N-20/92/MM

Modélisation symbolique complète des étapes intervenant dans la construction d'un codeur d'images orienté par la notion d'objet

-=-=-=-

M. VAN DROOGENBROEK

CMM Fontainebleau

August 1992

Modélisation symbolique complète des étapes intervenant dans la construction d'un codeur d'images orienté par la notion d'objet

Marc Van Droogenbroeck *

Août 1992

1 Introduction

L'objectif de ce court rapport est de développer et de présenter un formalisme général, capable de modéliser toutes les étapes de traitement d'images intervenant dans la construction d'un procédé de codage d'images évolué, qu'il s'agisse par exemple de la segmentation, d'un filtrage non-linéaire ou encore d'une simplification du contenu de l'image. La raison d'une telle tentative est qu'il n'existe pas à l'heure actuelle, du moins à notre connaissance, de formalisme universellement adopté; sans doute parce que certains domaines du traitement d'images demeurent fortement cloisonnés. Toute l'étude que nous mènerons restera cohérente à cette notation, ce qui permettra au lecteur d'établir aisément les liens entre des parties apparemment fort distinctes ou sans rapport aucun.

2 Cadre mathématique et notations

Soit une image I à niveaux de gris à traiter. Les opérations entreprises en traitement d'images sont nombreuses et parfois fort différentes. Ce qui nous concerne plus particulièrement est le codage d'images par objets. Pour coder une image par objets, il faut

- segmenter l'image, autrement dit identifier les objets présents dans la scène,

*M. Van Droogenbroeck est titulaire d'une bourse de l'Institut pour l'Encouragement de la Recherche Scientifique dans l'Industrie et l'Agriculture (I.R.S.I.A.– Belgique).

- qualifier les objets, les décrire précisément,
- puis il appartient de coder ces objets de la manière la plus compacte possible.

Plus l'image obtenue après décodage est conforme à l'image de départ, plus le codage est respectueux du contenu. Comme le codage est une technique qui supprime la redondance présente dans une image, l'image décodée est différente de l'image de départ et cela d'autant plus que les débits sont faibles. Présenté de la sorte, le codage opère une transformation de suppression de redondance d'information et fournit en fin de compte une *approximation de l'image* $\pi(I)$. Si l'écart entre l'image I et son approximation $\pi(I)$ est faible, la technique de codage est de bonne qualité. Idéalement, il y a égalité "perceptive" parfaite entre l'image et sa représentation, c'est-à-dire qu'un observateur ne perçoit aucune différence entre les deux. A ce stade de la discussion, il est utile de fournir un cadre mathématique plus rigoureux.

Cadre mathématique. Chaque image de luminance I définit de manière unique une application ni injective ni surjective qui, à chaque point du plan E (il s'agit du plan euclidien discrétisé $\mathbb{Z}^2$ ou continu $\mathbf{R}^2$) [1], associe une valeur de gris comprise dans un intervalle donné $\mathcal{G}$;

$$I:\ E \rightarrow \mathcal{G}:\ (i,j) \rightarrow I(i,j) \tag{1}$$

Si la position (i,j) est fixée, une nouvelle série d'applications à valeurs dans $\mathcal{G}$ s'obtient en définissant un opérateur ϕ qui transforme une image en une autre image:

$$\phi\ :\ \mathcal{G} \rightarrow \mathcal{G}:\ \forall (i,j) \in E,\ I(i,j) \rightarrow \phi(I(i,j)) \tag{2}$$

L'opérateur *approximation* est une application de ce type, tout comme l'est une opération de filtrage. L'*opérateur identité* noté 1 envoie un ensemble sur lui-même. Si $\circ$ représente la composition usuelle d'applications, $\phi \circ 1(I) = \phi(I) = 1 \circ \phi(I)$. Il arrive que l'ensemble de définition et l'ensemble image soient distincts, la définition de composition d'opérateurs reste valable à la condition que l'ensemble d'arrivée du premier opérateur concorde avec l'ensemble de départ du second opérateur. Les opérateurs de représentation et de reconstruction décrits plus loin constituent une paire d'opérateurs appartenant à cette famille.

[1] Pour la commodité des démonstrations, nombreux sont les auteurs qui recourent à $\mathbf{R}^2$ systématiquement. Ils perdent parfois de vue que la discrétisation de l'image transforme la nature de la modélisation et escamote ainsi les considérations.

Egalité entre deux opérateurs. L'égalité entre deux opérateurs ϕ et φ se traduit par une égalité en tout point du plan:

$$\phi(I) = \varphi(I) \Leftrightarrow \forall(i,j), \quad \phi(I(i,j)) = \varphi(I(i,j)) \tag{3}$$

Avec ces définitions, toute opération intervenant dans la formation d'une technique de codage acquiert une consistance mathématique à partir de laquelle il est possible de l'étudier en détail.

Une mesure de la fidélité de l'approximation $\pi(I)$ à l'image I est par exemple l'écart quadratique moyen

$$\zeta(I,\pi) = \frac{\sum_i \sum_j \{1[I(i,j)] - \pi[I(i,j)]\}^2}{M_i M_j} \tag{4}$$

où M_iM_j représente la taille de l'image.

L'approximation est parfaite si, en tout point, $\pi(I) = 1(I) = I$, autrement dit, il s'agit de l'opérateur identité. Il existe des techniques de codage à reconstruction parfaite, comme celle qui consiste à effectuer un codage différentiel sur les valeurs de luminance mais un tel codage offre des taux de compression peu élevés.

3 Représentation d'images

Le codage que nous désirons synthétiser repose sur la notion d'objet; une image est composée d'objets, véritables ensembles de points voisins ayant des propriétés homogènes entre-eux. Pour la modélisation, nous introduisons le concept de *représentation d'image* $\rho(I)$ qui par un procédé de *recomposition* adéquat noté τ conduira à une forme particulière d'*approximation* $\pi(I) = \tau(\rho(I))$. La représentation d'images est le concept-clef de tout algorithme s'appuyant sur la notion d'objet. Elle guide tous les traitements depuis la phase de détection d'objets à celle du codage en passant par l'étude du mouvement.

Insistons sur la nécessité d'adjoindre un procédé de reconstruction τ à toute représentation d'images $\rho(I)$. Cette remarque nous conduit à définir la représentation d'une image.

Définition 1 *En toute généralité, une représentation de l'image I est un outil de modélisation $\rho(I)$, qui par l'intermédiaire d'une paire représentation-reconstruction (ρ, τ), définit une application d'approximation $\pi = \tau \circ \rho$.*

Si la représentation guide fortement la formation d'un procédé de reconstruction, il n'en demeure pas moins –comme nous le verrons plus loin– qu'il existe un peu de liberté pour la caractérisation complète de ce dernier.

Toute représentation d'images $\rho(I)$ met certains aspects en lumière. C'est par son intermédiaire que se mène une réflexion d'interprétation de l'image, tantôt en termes de fréquences, tantôt en termes d'objets. Donnons quelques exemples d'approximations π générées par des paires (ρ, τ).

Exemples

1. Considérons la représentation naïve de l'image échantillonnée I, qui se compose des triplets $(i, j, I(i,j))$. On a donc $\rho(I) = \{(i, j, I(i,j))|\ (i,j) \in \mathbb{Z}^2\ et\ I(i,j) \in \mathcal{G}\}$. Le procédé de reconstruction est très simple: il suffit de placer la valeur $I(i,j)$ au point (i,j) du plan de l'image. Dans ce cas, la reconstruction de l'image conduit à une approximation parfaite $I = \pi(I)$. Une représentation telle que celle que nous venons de décrire présente peu d'intérêt parce que l'interprétation n'apporte aucune information nouvelle.

2. L'opération de filtrage linéaire $H(I)$ de l'image se transcrit sans difficulté dans le nouveau formalisme. Pour ce faire, la représentation d'images $\rho(I)$ est constituée de nombres complexes dans le plan transformé. Soient (u, v) les coordonnées du plan transformé et $\mathcal{F}$ la transformée de Fourier de l'image I, $\rho(I) = \{(u, v, \mathcal{F}(u,v))\}$, où

$$(u,v) \in E\ \ et\ \ \mathcal{F}(u,v) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} I(x,y) e^{-j(ux+vy)} dx dy\}$$

   Le filtrage s'opère à la reconstruction par un mécanisme à deux niveaux: (i) l'amplitude $\mathcal{F}(u,v)$ est multipliée par la réponse impulsionnelle de filtre $\mathcal{H}(u,v)$, (ii) vient ensuite la transformation inverse. Cet exemple montre clairement que si la représentation fixe le cadre de travail, toute liberté est allouée au procédé de reconstruction de fixer la réponse impulsionnelle du filtre.

3. Soit la représentation d'images regroupant les coefficients a_i obtenus par une transformation orthogonale de tous les blocs de l'image et associant une position dans la matrice bloc et dans l'image. Cette représentation est l'ensemble des coefficients ordonnés a_i ou encore $\rho(I) = \{(a_i, o_i)\}$ où o_i est un numéro d'ordre. La reconstruction procède primo au positionnement exact de chaque coefficient, secundo à la transformation inverse. La plupart des techniques de codage actuelles (codage sous-bande, codage par transformée en cosinus discrète, ...) s'appuyent sur des variantes plus ou moins complexes de l'approximation offerte par cette combinaison $\rho - \tau$. Remarquons la présence d'un concept d'ordre dans l'approximation. Bien des techniques englobent un ordre, ce qui n'est pas sans danger car toute perturbation dans l'ordre amène inéluctablement une distorsion de l'approximation. Dans la mise au point

d'une technique de codage, il faudra veiller à protéger l'ordre ou à éliminer au mieux toute la partie d'ordre qu'il n'est pas possible de connaître a priori.

4. Une image est représentée par une collection d'objets qui se caractérisent par leur forme X et par leur aspect, leur texture T. Une représentation possible est composée de la réunion de l'information concernant la forme de tous les objets X et leur position dans l'image (i, j): $\rho(I) = \{(X, i, j)\}$. La recomposition positionne correctement un ensemble de formes dans l'image, mais l'information de textures a disparu. L'approximation résultante n'est, en toute généralité, pas à reconstruction parfaite.
   Les images mosaïques illustrent un style de représentation d'image où la seule information de texture disponible est la valeur moyenne de la luminance sur chacune des régions.
   La figure 1 schématise un procédé de codage axé sur le concept d'objet et utilisé à des fins de transmission. Une remarque est importante: les contours et les textures sont deux sortes d'information pas toujours très bien décorrélées entre-elles. Dans la mesure du possible, on recupèrera le contenu d'une branche pour injecter l'information dans l'autre branche. Dans la figure, c'est l'information de contour qui est introduite dans la branche de codage des textures; rien n'empêche d'intervertir le croisement. Cette constatation influe considérablement sur les développements d'algorithmes de codage. C'est elle qui justifie *l'extrapolation* de textures, encore appelée *déconvolution*, dont il sera question plus loin. Elle est destinée à simplifier les textures d'une partie de l'information résiduelle de contour.

4 Etude de représentations d'images par objets

4.1 Représentations

Le procédé de codage, que nous désirons synthétiser, nécessite une représentation de la forme des objets X et de leur texture T: $\rho(I) = \rho_{X,T}(I)$. A des fins de transmission, c'est cette représentation qui sera codée et envoyée au récepteur. Caractériser la forme est aisé mais définir la notion de texture relève du défi. Nous proposons de donner une consistance ensembliste à ces concepts partant d'un ensemble abstrait $\mathcal{H}(I)$ qui englobe toute l'incertitude, et par voie de conséquence, toute l'information contenue dans l'image I. Ce choix n'est pas né du hasard car de la sorte nous espérons associer des entropies de source aux ensembles X et T nouvellement formés.

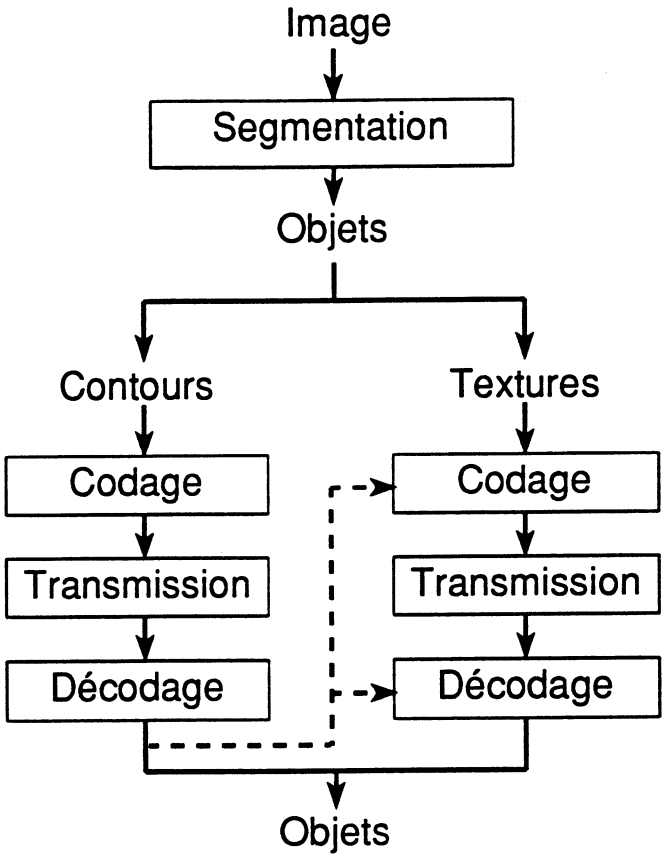

Figure 1: Schéma d'un algorithme de codage orienté par la notion d'objet.

Définition 2 *Soit $\mathcal{H}(I)$ l'ensemble des informations contenues dans l'image I. Les objets regroupent deux sous-ensembles X et T de $\mathcal{H}(I)$ décrivant respectivement la forme des objets et leur aspect. Les sous-ensembles X et T ne sont pas nécessairement disjoints; de plus, ils sont des éléments d'hypothétiques ensembles des parties relatives à la forme $\mathcal{P}(X)$ et l'aspect ou texture $\mathcal{P}(T)$ des objets de l'image I. Cette dernière précision signifie que toute description d'objets par X et T est un choix parmi d'autres descriptions.*

Placer un jeu de contours sur une image n'est pas toujours une mince affaire. La difficulté provient partiellement du fait que l'image est une projection d'un monde tridimensionnel. Il en résulte que, régulièrement, la décorrélation d'information entre la forme et la texture est imparfaite, tant et si bien que la représentation se compose en toute rigueur de plusieurs parties $\rho_{X,T}(I) = \rho_{X\setminus T}(I) + \rho_{T\setminus X}(I) + \rho_{X\cap T}(I)$ [2]:

[2]Les symboles $\setminus$ et $\cap$ identifient respectivement la différence et l'intersection ensemblistes.

- la partie descriptive de forme $\rho_{X\setminus T}(I)$,
- l'information de texture $\rho_{T\setminus X}(I)$, et
- une partie commune $\rho_{X\cap T}(I)$.

Pour la commodité des notations et par souci de clarté de l'exposé, nous noterons $\rho_{X\setminus T}$ par $\rho(X)$, $\rho_{T\setminus X}$ par $\rho(T)$ et par extension $\rho_{X\cap T}$ par $\rho(X\cap T)$. La partie commune $\rho(X\cap T)(I)$ de la représentation comprend par exemple la position du coin supérieur gauche du plus petit rectangle circonscrit à chaque objet ou un ordre de description commun. D'un point de vue algorithmique, $\rho(X\cap T)$ est la partie d'information extraite d'une branche et injectée dans l'autre branche, opération illustrée par la figure 1. Parfois il s'agit d'une part d'information plus difficile à expliciter. En effet, les représentations $\rho(X)$ et $\rho(T)$ codées empruntent physiquement ou virtuellement deux canaux distincts; alors, si le codage de la texture s'effectue par l'envoi de coefficients transformés, il importe de déconvoluer le spectre –autrement dit, d'étendre le signal de texture jusqu'à ce qu'il occupe un domaine rectangulaire– par une fonction qui décrit la forme de l'objet, sinon les coefficients contiendraient une partie d'information relative aux contours. Or la déconvolution est souvent imparfaite: toute information de texture contiendra résiduellement la forme du contour. La confusion entre $\rho(X)$ et $\rho(X)+\rho(X\cap T)$ est malheureusement si facile que nous choisirons souvent d'inclure inplicitement $\rho(X\cap T)$ dans $\rho(X)$. Il s'agit là certes d'un manque de rigueur critiquable mais ce choix privilégie le codage des contours ce qui, dans la pratique, s'explique parce que le débit nécessaire aux contours représente la plus grande part du débit total. Il paraît alors logique de concentrer les efforts sur le codage des contours et de récupérer tant que faire se peut cette information pour les textures.

Pour compléter la démarche de modélisation et l'étude du codage, nous devrons introduire ensuite des notions de la théorie de l'information car l'évaluation de la quantité à conserver dans la partie commune de la représentation dépendra des gains réalisés dans les parties $\rho(X)$ et $\rho(T)$. En effet, il est erroné de croire que la meilleure représentation de codage suppose que $\rho(X\cap T)=\emptyset$ car, par exemple dans la situation où la recomposition de la forme et de la texture nécessite une connaissance exacte de la position de l'objet, repousser les couples de position (i,j) dans les deux parties $\rho(X)$ et $\rho(T)$ grèverait inutilement le coût du codage.

4.2 Recompositions et approximations

L'approximation de l'image s'obtient au récepteur par recomposition des représentations. Elle se dédouble également en deux parties relatives à la forme et la texture des différents objets. En filigrane sont obtenues deux approximations d'image incomplètes si elles sont prises séparément et bien plus

précises si considérées simultanément. Les approximations en termes d'objets et de textures valent respectivement $\pi(X) = \tau(\rho(X) + \rho(X \cap T))$ et $\pi(T) = \tau(\rho(T) + \rho(X \cap T))$, alors que l'approximation d'une image codée par un schéma orienté objet est $\pi(I) = \tau(\rho(X) + \rho(T) + \rho(X \cap T))$.

Il est licite de s'interroger sur les propriétés d'une approximation en termes d'objets. Sans vouloir clore à présent le débat, il semble logique d'exiger que la représentation d'une approximation soit conforme à la représentation première, en d'autres termes, il est exigé que $\rho(I) = \rho \circ \tau \circ \rho(I)$. Dans ce cas, $\pi(I) = \pi \circ \pi(I)$, autrement dit l'approximation est un *opérateur idempotent* ou *opérateur de projection.*

5 Réduction de représentations et codage

Telles quelles les représentations $\rho(X)$ et $\rho(T)$ ne conviennent pas spécifiquement au codage; en cause, la redondance présente dans la représentation et le manque d'organisation. Pour le codage, nous *réduisons* les représentations afin d'obtenir des représentations épurées et ordonnées, indicées par le symbole r. Pour diminuer davantage le débit binaire nécessaire à la transmission, utilisation est faite de résultats de la théorie de l'information. Dans un schéma global de codage par objets, le codage $\mathcal{C}$ capable de supprimer la redondance agit séparément sur les contours et sur les textures. Plus précisément, c'est par l'intermédiaire de très bonnes représentations réduites $\rho_r(X)$, $\rho_r(T)$ et éventuellement $\rho_r(X \cap T)$ que nous comptons concurrencer dans certaines applications de codage les techniques actuelles.

A l'étape de codage $\mathcal{C}$ est associée l'étape de décodage $\mathcal{D}$ qui reconstitue la représentation de l'image.

Définition 3 *Une paire codage-décodage* $(\mathcal{C}, \mathcal{D})$ *d'une image* sans perte d'information *est un procédé judicieux de réduction de redondance qui, dans le canal, modifie la représentation réduite de l'image* $\rho_r(I)$ *de manière à la rendre plus compacte ou à l'adapter au canal. Formellement, cela signifie que*

$$\rho_r(\mathcal{D} \circ \mathcal{C}(I)) = \rho_r(I) \tag{5}$$

La recomposition τ ne joue apparemment aucun rôle. C'est inexact, car la liaison entre la représentation et la recomposition est établie par la définition d'une représentation (cfr. définition 1).

6 Segmentation d'images

Le premier problème à résoudre lors du codage d'une image par objets est la décomposition en objets appelée *segmentation.* Formellement, la segmentation s'apparente à un opérateur $\mathcal{E}$ agissant sur l'image I et fournissant, par

exemple, une image binaire $\mathcal{E}(I)$ qui différencie les points des contours des objets, ou un champ $\vec{\mathcal{E}}(I)$ de vecteurs d'orientations de contours semblable à un champ gradient. Exiger l'idempotence d'un opérateur de détection de contours nécessite d'obtenir un détecteur $\tilde{\mathcal{E}}(I)$ à valeurs dans $\mathcal{G}$, ce qui n'est pas toujours réalisable ou souhaitable. Notre propos est de tirer la représentation de forme $\rho(X)$ de cette étape de segmentation, sans oublier que la segmentation conditionne également du tout au tout la représentation de texture $\rho(T)$.

Définition 4 *L'obtention des objets compris dans une image I se réalise par application d'un opérateur noté $\mathcal{E}(I)$ et appelé opérateur de segmentation. Après application de l'opérateur, on dispose des représentations de forme $\rho(X)$ et de texture $\rho(T)$.*

De fait, l'adoption d'une démarche cohérente de codage par objets impose la caractérisation de tous les objets dès l'étape de détection des objets. Dès lors, le détecteur sera idempotent non pas sur I mais sur $\rho(I)$: $\mathcal{E}(\rho(I)) = \mathcal{E} \circ \mathcal{E}(\rho(I))$.

Malheureusement, les liens qui unissent l'étape de segmentation à la description par objets ne découlent pas naturellement. Pire, la plupart des techniques de segmentation d'images ne permettent pas la moindre liaison avec une représentation correcte et judicieuse de l'image. Ainsi, bon nombre d'outils de détection génèrent des contours non fermés, impossibles à considérer dans un modèle orienté objet. Pour l'étude d'un outil de segmentation adapté à notre problème, il faudra

(i) étudier les *propriétés* des outils de segmentation, mais surtout,

(ii) établir une *correspondance entre $\mathcal{E}(I)$ et $\rho(I)$* en gardant à l'esprit qu'une *description hiérarchique* en termes d'objets est indispensable. La description hiérarchique par multirésolution permet une description et un codage progressif de l'image, allant du trait grossier au détail précis. De plus, elle assure une certaine compatibilité de formats et acquiert tout son sens dans une approche par objets.

7 Mouvement d'objets

Jusqu'à présent, nous avons attiré l'attention sur la modélisation des opérations sur une image fixe. Hormis le cas particulier du changement de scène, la plupart des objets se retrouvent dans les images successives; quelques-uns se sont déplacés, d'autres sont demeurés immobiles. L'exploitation de la correspondance entre *blocs* similaires d'image à image se nomme *compensation de mouvement.* Dans une approche orientée par la notion d'objet, ce n'est

pas le bloc qui est considéré comme l'entité de comparaison mais l'objet. Passer à la notion d'objet complique le traitement parce que la forme n'est plus fixe mais dépendante du contenu de l'image, des objets. Pour utiliser l'information de mouvement d'objets dans un codage, plusieurs phases sont indispensables:

1. Identifier les objets présents, ce qui revient à découvrir une représentation au temps t de l'image, $\rho^*(I,t) = \rho^*(X,t) + \rho^*(T,t) + \rho^*(X \cap T, t)$ adaptée à l'étude temporelle. Concrètement, il s'agira de la représentation $\rho(I)$ obtenue au terme de la détection de régions légèrement modifiée.

2. Analyser l'image au temps t^+ (t^+ est le moment qui suit immédiatement t) par le même outil afin d'obtenir la représentation $\rho^*(I,t^+)$.

3. Etablir la correspondance entre $\rho^*(I,t)$ et $\rho^*(I,t^+)$.

4. Coder le mouvement des objets retrouvés dans la scène au temps t^+ et les nouveaux objets.

Il va sans dire que l'avant-dernière étape s'avère particulièrement difficile. Non seulement, comme dans l'approche classique où s'utilise le critère d'identification par sommation de valeurs absolues, les objets à faire correspondre entre $\rho^*(I,t)$ et $\rho^*(I,t^+)$ possèdent des textures identiques, mais de plus, les objets se ressemblent par leur forme, qu'ils aient subi un mouvement de translation, de rotation ou de déformation. Nous espérons utiliser des notions d'extrapolation spatiale pour établir une correspondance de formes entre objets ayant subi un mouvement planaire quelconque ou un mouvement de déformation homothétique.