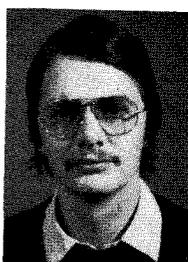


Algorithmes

de traitement d'image et réseaux systoliques

A real-time 2 level systolic 2 D convolution chip



Michel FOUQUES

CETIA : Centre d'Études des Techniques de l'Informatique et d'Analyses de Systèmes.

CSEE : Compagnie de Signaux et d'Entreprises Électriques.

150, rue Marcellin-Berthelot, ZI Toulon Est, 83088 TOULON CEDEX.

Points forts des réalisations et travaux en cours : calculateurs conformes et architectures à haut degré de parallélisme; réalisation d'opérateurs spécialisés de type data flow et systolique; intégration VLSI en technologie CMOS; développement de logiciels d'aide à la conception et à la simulation de réseaux systoliques.



Roger SAMY

CETIA : Centre d'Études des Techniques de l'Informatique et d'Analyses de Systèmes.

CSEE : Compagnie de Signaux et d'Entreprises Électriques.

150, rue Marcellin-Berthelot, ZI Toulon Est, 83088 TOULON CEDEX.

Points forts des réalisations et travaux en cours : traitement de séquences d'images télévision en temps réel, en vue d'application à des problèmes de poursuite; étude d'algorithmes de traitement et d'analyse d'images orientées vers l'intégration en composant VLSI; application des techniques de reconnaissance de forme au codage d'images.

RÉSUMÉ

Cet article présente l'architecture d'un circuit réalisant l'opération de convolution bidimensionnelle, dans le cas usuel d'un masque de dimension 3×3 .

Les contraintes d'intégrabilité (sous forme d'un composant VLSI) ont amené à choisir une architecture systolique à deux niveaux :

- association systolique de cellules élémentaires de multiplication-addition de mots binaires;
- association systolique de microcellules opérant sur les bits.

MOTS CLÉS

Convolution 2D, architecture systolique, intégration VLSI.

SUMMARY

This paper describes the architecture of a chip for performing the 2D convolution.

The systolic architecture proposed by H. T. Kung [1] has been chosen for the array of basic cells but a second level of systolic architecture has been introduced with the internal structure of a basic cell. The basic cell performs multiplication/accumulation in a serial way with a systolic architecture at the bit level.

KEY WORDS

2D convolution, Systolic architecture, VLSI integration.

TABLE DES MATIÈRES**Introduction****1. Filtrage linéaire d'images****2. Réseaux systoliques appliqués au filtrage linéaire bidimensionnel****Conclusion****Remerciements****Bibliographie****Annexe****Introduction**

En vision par ordinateur, on distingue généralement quatre grandes étapes dans le traitement des informations :

- (a) les prétraitements;
- (b) l'extraction d'attributs;
- (c) la segmentation;
- (d) la reconnaissance des formes.

Les prétraitements et l'extraction d'attributs font appel à des opérations sur un voisinage de points, une classe de ces opérations étant basés sur le filtrage linéaire, aisément décomposable dans le cas de signaux multidimensionnels en opérations élémentaires de convolution. D'une manière générale, on peut considérer que la plupart des algorithmes de prétraitement et d'extraction d'attributs mettent en œuvre une ou plusieurs convolutions élémentaires par un masque 3×3 ou 5×5 .

Le filtrage linéaire des signaux bidimensionnels, tel qu'une image 2D, est très coûteux en nombre d'opérateurs élémentaires.

En effet, pour une image 2D, une convolution par un noyau de dimension $n \times m$ dont les coefficients sont quelconques nécessite $n \times m$ multiplication et $(n \times m) - 1$ addition par point.

Des efforts importants ont été effectués pour résoudre les problèmes associés à la complexité intrinsèque des calculs. Sur le plan des choix algorithmiques on peut citer, par exemple :

- les transformations orthogonales (FFT) qui sont efficaces pour les noyaux de convolution dont les dimensions sont supérieures à 10×10 .
- la décomposition SVD/SGK (Singular Value Decomposition/Small Generating Kernel) 3 du noyau afin d'obtenir une approximation de la réponse impulsionnelle du filtre mis en œuvre moins coûteuse en nombre d'opérations.

Sur le plan des solutions technologiques, les processeurs de tableau (Array Processor) dont l'architecture est du type SIMD (Single Instruction Multiple Data) sont largement utilisés dans les centres d'études et de recherches et des opérateurs câblés spécifiques sont souvent développés pour les applications industrielles et militaires.

L'évolution des technologies d'intégration des composants électroniques permet aujourd'hui de réaliser un opérateur de convolution par un noyau de dimension 3×3 ou 5×5 dans des conditions acceptables pour ce qui concerne le compromis « coût-efficacité ».

Dans ce qui suit, on présente l'architecture de type systolique d'un tel opérateur dans le cas 3×3 , l'étude s'étendant aisément à un masque 5×5 .

1. Filtrage linéaire d'images

On présente brièvement le filtrage linéaire et les filtres de dimensions 3×3 classiques utilisés en traitement et analyse d'images. L'algorithme de filtrage linéaire d'images 2D repose sur la convolution bi-indicielle :

$$(1) G(j, k) = \sum_{m=1}^M \sum_{n=1}^N F(m, n) H(j-m+1, k-n+1),$$

où $F(m, n)$ désigne l'image originale; $H(j, k)$ désigne le noyau de convolution de taille $N \times M$ et $G(j, k)$, l'image filtrée.

Une formule plus condensée utilise le symbole \otimes pour désigner la convolution 2D :

$$(2) G(j, k) = F(j, k) \otimes H(j, k).$$

On trouvera en annexe une liste des filtres utilisés classiquement en traitement d'images. Ces filtres qui correspondent à des fonctionnalités différentes peuvent être classés de la façon suivante :

- filtres passe bas ($H_1 \dots H_3$);
- filtres passe haut ($H_4 \dots H_6$);
- filtres laplaciens ($H_7 \dots H_9$);
- filtres de gradient ($H_{10} \dots H_{26}$).

2. Réseaux systoliques appliqués au filtrage linéaire bidimensionnel

La notion d'architecture systolique a été introduite par H. T. Kung [2] en 1978 pour faciliter la réalisation sous forme de composants VLSI d'algorithmes présentant un taux de parallélisme important et faisant intervenir des opérations élémentaires répétitives.

Un réseau systolique est constitué d'un ensemble de processeurs identiques, appelés cellules élémentaires de calcul, interconnectés entre eux de façon simple et

APPLICATIONS

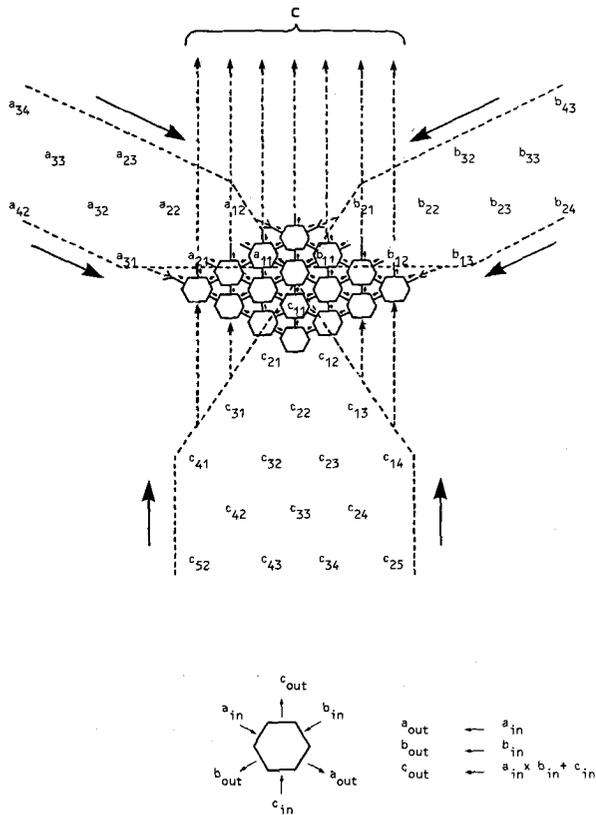


Fig. 1.

rigide et fonctionnant de façon synchrone. La figure 1 représente un exemple de réseau systolique réalisant le produit de deux matrices, la figure 2 représentant trois phases du calcul de ce réseaux.

Les avantages de ces réseaux sont les suivantes :

- modularité du réseau et facilité d'intégration;
- taux de parallélisme important;
- flux de données et d'instruction de contrôle ayant une définition simple et par essence réguliers;
- multitraitement de l'ensemble des données (une donnée entrant dans le réseau participe à plusieurs calculs élémentaires).

RÉSEAUX SYSTOLIQUES DE CONVOLUTION 2 D

Nous présentons dans cette étude l'architecture d'un opérateur de convolution par un masque 3×3 , les principes retenus autorisant l'implémentation de cet opérateur sous forme de composants VLSI. De plus, pour réduire le nombre et la complexité des processeurs utilisés tout en permettant la réalisation de la plupart des masques 3×3 usuels, on a choisi de limiter la valeur des coefficients à des valeurs entières entre -15 et $+15$.

L'algorithme de convolution à deux dimensions présente des caractéristiques (calculs simples et répétitifs pour un grand nombre de données à traiter) facilitant la réalisation de réseaux systoliques adaptés. H. T.

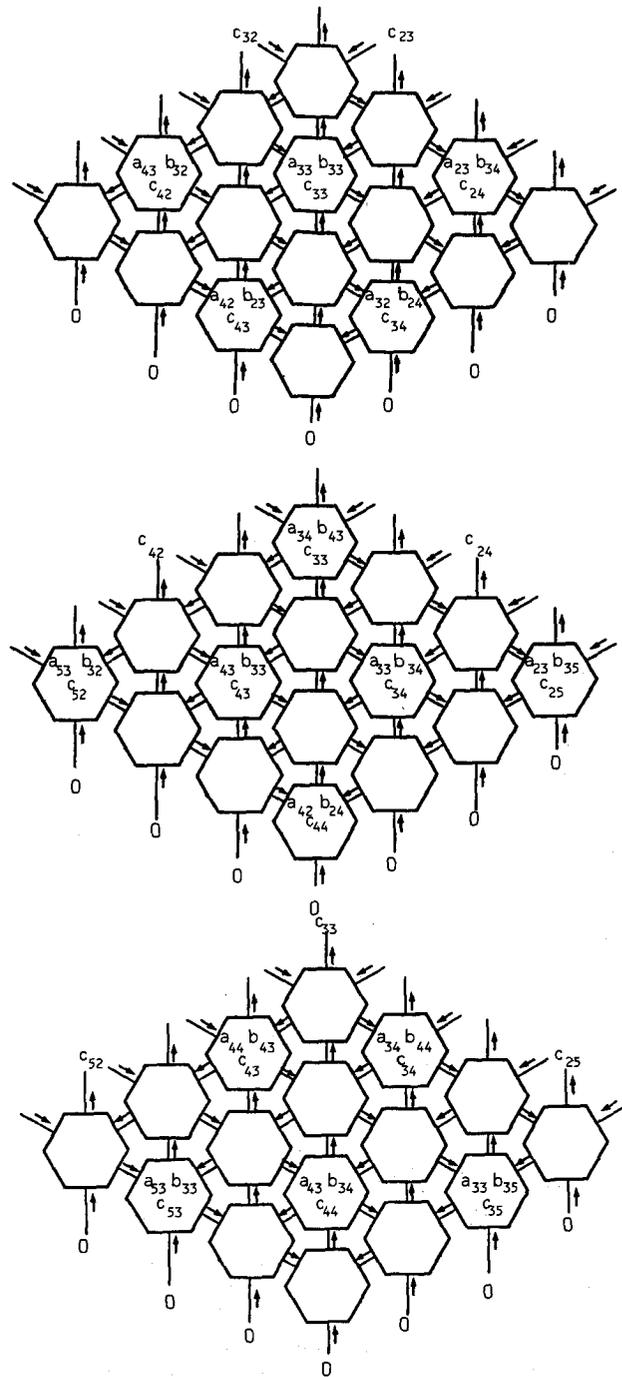


Fig. 2. - Trois étapes successives du calcul $A \times B \rightarrow C$.

Kung a proposé deux réseaux réalisant la convolution par un masque 3×3 de coefficients [2].

Première approche (fig. 3)

Les coefficients W_{ij} sont préchargés dans chaque cellule, et trois lignes de pixels décalées d'un pixel sont introduites aux entrées 1, 2, 3. Dans cette solution un pixel est introduit dans une entrée de réseau tous les deux cycles de calcul d'une cellule élémentaire.

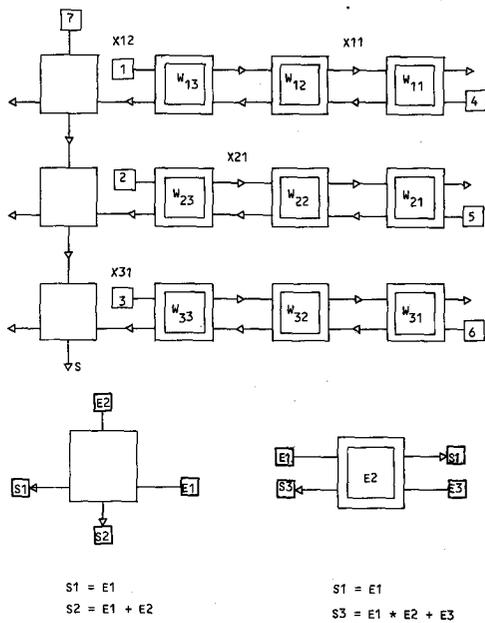


Fig. 3.

Deuxième approche

Les coefficients du masque sont préchargés comme dans l'approche précédente, les trois lignes de pixels sont décalées dans le temps d'un pixel mais un seul cycle de calcul sépare l'introduction de deux pixels voisins (fig. 4).

Ces deux approches nécessitent neuf cellules de calcul comportant des multipliers-additionneurs, disposées en trois lignes, ce qui facilite l'intégration et la modularité par rapport à une architecture utilisant neuf multipliers suivis d'un étage de sommation des neuf multipliers.

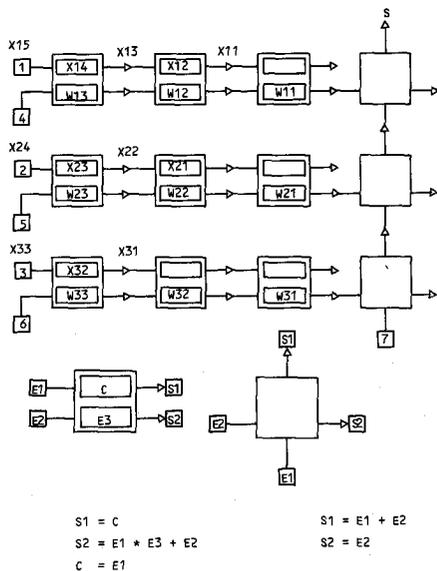


Fig. 4.

IMPLÉMENTATION

Nous présentons ici une réalisation de la cellule élémentaire de calcul de $a * b + c$ appliquée à la convolution 2D sous la forme d'un réseau systolique comportant des « multipliers » basées sur des additionneurs un bit.

Le calcul binaire de la multiplication et de l'addition est traité de façon série, ce qui permet l'intégration de neuf multipliers additionneurs dans le même circuit, par opposition à un traitement parallèle.

La multiplication binaire d'un pixel codé sur 8 bits par un coefficient du masque codé sur 4 bits (plus 1 bit de signe : coefficients entiers variant de +15 à -15) utilise quatre « microcellules ».

Deux types de « microcellules » peuvent être définis conduisant à deux réseaux systoliques multipliers différents.

Première solution

Les termes en B_i représentent les différents bits caractérisant la valeur du pixel.

Les termes en A_i représentent les termes composant la valeur absolue du coefficient du masque. Chaque bit est préchargé dans chacune des « microcellules ».

Les bits B_i du pixel sont introduits tous les deux cycles de calcul des « microcellules ».

Le premier bit S_0 du résultat est obtenu après un cycle de calcul, et les S_i sont élaborés tous les deux cycles de calcul (fig. 5).

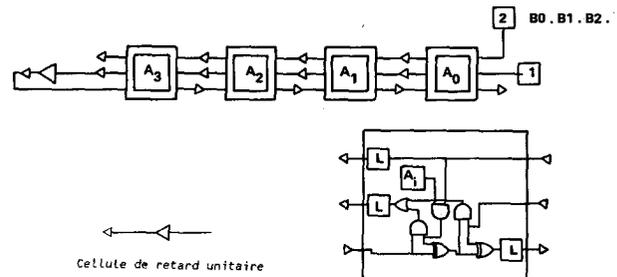


Fig. 5.

Deuxième solution

Les bits A_i du coefficient du masque sont toujours préchargés dans les microcellules.

Les bits B_i du pixel sont introduits à chaque cycle de calcul.

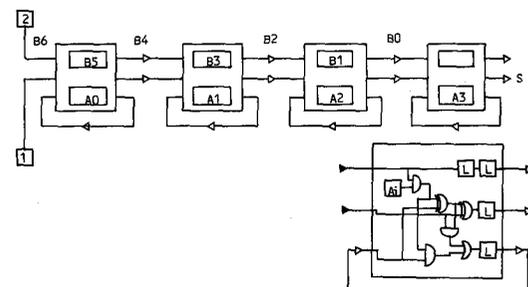


Fig. 6.

APPLICATIONS

Le premier bit du résultat est obtenu après quatre cycles de calcul des « microcellules » (fig. 6).

L'addition avec un résultat précédent est effectuée pour les deux solutions en complément à deux, mais suivant deux techniques différentes.

Pour la première solution (fig. 7), le résultat de la multiplication est complété à deux suivant le bit de signe du coefficient du masque, puis l'addition est réalisée en complément à deux avec le résultat précédent.

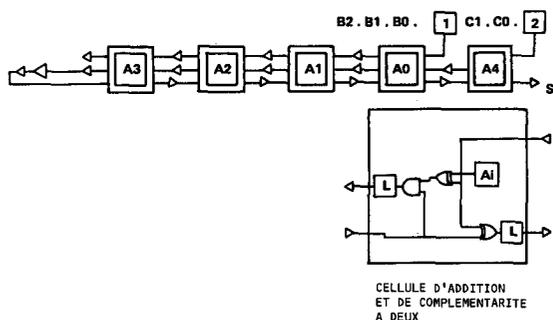
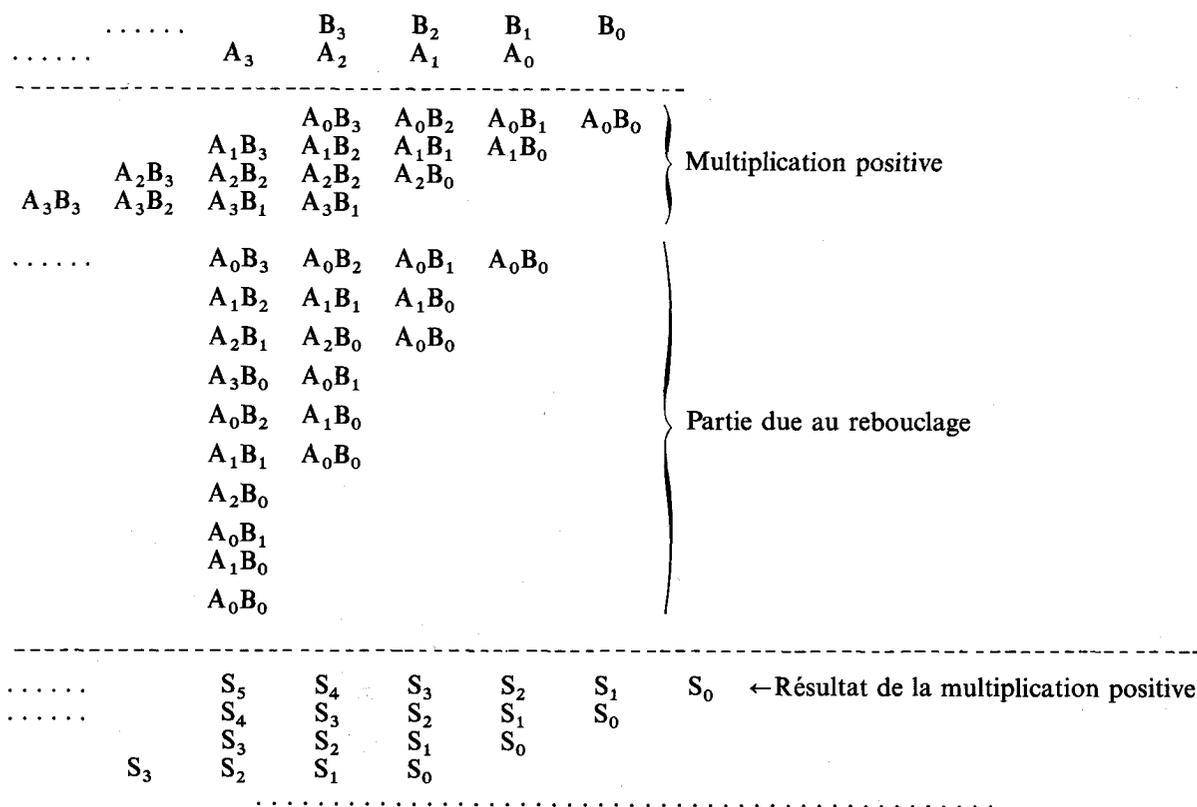


Fig. 7. — Multiplier-additionneur.

La complémentation à deux du résultat de la multiplication est réalisé en rebouclant la sortie du multiplieur sur son entrée 1 à travers une cellule de retard ce qui conduit au calcul suivant :



qui vaut :

$$= S + S \times 2 + S \times 4 + S \times 8 \dots$$

S étant le résultat de la multiplication sans rebouclage (multiplication positive) ou encore

$$= S(2 + 2^2 + 2^3 + 2^4 \dots 2^n)$$

n nombre de bits de la multiplication au positif

$$= S \times (2^{n+1} - 1).$$

Or $2^{n+1} - 1$ est égal à -1 en décimal codé sur n bit en complément à deux.

Le résultat de la multiplication avec rebouclage égal à $-S$ est donc le complément à deux du résultat de la multiplication positive.

Les bits B_i du pixel sont introduits sur l'entrée 1 du réseau un cycle de calcul sur deux, les bits du résultat d'un calcul précédent étant introduits sur l'entrée 2 du réseau avec un retard d'un cycle de calcul par rapport aux B_i et étant également présents un cycle sur deux. Le premier bit S_0 du résultat est disponible deux cycles de calcul après l'entrée du premier bit B_0 du pixel.

L'addition avec un résultat précédent pour la deuxième réalisation de multiplieur (fig. 8) est effectuée en complément à deux les bits du pixel si le signe du coefficient est négatif puis la multiplication et l'addition sont effectuées en complément à un.

La complémentation à deux du pixel est réalisée en additionnant -1 codé en complément à deux puis en calculant le complément à deux.

Les B_i sont introduits à chaque cycle de calcul à l'entrée 1. Les bits d'un calcul précédent sont introduits à chaque cycle à l'entrée 2 décalés dans le temps d'un cycle par rapport aux bits B_i du pixel.

Le premier bit du résultat est obtenu après cinq cycles de calcul.

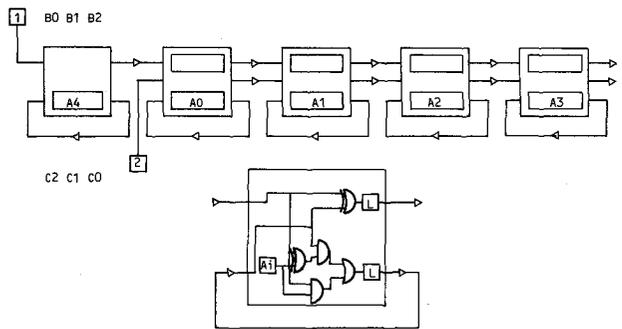


Fig. 8.

ASSOCIATION DES CELLULES ET DES MICROCELLULES

Pour la réalisation du convolveur, il faut donc associer neuf cellules élémentaires (première ou deuxième solution) dans une des deux approches proposées par H. T. Kung. Il est nécessaire d'introduire un certain nombre de cellules de retard d'un bit entre les cellules pour synchroniser les flots de données traversant le convolveur.

La solution retenue a consisté à associer la première approche de H. T. Kung pour les cellules et la première solution de multiplieur-additionneur pour les microcellules (fig. 9). Cette solution, originale à notre connaissance, a été retenue pour son faible nombre de portes logiques comparée aux autres solutions. En particulier les microcellules du « multiplieur-additionneur » fonctionnent sur niveaux d'horloge et non sur front (deux cellules contiguës fonctionnent sur des niveaux inverses), ce qui permet d'utiliser des bascules très simples, contrairement à la deuxième solution de multiplieur qui nécessite des bascules maître-esclave plus importantes en portes.

Il est nécessaire, pour synchroniser les différentes données d'intercaler 27 retards d'un demi-cycle d'horloge entre les cellules élémentaires.

Dans une image numérisée en 512×512 pixels codés sur 256 niveaux de gris, le convolveur nécessite à chaque période de 100 ns le chargement en parallèle dans les registres de sérialisation de trois pixels appartenant à trois lignes d'images différentes. Dans le même temps un résultat d'un calcul précédent est obtenu sur 16 bits dans le registre de parallélisation. Extérieurement le circuit fonctionne à la fréquence

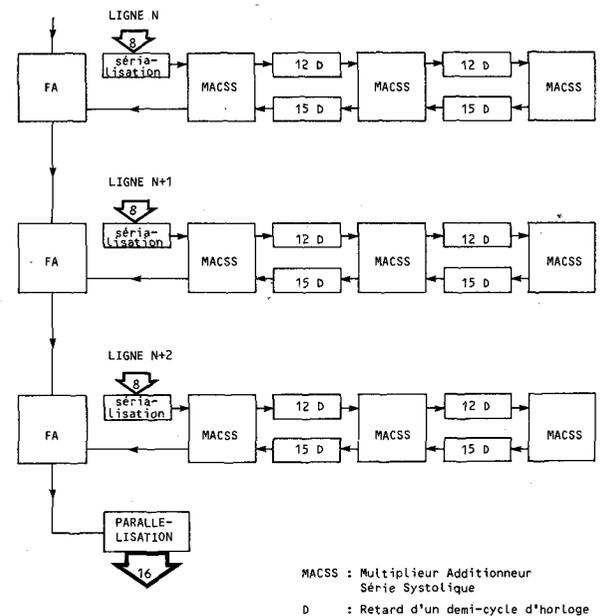


Fig. 9. — Schéma complet avec registre d'entrée-sortie.

d'échantillonnage (10 MHz). Mais intérieurement, le calcul s'effectuant en série sur 16 bits, la fréquence de fonctionnement interne est de 160 MHz ce qui nécessite l'utilisation d'une technologie type ECL.

L'intégration de ce processeur de convolution nécessite environ 2500 portes logiques équivalentes. Le circuit correspondant a été entièrement validé et est en cours d'industrialisation chez un fondeur de silicium.

Conclusion

Cette architecture de convolveur 2D par un masque 3×3 est facilement intégrable en raison de sa structure « microcellulaire » répétitive et de par la simplicité de la « microcellule » équivalente à un additionneur un bit.

Remerciements

Nous remercions M. Claude Bozzo, Directeur Scientifique de la CSEE, pour les conseils qu'il nous a prodigués pour la rédaction de cet article et pour les additifs et compléments qu'il y a apportés.

Nous tenons également à remercier MM. Michel Soufflet, Michel Dumas et Edouard Siffredi (CSEE/CETIA) pour leur aide précieuse qui nous a permis de faire aboutir cette étude.

APPLICATIONS

BIBLIOGRAPHIE

- [1] H. T. KUNG, *Why Systolic Architecture Technical Report*, 1981, Carnegie Mellon University.
- [2] H. T. KUNG et S. W. SONG, *A Systolic 2D Convolution Chip*, 1981, Carnegie Mellon University.
- [3] J. F. ABRAMATIC, *Approximation des filtres bi-indiciels et traitement des images, Thèse d'état*, INRIA, Rocquencourt.

Annexe

Quelques filtres utilisés couramment en traitement d'images

1. FILTRES PASSE-BAS

Les *filtres passe-bas* utilisés pour « nettoyer » l'image afin de réduire l'effet du bruit sont en général les suivants :

$$H_1 = \frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix},$$

$$H_2 = \frac{1}{10} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix},$$

$$H_3 = \frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}.$$

2. FILTRES PASSE-HAUT

Les *filtres passe-haut* sont utilisés pour accentuer les contours d'une image qui sont à nature haute fréquence. Il est recommandé de n'utiliser ces filtres que pour des images dont le rapport signal à bruit est assez grand :

$$H_4 = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix},$$

$$H_5 = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 9 & -1 \\ -1 & -1 & -1 \end{bmatrix},$$

$$H_6 = \begin{bmatrix} 1 & -2 & 1 \\ -2 & 5 & -2 \\ 1 & -2 & 1 \end{bmatrix}.$$

3. FILTRES LAPLACIENS

Les masques des *filtres laplaciens* sont utilisés pour l'extraction des contours pour décorréler une image

afin d'analyser la texture et également pour accentuer les pics de corrélation afin de recaler deux images :

$$H_7 = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix},$$

$$H_8 = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix},$$

$$H_9 = \begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{bmatrix}.$$

4. FILTRES DE GRADIENT

D'autres *filtres de dimensions 3 × 3* sont utilisés pour exploiter des primitives comme des segments :

$$H_{10} = \begin{bmatrix} -1 & 2 & -1 \\ -1 & 2 & -1 \\ -1 & 2 & -1 \end{bmatrix},$$

$$H_{11} = \begin{bmatrix} -1 & -1 & -1 \\ 2 & 2 & 2 \\ -1 & -1 & -1 \end{bmatrix},$$

$$H_{12} = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix},$$

$$H_{14} = \begin{bmatrix} -1 & -1 & 2 \\ -1 & 2 & -1 \\ 2 & -1 & -1 \end{bmatrix}.$$

Les *filtres* les plus fréquemment utilisés *en vision* sont les filtres qui permettent d'extraire les contours.

En effet, un opérateur de contour est formé de au moins deux filtres de dimension 3 × 3 suivi d'un opérateur de norme :

Opérateur de Prewitt

$$H_{14} = \begin{bmatrix} +1 & 0 & -1 \\ +1 & 0 & -1 \\ +1 & 0 & -1 \end{bmatrix}, \quad H_{15} = \begin{bmatrix} +1 & +1 & +1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}.$$

Opérateur de Sobel

$$H_{16} = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & 2 \\ 1 & 0 & -1 \end{bmatrix}, \quad H_{17} = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}.$$

Opérateur de Chen et Frei

$$\begin{aligned}
 H_{18} &= \begin{bmatrix} 1 & 0 & -1 \\ \sqrt{2} & 0 & -\sqrt{2} \\ 1 & 0 & -1 \end{bmatrix}, & H_{19} &= \begin{bmatrix} 1 & \sqrt{2} & 1 \\ 0 & 0 & 0 \\ -1 & -\sqrt{2} & -1 \end{bmatrix}, \\
 H_{20} &= \begin{bmatrix} \sqrt{2} & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & -\sqrt{2} \end{bmatrix}, & H_{21} &= \begin{bmatrix} 0 & -1 & \sqrt{2} \\ 1 & 0 & -1 \\ -\sqrt{2} & 1 & 0 \end{bmatrix}, \\
 H_{22} &= \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}, & H_{23} &= \begin{bmatrix} -1 & 0 & +1 \\ 0 & 0 & 0 \\ 1 & 0 & -1 \end{bmatrix}, \\
 H_{24} &= \begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{bmatrix}, & H_{25} &= \begin{bmatrix} -2 & 1 & -2 \\ 1 & 4 & 1 \\ -2 & 1 & -2 \end{bmatrix}.
 \end{aligned}$$

Opérateur de Kirsch

$$H_{26} = \begin{bmatrix} 5 & 5 & 5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix}.$$

Les autres masques de filtrage s'obtiennent par permutation circulaire du masque H_{26} .

Manuscrit reçu le 25 juin 1984.