$$

où a , resp. b , est le nombre de pixels de la région A , resp. B . Cette propriété permet de propager les P-canaux intégraux de manière récursive sur toute l'image en un seul parcours :

Le parcours de l'image calcule la valeur des P-canaux intégraux par incréments. Pour le calcul de la valeur d'un pixel, il suffit que ses voisins de gauche, du haut et de diagonale gauche-haut aient déjà été calculés. Pour la première ligne et la première colonne, les pixels voisins ont une valeur nulle par défaut. Ainsi, le parcours de l'image peut être

effectué ligne par ligne ou colonne par colonne, en commençant par le coin supérieur gauche. La figure 3 représente la propagation des P-canaux intégraux pour un pixel donné de coordonnées (x_0, y_0) . Pour ce pixel, les vecteurs de P-canaux intégraux \mathbf{r}_i sont la somme de deux termes. Le premier est le terme de propagation et est calculé à partir des pixels voisins (eq. (11)). Le second correspond à la contribution du pixel local et est calculé en utilisant l'encodage du vecteur de caractéristiques locales en P-canaux (eq. (12)) :

$$\mathbf{R}(x_0, y_0) \Leftarrow \mathbf{R}(x_0 - 1, y_0) + \mathbf{R}(x_0, y_0 - 1) - \mathbf{R}(x_0 - 1, y_0 - 1) \quad (11)$$

$$\mathbf{R}(x_0, y_0) \Leftarrow \mathbf{R}(x_0, y_0) + \mathbf{Q}(\mathbf{f}(x_0, y_0)) \quad (12)$$

où $\mathbf{Q}(\mathbf{f}(x, y))$ est la matrice de contribution locale dont les éléments $\mathbf{q}_i(\mathbf{f}(x, y))$ sont donnés par l'équation (6). Avec cette méthode, le calcul de l'image intégrale peut être effectué en un seul parcours de l'image.

Calcul des P-canaux intermédiaires à partir des P-canaux intégraux. Une fois que les P-canaux intégraux \mathbf{R} ont été calculés, il est possible de calculer la matrice de P-canaux intermédiaires $\mathbf{Q}(T)$ pour n'importe quelle région rectangulaire T en quelques opérations arithmétiques. Si la région est définie par les coins supérieur gauche (x_0, y_0) et inférieur droit (x_1, y_1) , les vecteurs de P-canaux $\mathbf{q}_i(T)$ sont donnés par :

$$\mathbf{q}_i(T) = \mathbf{r}_i(x_1, y_1) - \mathbf{r}_i(x_1, y_0) - \mathbf{r}_i(x_0, y_1) + \mathbf{r}_i(x_0, y_0) \quad (13)$$

Calcul des P-canaux à partir des P-canaux intermédiaires. Montrons maintenant comment calculer la représentation par P-canaux $\mathbf{P}(T)$ (avec coordonnées spatiales normalisées) à partir de $\mathbf{Q}(T)$. La région T est d'abord divisée en $n_x \times n_y$ cellules $C_{j,k}, (j, k) \in [1 \dots n_x] \times [1 \dots n_y]$. Pour chacune des cellules, une représentation intermédiaire $\mathbf{Q}_{j,k}$ est extraite des P-canaux intégraux. Les P-canaux $\mathbf{p}_{i,j,k}$ sont alors égaux aux P-canaux $\mathbf{q}_{i,j,k}$, à une normalisation spatiale près :

$$p_{i,j,k}^d = \frac{1}{t} q_{i,j,k}^d \quad \text{if } d \in [1, D] \cup \{D+3\}$$

$$p_{i,j,k}^{D+1} = \frac{1}{t(x_1-x_0)} (q_{i,j,k}^{D+1} - \frac{x_0+x_1}{2} q_{i,j,k}^{D+3}) \quad (14)$$

$$p_{i,j,k}^{D+2} = \frac{1}{t(y_1-y_0)} (q_{i,j,k}^{D+1} - \frac{y_0+y_1}{2} q_{i,j,k}^{D+3})$$

où (x_0, x_1, y_0, y_1) sont les coordonnées de la cellule $C_{j,k}$, et t est le nombre de pixels de la région T . La normalisation de l'équation 14 effectue l'opération suivante : pour chaque canal, l'origine de la cellule est soustrait un nombre de fois correspondant au nombre de pixels dans la composante histogramme, et le résultat est recentré et mis à l'échelle. On peut montrer facilement que la représentation résultante est

Algorithme 1 Mise en correspondance de régions par P-canaux intégraux

```

1: pour chaque image point  $j$  faire
2:   Propage P-canaux intégraux (eq. 11)
3:   pour  $d = 1$  à  $D$  faire
4:      $i_d \Leftarrow \text{arrondi}(f_{j,d})$ 
5:      $o_d \Leftarrow f_{j,d} - i_d$ 
6:   stop
7:    $l \Leftarrow \text{indice\_linaire}(i)$ 
8:   pour  $d = 1$  à  $D$  faire
9:      $r_{l,d} \Leftarrow r_{l,d} + o_d$ 
10:  stop
11:   $r_{l,D+1} \Leftarrow r_{l,D+1} + x$ 
12:   $r_{l,D+2} \Leftarrow r_{l,D+1} + y$ 
13:   $r_{l,D+3} \Leftarrow r_{l,D+1} + 1$ 
14: stop
15: pour chaque centre de région possible faire
16:   pour chaque échelle possible faire
17:     calcule  $n_x n_y$  P-canaux  $\mathbf{Q}$  (eq. 13)
18:     calcule les P-canaux de région  $\mathbf{P}$  (eq. 14)
19:     calcule la distance entre les P-canaux
20:   stop
21: stop

```

strictement équivalente à la représentation par P-canaux des vecteurs caractéristiques \mathbf{f}' pour la région T . L'algorithme 1 contient le pseudo code pour une recherche exhaustive utilisant les P-canaux intégraux. On peut noter en particulier que contrairement à la recherche exhaustive utilisant la méthode conventionnelle, cette méthode ne répète pas le calcul de la représentation par P-canaux pour chaque région cible possible.

4 Résultats

Présentons maintenant nos résultats expérimentaux. Nous comparons les résultats de notre méthode avec la mise en correspondance de régions par les histogrammes intégraux [12] et la covariance de région [14].

Pour la méthode des P-canaux et la méthode des histogrammes, le facteur limitant est le nombre d'images intégrales devant être calculées. En effet, une image intégrale doit être calculée pour chaque valeur mémorisée dans chaque classe de l'histogramme (1 valeur par classe pour les histogrammes, et $(D+3)$ par canal pour la représentation par P-canaux). Nous avons donc choisi un nombre limité de caractéristiques et de classes. Nous utilisons des images en couleurs et les caractéristiques sont la teinte (*hue*), la saturation et l'orientation locale du gradient (calculée sur une image monochrome). Pour les P-canaux, le nombre de canaux est $n = 3$, et les dimensions spatiales sont divisées en $n_x = n_y = 2$ cellules, ce qui conduit au calcul de 162 images intégrales. Pour les histogrammes, le nombre de classes est $n = 5$, avec un nombre total de 125 images intégrales. Pour chacune des ces deux méthodes, la distance euclidienne est utilisée. Nous avons choisi un

nombre limité de canaux et de cellules spatiales, ainsi que la distance euclidienne afin de diminuer le temps de calcul. Nos expériences ont montré que la mise en correspondance reste robuste malgré cette limitation.

Pour la méthode de covariance de région, beaucoup plus de caractéristiques peuvent être utilisées, car la distribution des caractéristiques est encodée seulement par la matrice de covariance des vecteurs. Si nous prenions seulement 3 caractéristiques, la comparaison ne serait pas équitable, puisque seulement 9 images intégrales seraient nécessaires. Nous choisissons donc d'utiliser plus d'information et utilisons le vecteur de caractéristiques de dimension 9 utilisé dans la publication originale : la position des pixels (x, y) , les 3 canaux de couleur (RGB), et la norme de la dérivée première et seconde de l'intensité dans les directions x et y . Le nombre total d'images intégrales est alors 54. La distance entre deux matrices de covariance est la distance de Förstner [6].

Pour les trois méthodes, nous appliquons la méthode d'affinage des résultats suivante : Les meilleures 1000 régions trouvées sont comparées à nouveau à la région de référence en utilisant 5 sous-régions (voir [14] pour plus de détails).

4.1 Recherche exhaustive

Pour chaque test de recherche de région, nous utilisons une image comme référence et définissons une manuellement région rectangulaire de référence. Pour chacune des autres images représentant le même objet, nous recherchons la meilleure région de manière exhaustive en comparant de nombreuses régions à la région de référence. Les images sont de dimension 320 par 240 pixels. La première recherche est effectuée avec une région rectangulaire de la même taille que la région de référence et avec un déplacement de 6 pixels entre deux positions de recherche. Nous recommandons ensuite la recherche pour 9 échelles plus petites et 9 échelles plus grandes, avec une différence de 15% entre deux échelles consécutives (le déplacement entre les positions de recherche est également mis à l'échelle et arrondi). Pour une région de référence de 100 par 100 pixels, un total d'environ 70.000 régions sont testées pour chaque image. Dans nos tests, bien que les images forment une séquence video, chaque image est traitée séparément, et l'information temporelle n'est pas utilisée.

4.2 Séquences

Les trois méthodes ont été testées avec 4 séquences différentes¹. La figure 4 montre quelques images issues de ces séquences avec les régions trouvées par les trois méthodes. (Note : les temps requis par chacune des méthodes diffèrent, mais les vidéos ont été encodées à une vitesse fixe de 15 images par seconde).

La première séquence ("bouteille") montre une application de notre méthode pour une région semi-plane avec changement de pose partiel. Notre méthode trouve la région cor-

1. Séquences disponibles sur le site web des auteurs à l'adresse <http://www.dfki.uni-kl.de/~pagani/video/orasis09>

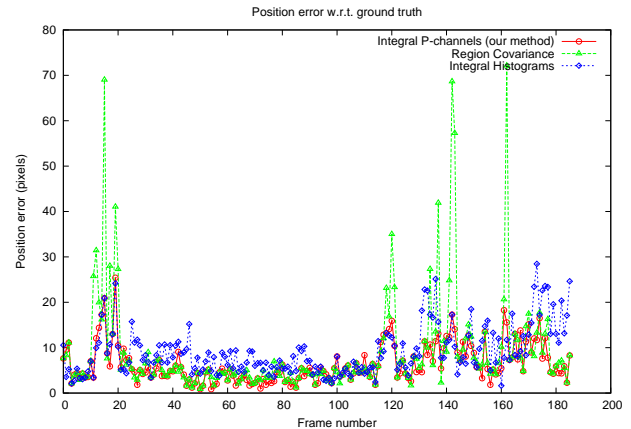


FIGURE 5 – Erreur par rapport à la région réelle

recte même en présence de grande variation d'échelle et de flou. La deuxième séquence ("puzzle") montre le résultats avec variations de la lumière incidente sur l'objet. Les séquences "jouets" et "voitures" (troisième et quatrième lignes) montrent les résultats avec un objet en 3 dimensions.

La figure 5 montre l'erreur moyenne sur la position des coins par rapport à la région réelle marquée soigneusement à la main pour une autre séquence. L'erreur importante obtenue par le méthode de la covariance pour certaines images est due au flou de mouvement de la caméra. Dans toutes les séquences, les résultats obtenus par notre méthode sont légèrement meilleurs que ceux de la méthode des covariances, tandis que la méthode des histogrammes conduit à des erreurs plus grandes que celles obtenues par les autres méthodes. Ceci peut s'expliquer par le fait que les histogrammes intégraux n'utilisent pas d'information spatiale, tandis que notre méthode et celle des covariances l'utilisent.

4.3 Consommation des ressources

Nous avons implémenté les trois méthodes de mise en correspondance en C++ en utilisant un processeur Intel Core2 à 2.4 GHz. Les trois méthodes utilisent la même architecture, et seulement la partie spécifique de chaque méthode diffère. Il faut noter que notre code n'est pas optimisé et que les indications temporelles données ici servent à comparer les méthodes entre elles seulement. Le tableau 1 montre le temps nécessaire à chaque méthode pour l'étape préliminaire (calcul des images intégrales), pour le calcul des distances et le temps total de la recherche pour une image et pour chacune des trois méthodes : histogrammes intégraux (HI), covariance de région (CR), P-canaux intégraux (PCI). Nous avons utilisé une région de référence de 100 par 100 pixels, avec des facteurs d'échelle de 0.57 à 1.75, conduisant à 13317 régions candidates. Les résultats montrent que même si notre méthode soit plus lente que les histogrammes intégraux, elle est plus rapide que la

Méthode	HI	CR	PCI
Etape préliminaire	100.73	47.35	122.07
Distance	28.15	349.5	137.81
Total	128.88	396.85	259.88

TABLE 1 – Temps nécessaire (en msec) pour la comparaison de 13317 régions avec une région de référence. (HI) : histogrammes intégraux, (CR) : covariance de région, (PCI) : P-canaux intégraux. Voir texte pour plus de détails.

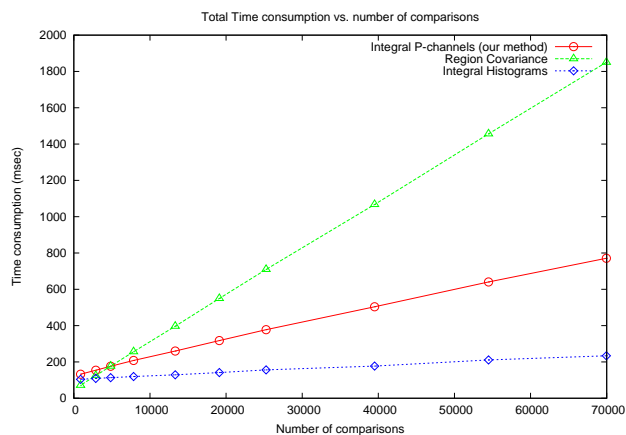


FIGURE 6 – Consommation en temps en fonction du nombre de régions candidates.

covariance de région tout en obtenant des résultats légèrement meilleurs. Le gain de temps est approximativement de 50%.

Nous avons comparé le temps nécessaire pour chacune des méthodes pour différents nombres de régions candidates (voir figure 6). La différence entre les pentes de ces courbes s’explique par l’utilisation de la distance de Förstner pour la méthode des covariances, où un problème de valeurs propres généralisées doit être résolu pour chaque distance à calculer. Par opposition, les P-canaux intégraux et les histogrammes intégraux utilisent la distance euclidienne qui est beaucoup plus rapide à calculer. Ainsi, les P-canaux et les histogrammes sont d’autant plus intéressants que le nombre de régions candidates augmente.

5 Conclusion

Nous avons présenté une nouvelle méthode pour le calcul rapide de la représentation par canaux projectifs (P-canaux) d’un grand nombre de régions dans une image. Cette méthode utilise la structure des images intégrales pour construire une représentation intégrale par P-canaux, à partir de laquelle la représentation par P-canaux de n’importe quelle région de l’image peut être calculée de manière extrêmement rapide (environ 100 reconstructions de P-canaux par msec). Cette méthode rapide sert de base à

un algorithme de mise en correspondance de régions utilisant la méthode du plus proche voisin sur l’ensemble des régions candidates. Nous avons comparé les P-canaux intégraux avec deux autres méthodes pour la mise en correspondance de régions et nous avons montré que notre méthode obtient de meilleurs résultats que les histogrammes intégraux et des résultats similaires à ceux de la méthode de la covariance, tout en étant plus rapide de 50%. Dans sa version actuelle, notre algorithme effectue une recherche exhaustive parmi toutes les régions possibles dans une image. Notre recherche s’oriente maintenant vers l’utilisation de l’information temporelle pour réduire l’espace de recherche et affiner les résultats au moyen de filtres temporels.

Références

- [1] A. Adam, E. Rivlin, and I. Shimshoni. Robust fragments-based tracking using the integral histogram. In *CVPR*, 2006.
- [2] S.T. Birchfield and S. Rangarajan. Spatiograms versus Histograms for Region-Based Tracking. In *CVPR*, 2005.
- [3] M. Felsberg and G. Granlund. P-channels : Robust multivariate m-estimation of large datasets. In *ICPR*, 2006.
- [4] M. Felsberg and J. Hedberg. Real-time view-based pose recognition and interpolation for tracking initialization. *J. of Real-Time Image Proc.*, 2 :103–115, 2007.
- [5] M. Felsberg and J. Hedberg. Real-time visual recognition of objects and scenes using p-channel matching. In *Scand. Conf. on Image Analysis*, 2007.
- [6] W. Förstner and B. Moonen. A metric for covariance matrices. *Quo vadis geodesia*, pages 113–128, 1999.
- [7] G.H. Granlund. An Associative Perception-Action Structure Using a Localized Space Variant Information Representation. In *AFPAC*, 2000.
- [8] C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey vision conference*, 1988.
- [9] V. Lepetit, J. Pilet, and P. Fua. Point matching as a classification problem for fast and robust object pose estimation. In *CVPR*, 2004.
- [10] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comp. Vision*, 60 :91–110, 2004.
- [11] K. Mikolajczyk and C. Schmid. Scale & Affine Invariant Interest Point Detectors. *Int. J. Comp. Vision*, 60 :63–86, 2004.
- [12] F. Porikli. Integral histogram : a fast way to extract histograms in Cartesian spaces. In *CVPR*, 2005.
- [13] J. Shi and C. Tomasi. Good Features to Track. In *CVPR*, 1994.
- [14] O. Tuzel, F. Porikli, and P. Meer. Region covariance : A fast descriptor for detection and classification. In *ECCV*, 2006.
- [15] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *CVPR*, 2001.

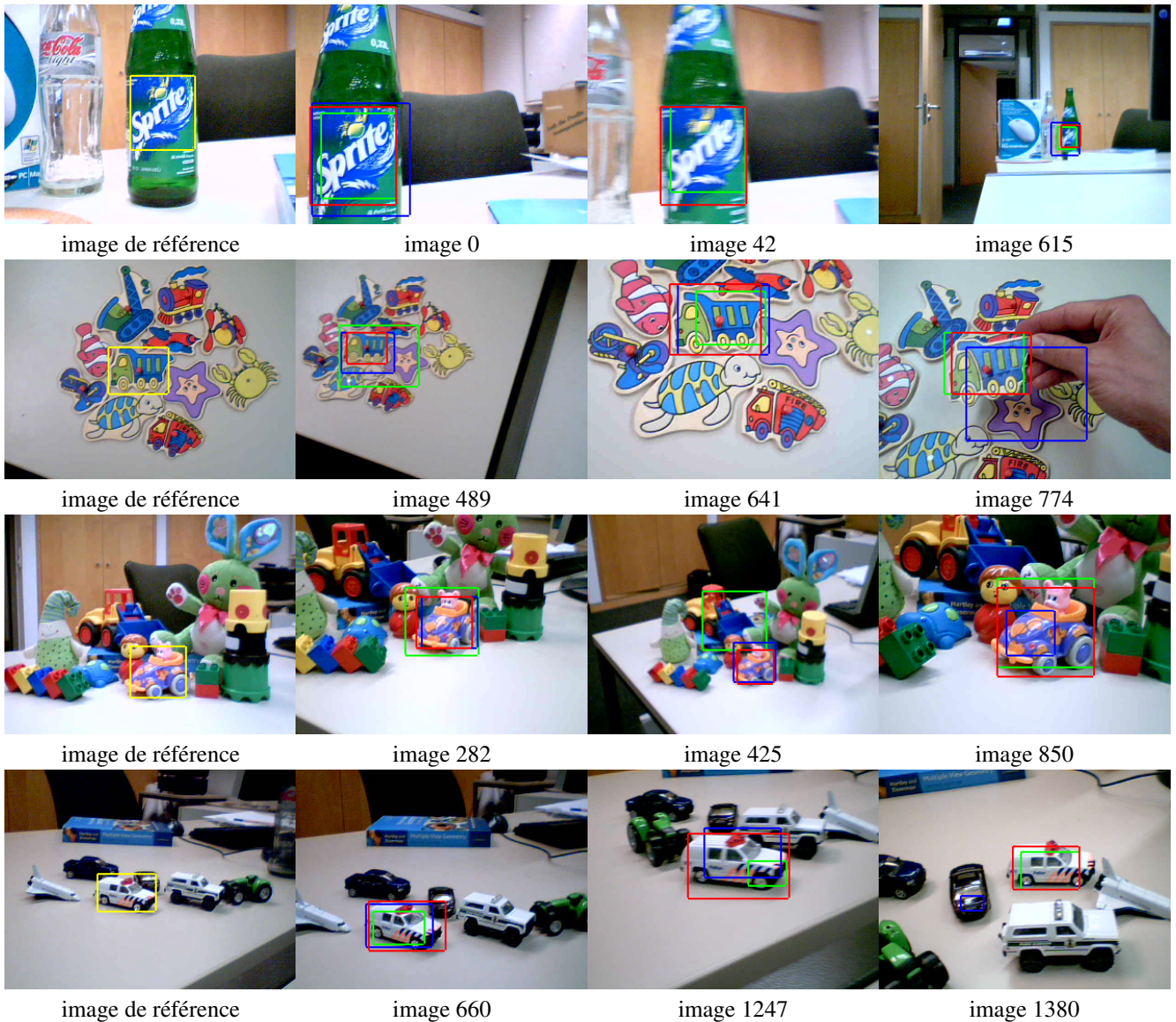


FIGURE 4 – Comparaison de trois méthodes de mise en correspondance de régions. La première colonne montre les images de référence et les régions de référence. Les autres images montrent les résultats obtenus avec notre méthode des P-canaux (rouge), la méthode de covariance de régions (vert), et la méthode des histogrammes (bleu). Notre méthode obtient toujours le bon résultat alors que la méthode des covariances, bien qu'elle donne de bons résultats, n'est pas performante dans certains cas, notamment lorsqu'il y a un changement d'échelle. Les séquences complètes sont disponibles sur le site web des auteurs (<http://www.dfki.uni-kl.de/~pagani/video/orasis09>).