

# Localisation et Reconnaissance de visages : Vers une implantation sur silicium

Dominique GINHAC, Eri Prasetyo, Michel Paindavoine  
Laboratoire LE2I - UMR CNRS 5158  
Université de Bourgogne  
21078 Dijon Cedex - FRANCE

E-mail : (dginhac,prasetyo,paindav)@u-bourgogne.fr

**Abstract**— Depuis de nombreuses années, le laboratoire LE2I de l'Université de Bourgogne s'est intéressé à la mise en œuvre de systèmes électroniques et informatiques dédiés à la localisation et la reconnaissance de visages en temps réel.

Ce papier présente les premiers travaux relatifs au développement microélectronique d'un capteur CMOS intelligent embarqué dédié à l'extraction de paramètres pour ce type d'applications. L'objectif, ici, est de définir les concepts architecturaux retenus pour la conception du capteur CMOS dédié.

**Keywords** : *Capteurs CMOS intelligents - Systèmes embarqués - Traitement d'images Temps réel*

## I. INTRODUCTION

La mise en œuvre d'un système de localisation et de reconnaissance de visages en temps réel permet d'envisager dans un avenir très proche de multiples applications dans les domaines des interfaces homme machine, de la vidéo-conférence, de la téléphonie mobile ou du contrôle d'accès sécurisé par exemple.

Définir un tel système répondant à des critères d'embarquabilité (faible volume, consommation maîtrisée), de fiabilité (taux de reconnaissance très élevé dans n'importe quel environnement) et de rapidité (reconnaissance temps réel d'une ou plusieurs personnes en mouvement dans une scène) reste encore un problème très ouvert de nos jours.

Notre approche s'inscrit dans une telle problématique et a pour objectif principal le développement d'une architecture matérielle et logicielle reposant sur l'utilisation d'un réseau de neurones de type RBF (Radial Basis Function).

Ce papier vise à présenter les premiers travaux relatifs à la création d'un capteur d'images – de type CMOS – dédié à l'extraction des paramètres destinés au réseau de neurones.

## II. PRÉSENTATION DU SYSTÈME GLOBAL DE RECONNAISSANCE

Le traitement automatique d'images de visages a été abordé de multiples manières dans la littérature et, en particulier à l'aide de réseaux de neurones artificiels [1][2]. Les réseaux de type RBF donnent des résultats très satisfaisants dans les applications de reconnaissance de visage[3][4]. De plus, il a été montré dans [5] que les classifieurs de type RBF possèdent une grande capacité de généralisation permettant l'identification des visages de 3/4 profil ou de profil. En se basant sur ces recherches antérieures, nous avons développé un système de vision temps réel de localisation de visages dans des séquences vidéo et de vérification de leur identité[6]. L'architecture du système est présentée sur la figure 1.

Pour réaliser la phase de détection globale, l'image est découpée en imagettes successives de taille 40 x 32 pixels (balayage de l'image avec un masque). Le module de pré-traitement est chargé de transformer chacune des imagettes en vecteur d'entrée du réseau RBF. L'extraction de caractéristiques se fait sous la forme d'un moyennage des ensembles de quatre pixels consécutifs sur les lignes des imagettes, ce qui donne un total de 320 composantes en entrée du réseau de neurones. L'architecture du réseau RBF est prédéfinie par le nombre de personnes à vérifier, avec une structure modulable. Enfin, le module de décision détermine la présence, la position, l'identité et l'échelle des visages en cherchant les points maxima parmi les réponses du réseau.

## III. INTÉRÊT D'UN CAPTEUR DÉDIÉ

Les deux réalisations matérielles à base de FPGA présentées dans [6] ont permis de mettre en évidence que la phase d'extraction des caractéristiques des imagettes nécessite un grand nombre d'opérations élémentaires de

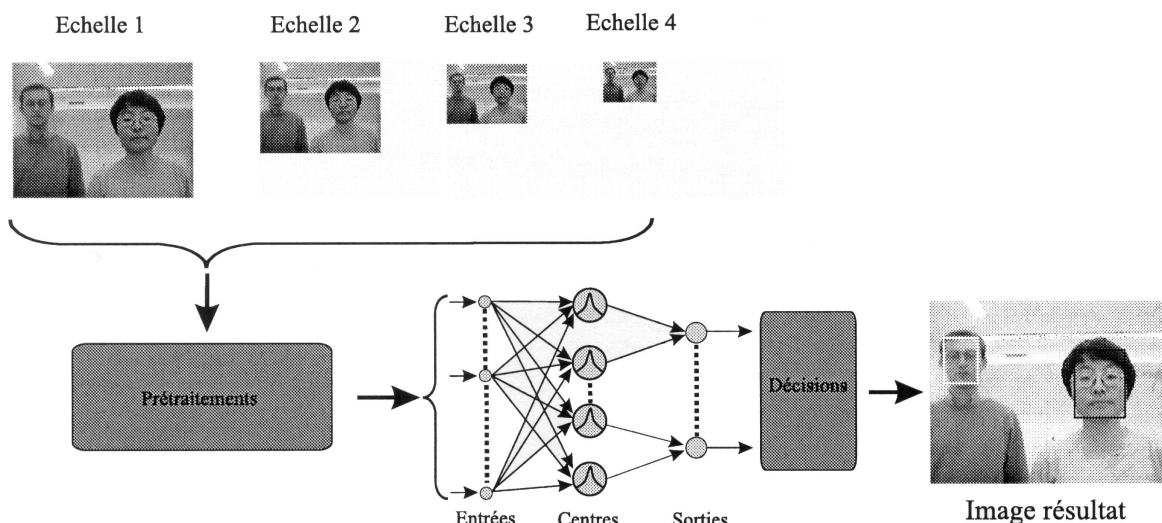


Fig. 1. Architecture de localisation et de reconnaissance de visages

calcul. De fait, l'idée retenue est de délocaliser l'extraction des vecteurs à analyser par le réseau de neurones directement sur le capteur d'acquisition des images. Celui-ci ne se contente plus de fournir au système de traitement un flot d'images brutes mais des informations de plus haut niveau sous la forme d'un ensemble de valeurs pour une position donnée de la fenêtre d'intérêt à évaluer. Pour cela, il est nécessaire d'intégrer au plus près de la zone photosensible du pixel un ensemble de transistors capables de réaliser la fonction désirée[7][8].

Le capteur à mettre en œuvre doit posséder certaines fonctions programmables rendant son utilisation la plus large possible. On peut citer par exemple :

- la taille d'acquisition des fenêtres d'intérêt doit être variable car il semble indispensable de pouvoir traiter des images de taille programmable (et non pas uniquement de taille 40 x 32 pixels).
- la position de la fenêtre d'intérêt doit être variable afin de pouvoir effectuer un balayage automatique de l'image.
- la résolution doit être variable sur les lignes et les colonnes de l'image afin de pouvoir soit traiter individuellement un ensemble de pixels consécutifs, soit réaliser un sous-échantillonnage à différentes échelles dans le but de diminuer le volume d'informations.

#### IV. ARCHITECTURE DU CAPTEUR

##### A. Choix et principe de fonctionnement d'un pixel

Le capteur consiste principalement en une matrice carrée de pixels, chacun constitué d'une photodiode et de

4 transistors de commande fonctionnant en interrupteurs (cf. fig 2). Ce type de pixel est communément employé pour la réalisation de capteurs APS[9].

La phase de lecture d'un pixel comprend 4 phases principales :

- La fermeture des transistors R de reset et L du contrôle du temps d'obturation permet d'imposer la tension  $V_{Reset}$  aux bornes de la photodiode,
- L'ouverture du transistor de Reset permet l'intégration du courant généré par la photodiode dans la capacité parasite de celle-ci,
- En fin d'intégration, l'ouverture du transistor L permet de geler le signal  $V_{Pix}$  sur la grille du transistor T,
- La fermeture du transistor S de sélection permet d'acheminer le courant  $I_{Col}$  dépendant de la tension aux bornes de la diode vers l'amplificateur situé en bout de chaque colonne.

##### B. Choix de l'architecture

L'architecture comprend plusieurs blocs fonctionnels interconnectés :

- la matrice composée d'un ensemble de pixels — présentés au paragraphe précédent — chargée de faire l'acquisition des images
- les décodeurs de ligne chargés d'envoyer à chaque ligne de la matrice les signaux de commande permettant de réaliser l'acquisition.
- les amplificateurs de colonnes qui sont chargés d'amplifier le signal issu des pixels sur la colonne correspondante,

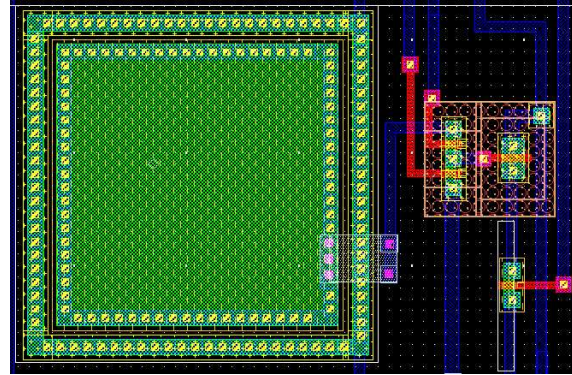
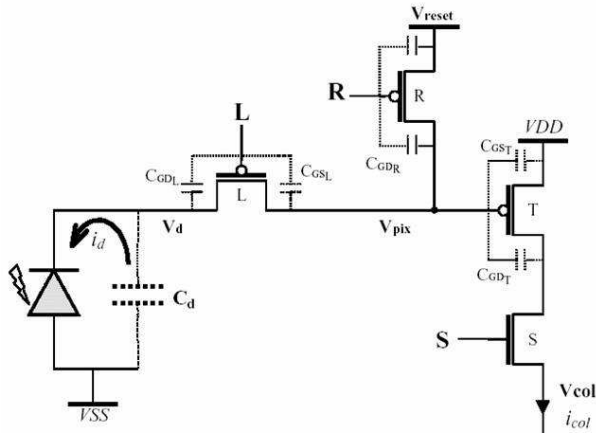


Fig. 2. Schéma et Layout d'un pixel à photodiode

- les convertisseurs analogiques numériques chargés de transformer le signal analogique en signal numérique sur chacune des colonnes de la matrice,
- la mémoire RAM sur 8 bits chargée de stocker les valeurs produites par les convertisseurs,
- les opérateurs de calcul chargés d'évaluer par exemple les valeurs moyennes de plusieurs pixels consecutifs d'une même ligne.
- l'électronique de commande qui séquence l'ensemble.

Cette architecture, relativement générique, doit être adaptée afin de satisfaire pleinement le cahier des charges. En particulier, l'aspect massivement parallèle des traitements à effectuer nous permet d'envisager un fonctionnement non pas uniquement sur des pixels mais sur des lignes compètes (ou du moins sur des ensembles de pixels consecutifs correspondant à la largeur des fenêtres d'intérêt). L'adressage des pixels se faisant par ligne — c'est-à-dire que les signaux de commande sont envoyés à tous les pixels d'une même ligne —, nous pouvons donc récupérer simultanément les valeurs analogiques d'une même ligne en sortie des amplificateurs de colonnes. Ainsi, le choix de réaliser un convertisseur analogique numérique par colonne de type sigma-delta plutôt qu'un seul convertisseur (de type flash par exemple) nous donne l'opportunité d'obtenir rapidement une ligne de valeurs numériques qui seront stockées dans une ligne de cellules de mémoire RAM statique à 6 transistors par bit.

## V. PREMIERS RÉSULTATS

Pour démontrer la faisabilité de l'approche retenue dans le cadre de ce projet, nous avons mis l'accent sur deux aspects majeurs à savoir d'une part la cellule

élémentaire du pixel analogique composée de la photodiode et de ses transistors de commande et d'autre part sur la structure du convertisseur analogique numérique.

Le premier aspect a conduit à la réalisation d'un circuit test en technologie AMS CMOS  $0.6 \mu$  comportant 2 matrices élémentaires de  $8 \times 8$  pixels chacune avec leur logique complète de commande permettant d'accéder à n'importe quelle fenêtre d'intérêt (cf. figure 3). A l'heure actuelle, des séries de mesures sont effectuées sur ce prototype permettant à très court terme de valider les solutions retenues pour la conception des pixels analogiques.

Le deuxième aspect est en cours d'étude pour l'instant. L'objectif majeur est ici de pouvoir réaliser le meilleur convertisseur analogique numérique tout en étant fortement contraint par la surface (largeur du convertisseur identique à la largeur d'un pixel élémentaire). Le choix d'un convertisseur de type sigma-delta semble approprié à notre problématique et est en cours de validation par simulation.

## VI. CONCLUSION ET PERSPECTIVES

Dans ce papier, nous avons présenté les fondements d'une rétine analogique CMOS dédiée à l'extraction des paramètres destinés au réseau RBF. Notre objectif majeur est de diminuer fortement la charge de travail de l'architecture de calcul effectuant la phase de reconnaissance de visages.

L'approche proposée repose sur le développement d'un capteur CMOS intégré réalisant successivement la phase d'acquisition de zones d'intérêt dans l'image, l'amplification et la conversion numérique des signaux issus des pixels puis leur traitement dans le but de produire un flot temps réel de caractéristiques qui serviront de point d'entrée d'un réseau de neurones.

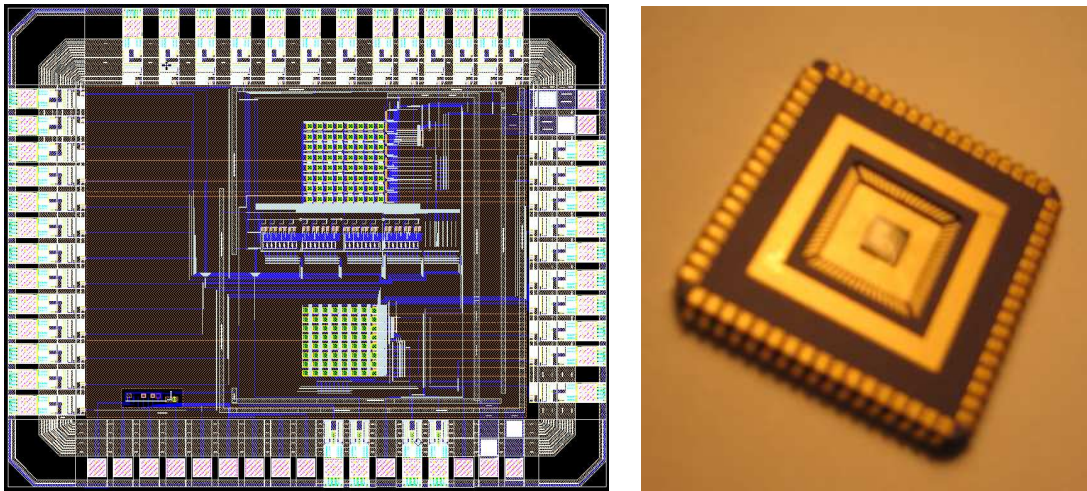


Fig. 3. Layout du circuit complet et circuit réalisé

Les deux points clés (cellule élémentaire d'acquisition et convertisseur analogique numérique) sont en cours de validation. Les poursuites à court terme concernent le dessin complet du convertisseur et de la mémoire associée et plus généralement l'intégration des différents blocs composant le système global.

#### REFERENCES

- [1] R. Férand and O. Bernier, "A fast and accurate face detector based on neural network," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 1, pp. 42–53, Jan 2001.
- [2] N. Intrator, D. Reisfeld, and Y. Yeshurun, "Face recognition using a hybrid supervised/unsupervised neural network," *Pattern Recognition Network*, vol. 17, no. 1, pp. 67–76, Jan 1996.
- [3] M. Rosenblum, Y. Yacoob, and L. David, "Human expression recognition from motion using a radial basis function network architecture," *IEEE Transactions On Neural Network*, vol. 7, no. 5, pp. 1121–1138, 1996.
- [4] A. Howell and H. Buxton, "Learning identity with radial basis function networks," *Neurocomputing*, vol. 20, pp. 15–34, 1998.
- [5] A. O'Toole, S. Edelman, and H. Bulthoff, "Stimulus-specific effects in face recognition from motion using a radial basis function network architecture," *Vision research*, vol. 38, pp. 2351–2363, 1998.
- [6] N. Malasné, "Localisation et reconnaissance de visages en temps réel : algorithmes et architectures," Ph.D. dissertation, Université de Bourgogne, Nov 2002.
- [7] A. Moini, *Vision Chips*. Kluwer Academic Publishers, 2000.
- [8] R. Nixon, S. Kemeny, C. Staller, and E. Fossum, "128x128 cmos photodiode-type active pixel sensor with on-chip timing, control and signal chain electronics," *SPIE Charge Coupled Devices and Solid State Optical Sensors*, vol. 2415, no. 34, 1995.
- [9] C. H. Aw and B. Wooley, "A 128x128 pixel standard cmos image sensor with electronic shutter," *IEEE Journal of Solid State Circuits*, vol. 31, no. 12, pp. 1922–1930, 1996.