

Capteur de vision avec extraction du contraste, de l'orientation et du mouvement local

FRANÇOIS KAESS¹, PASCAL HEIM¹, PIERRE-FRANÇOIS RÜEDI¹, STEVE GYGER¹

¹ Sensory Information Processing
Centre Suisse d'Electronique et de Microtechnique
Jaquet-Droz 1, 2002 Neuchâtel, Suisse

¹Francois.Kaess@csem.ch, Pascal.Heim@csem.ch,
Pierre-Francois.Ruedi@csem.ch, Steve.Gyger@csem.ch

Résumé – Cette communication présente un capteur de 128×128 pixels qui extrait en temps réel l'amplitude des contrastes de la scène observée, leur orientation et l'information de mouvement local, par un calcul analogique parallèle effectué dans le plan focal, avec une excellente stabilité de représentation dans une gamme dynamique de 130dB. En outre, la sortie des informations est ordonnée sous forme de liste par ordre d'amplitude décroissante des caractéristiques, ce qui permet de construire des systèmes de traitement en temps réel extrêmement réactifs, car ne prenant en compte que l'information la plus pertinente. Ce capteur est particulièrement approprié pour des applications où les conditions d'illumination ne sont pas contrôlées ou changent rapidement, par exemple pour des applications de sécurité, automobile et domotique.

Abstract - This paper describes a 128 by 128 pixel vision sensor which computes at the pixel level the magnitude and direction of image contrasts and local motion information independently of the illumination level. It is able to capture a scene with a dynamic range of 130 dB. A temporal ordering of this information according to the contrast or motion magnitude is used to reduce the amount of data delivered and facilitate subsequent processing. This sensor is ideally suited for applications where illumination conditions are not under control or rapidly changing, such as automotive, security and home automation applications.

1 Introduction

Il existe une demande croissante pour des systèmes de vision en temps réel robustes, autonomes, fonctionnant dans des environnements non-contrôlés et de coût modéré. Prises ensemble, ces contraintes semblent difficiles à remplir simultanément par les systèmes de prise et de traitement d'images conventionnels. Une difficulté significative consiste à extraire uniquement les éléments caractéristiques nécessaires à la réalisation d'une tâche donnée et d'en obtenir une représentation stable indépendante des variations d'illumination.

Une approche possible [1-7] consiste à intégrer une partie du traitement du signal directement dans le capteur, dans la matrice de pixels, où le traitement analogique parallèle permet d'exploiter au mieux les données à traiter. Dans [7], nous avons décrit un capteur de vision qui calcule analogiquement la norme et la phase du contraste au niveau des pixels. Sa grande dynamique est obtenue par le biais d'un système d'auto-exposition appliqué à chaque pixel séparément et la représentation obtenue est particulièrement stable puisque le contraste est indépendant des variations d'illumination et représente les différences de réflectance entre les pixels voisins. Récemment, d'autres capteurs ont été proposés mettant en œuvre des stratégies similaires : le capteur [8] extrait de façon asynchrone la dérivée temporelle du logarithme de la luminance au niveau des pixels, ce qui restitue une information de mouvement, mais pas le contraste spatial

ni l'orientation, alors que le capteur [9] détecte uniquement le contraste spatial de façon binaire.

Cette communication décrit un circuit qui ajoute aux précédentes caractéristiques l'estimation de la dérivée temporelle de l'amplitude du contraste, représentative du mouvement local. Cette information additionnelle est utilisée pour ordonner temporellement la sortie des caractéristiques de l'image par ordre d'amplitude de mouvement décroissante, ce qui présente deux avantages : d'une part, l'information considérée comme la plus significative est restituée en priorité et d'autre part, cela permet de limiter la quantité de données à traiter en interrompant la sortie lorsque l'information est suffisante pour l'exécution d'une tâche donnée. En outre, les performances de cette nouvelle génération ont été significativement améliorées en termes de dynamique et de consommation.

2 Principe de fonctionnement

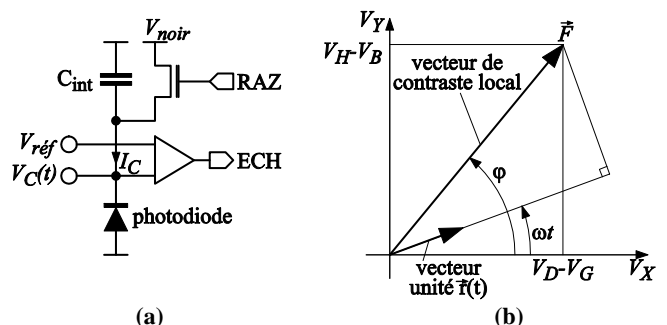


Figure 1 : principe d'intégration du photocourant (a) et de calcul de l'amplitude et de la direction du contraste (b).

L'objectif du traitement effectué en parallèle au niveau de chaque pixel est de construire trois cartes de la scène observée : contraste, orientation et mouvement.

Dans chaque pixel C, le photocourant I_C est intégré sur un condensateur C_{int} . La tension résultante $V_C(t)$ est comparée à une tension de référence $V_{réf}$ et distribuée aux pixels voisins. Lorsqu'elle atteint cette tension de référence, le pixel échantillonne les tensions V_D , V_G , V_H et V_B reçues des pixels voisins et les mémorise dans des condensateurs. Le temps d'intégration est donné par :

$$T_S = \frac{C_{int} \cdot V_{réf}}{I_C} \quad (1)$$

Si I_N est le photocourant d'un pixel voisin N, la tension correspondante échantillonnée vaut :

$$V_N = \frac{I_N \cdot T_S}{C_{int}} = \frac{I_N}{I_C} V_{réf} \quad (2)$$

Cette tension est indépendante du niveau d'illumination et dépend uniquement des rapports des photocourants générés dans les pixels N et C, à la tension constante $V_{réf}$ près. En faisant l'hypothèse que le photocourant du pixel central C correspond à la moyenne de ses quatre voisins, ce qui est approximativement vérifié localement compte-tenu des propriétés de filtrage passe-bas de l'optique, les tensions $V_D - V_G$ et $V_H - V_B$ représentent les composantes X et Y du vecteur de contraste local.

Le calcul de la norme et de la phase de ce vecteur est réalisé en le projetant sur un vecteur unité tournant comme illustré sur la figure 1(b). Le résultat de cette projection est donné par le produit scalaire suivant :

$$C(t) = \vec{F} \cdot \vec{r}(t) = (V_D - V_G) \cos \omega t + (V_H - V_B) \sin \omega t \quad (3)$$

où $\cos \omega t$ et $\sin \omega t$ sont deux fonctions distribuées à tous les pixels en parallèle qui correspondent aux composantes du vecteur unité \vec{r} . Lors de la première période de rotation, la fonction atteint un maximum correspondant à la norme du contraste. Cette valeur est détectée et mémorisée au sein du pixel par un détecteur de maximum. En notant $\vec{C}(t)$ et $\vec{C}(t - \Delta t)$ les vecteurs de contraste respectivement estimés lors des prises d'image courante et précédente, la dérivée temporelle de la norme du contraste est donnée par :

$$\frac{\partial |\vec{C}(t)|}{\partial t} \approx \frac{|\vec{C}(t)| - |\vec{C}(t - \Delta t)|}{\Delta t} \quad (4)$$

La phase φ est encodée temporellement par le passage par zéro de la fonction de contraste $C(t)$ avec une pente positive, ce qui advient à $\omega t = \varphi_0$. En notant que le passage par zéro est décalé de -90° par rapport au maximum de $C(t)$, la phase effective vaut : $\varphi = \varphi_0 + 90^\circ$

3 Mise en œuvre

Le schéma de principe du pixel est présenté sur la figure 2. Le bloc d'intégration du photocourant et échantillonnage est constitué d'une photodiode, d'un condensateur d'intégration C_{int} , d'un circuit de remise à zéro et d'un comparateur (figure 1(a)). Au début de

chaque prise d'image le signal RAZ est porté à l'état haut afin d'initialiser la tension d'intégration $V_C(t)$ à une tension de référence V_{noir} . Après cette phase initiale, le photocourant est intégré sur le condensateur C_{int} et la tension résultante $V_C(t)$ est comparée à une valeur de référence $V_{réf}$. Au début de l'intégration le signal ECH est à l'état haut. Lorsque $V_C(t)$ atteint $V_{réf}$, le comparateur bascule, le signal ECH passe à l'état bas, ce qui termine l'échantillonnage des tensions V_D , V_G , V_H et V_B fournies par les pixels voisins.

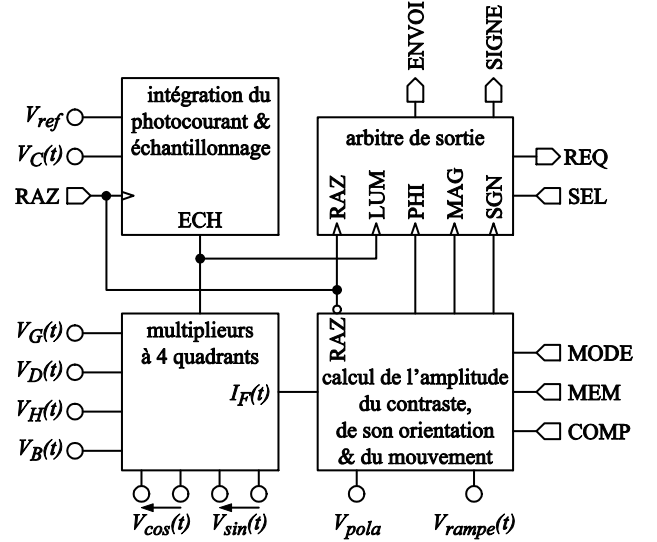


Figure 2 : schéma de principe du pixel.

Deux multiplieurs à quatre quadrants [10] réalisent la multiplication de ces tensions par des fonctions cosinus et sinus générées sous la forme de tensions différentielles $V_{cos}(t)$ et $V_{sin}(t)$. La sortie des multiplieurs est un courant sinusoïdal $I_F(t)$ représentant la fonction $C(t)$ décrite dans l'équation (3).

Ce courant est fourni au bloc d'extraction du contraste, de l'orientation et du mouvement local, illustré sur la figure 3. Ce bloc comprend trois comparateurs de courant, un détecteur de maximum, une cellule de mémorisation analogique et un redresseur.

Les transistors M_{P1} , M_{N1} , M_{N2} et M_{P14} constituent un comparateur de courant [11]. Lorsque le courant $I_F(t)$ est positif, le nœud PHI se trouve à l'état bas, ce qui rend M_{P14} passant et bloque M_{N2} . Le courant $I_F(t)$ est alors collecté par V_{SS} à travers M_{P14} . Lorsque $I_F(t)$ est négatif, le nœud PHI est à l'état haut, ce qui bloque M_{P14} et rend M_{N2} passant. Le courant $I_F(t)$ est alors collecté depuis V_{DD} à travers M_{P5} et M_{N1} .

La détection de maximum est réalisée par les transistors M_{P2} , M_{P5} - M_{P10} et le condensateur C_{max} . Au début de chaque prise d'image, le condensateur C_{max} est déchargé en passant temporairement le signal RAZ à l'état bas. Puis, si le courant $I_{crête}$ passant à travers M_{P5} augmente, M_{P9} est passant et C_{max} est chargée jusqu'à ce qu' I_{max} soit égal à $I_{crête}$. Si $I_{crête}$ diminue, M_{P9} se bloque et I_{max} est mémorisé et représente la norme du contraste pour l'image courante.

La mémorisation de l'amplitude du contraste de la prise d'image précédente est réalisée par les transistors M_{N3} - M_{N6} ainsi que le condensateur $C_{préc}$.

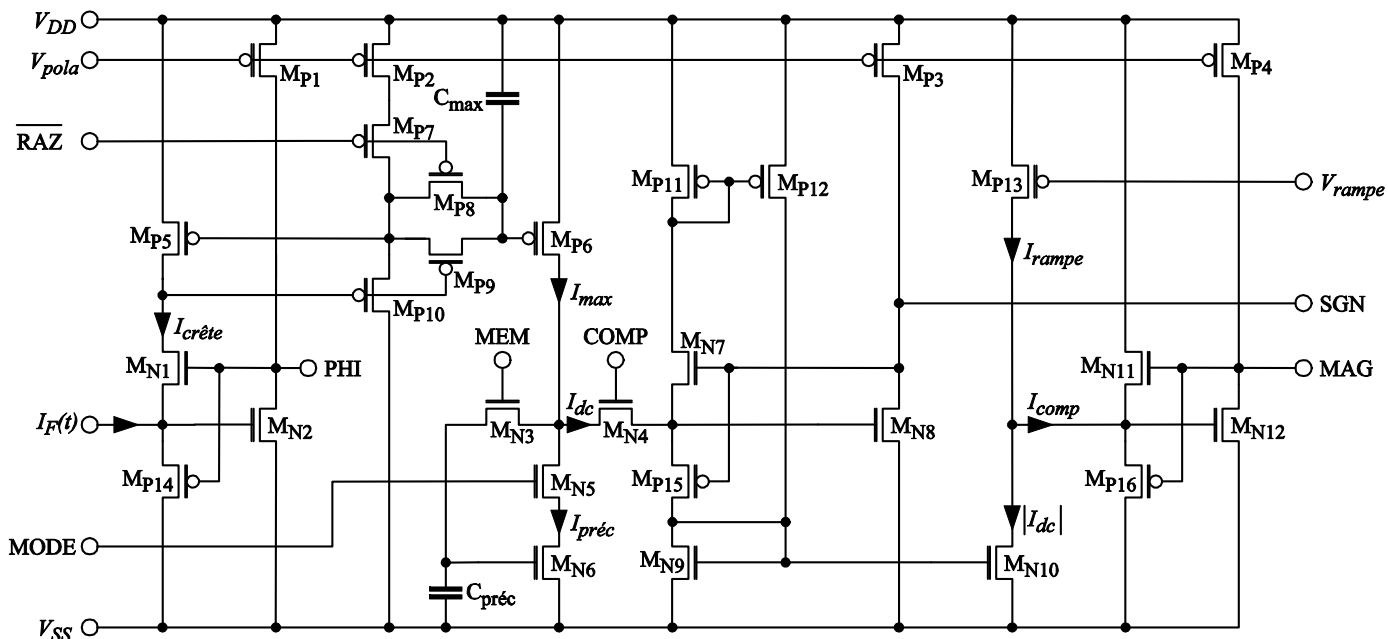


Figure 3 : schéma du circuit d'extraction du contraste, de l'orientation et du mouvement local.

Considérant que le signal de commande MODE est à l'état haut à la fin de la prise d'image précédente et que le signal COMP est à l'état bas alors que MEM est haut, le condensateur $C_{préc}$ est chargé jusqu'à ce que le courant $I_{préc}$ égale I_{max} . Puis MEM est mis à l'état bas avant de redémarrer une nouvelle phase d'intégration. Lors de la phase de calcul correspondant à la prise d'image courante, COMP est placé à l'état haut et le courant I_{dc} représentant la dérivée de la norme du contraste est donné par :

$$I_{dc} = I_{max} - I_{préc} \quad (6)$$

Ce courant I_{dc} est comparé à zéro par un deuxième comparateur de courant (M_{P3} , M_{N7} , M_{N8} et M_{P15}). Associés au comparateur de courant précité, les transistors M_{P11} , M_{P12} , M_{N9} et M_{N10} réalisent un redresseur de courant : si I_{dc} est positif, il passe à travers M_{N9} . Dans le cas contraire, il est d'abord redressé par le miroir de courant M_{P11} et M_{P12} avant de passer à travers M_{N9} . Dans les deux cas, la valeur absolue du courant I_{dc} est recopiée par M_{N10} , tandis que la sortie SGN du comparateur en indique le signe.

Le bloc de contrôle du circuit génère une fonction de seuil monotone décroissante qui est convertie en une tension de seuil V_{ramp} par un convertisseur numérique-analogique 8 bits. Cette tension est convertie dans chaque pixel en un courant correspondant I_{ramp} . La valeur absolue du courant I_{dc} est soustraite de ce courant I_{ramp} . Le courant résultant I_{comp} est comparé à zéro par un troisième comparateur de courant (M_{P4} , M_{N11} , M_{N12} et M_{P16}). La valeur de la dérivée de l'amplitude du contraste est définie par l'instant où la sortie MAG de ce comparateur passe à l'état haut.

Le séquençage approprié des signaux de commande MODE, MEM et COMP permet de sortir soit l'amplitude du contraste, soit la valeur absolue de la différence d'amplitude du contraste ainsi que son signe, soit les deux informations à chaque prise d'image.

Au niveau du circuit global, la sortie des données est arbitrée de manière asynchrone par un double système de jetons vertical et horizontal afin d'éviter collisions et pertes d'information. Lorsqu'un pixel est prêt à sortir l'information requise, il active le signal REQ sur une ligne horizontale de la matrice. Quand le jeton vertical atteint cette ligne, il interrompt sa circulation et autorise la sortie des données de cette ligne en activant le signal SEL. A la réception de cet acquittement, les pixels de la ligne ayant une donnée valide à transmettre activent le signal ENVOI, propagé verticalement le long des colonnes de la matrice. Le jeton horizontal commence alors à circuler en s'arrêtant à chaque fois que le signal ENVOI est activé. Les adresses des ligne et colonne respectives ainsi que la valeur courante codée sur 8 bits de l'information considérée (contraste, orientation ou mouvement) constituent un événement, qui est ajouté à la liste de sortie du capteur.

4 Résultats

La figure 4 représente la photographie du capteur qui comprend une matrice de 128×128 pixels intégré dans une technologie CMOS standard ($0.5 \mu\text{m}$). Le facteur de conversion est de $24 \mu\text{V}/e^-$. A température ambiante, le courant noir est de $1 \text{ nA}/\text{cm}^2$. Les caractéristiques du capteur sont résumées dans le tableau 1.

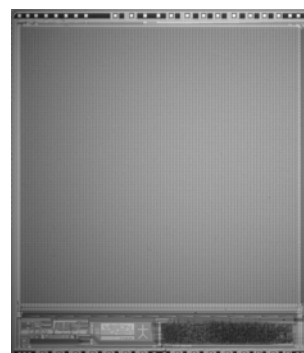


Figure 4 : photographie du capteur.

Tableau 1 : Spécifications du circuit

Technologie	0.5 μm , 3 métaux, 2 poly
Taille du pixel / facteur de remplissage	64 \times 64 μm^2 / 7.4 %
Taille du circuit	9.2 \times 10.6 mm ²
Temps d'intégration T _i (25°C)	1 μs ... 100 ms
Gamme dynamique (25°C)	130 dB
Calcul et sortie des cartes caractéristiques (T _c)	< 3.5 ms / carte
Cadence des prises d'images	1/(T _i + T _c) (max 250 images /s)
Sensibilité de l'estimation du contraste	2 %
Sensibilité de l'estimation du mouvement	3 %
Précision du calcul de l'orientation	$\pm 3^\circ$
Consommation	150 mW @ 3.3 V
Température de fonctionnement	-40°C à +85°C

Le circuit est contrôlé par une interface sérielle permettant d'ajuster les modes et paramètres de fonctionnement. Les données sont sorties par une interface parallèle, chaque caractéristique (contraste, orientation, mouvement) peut être sortie successivement et indépendamment à chaque prise d'image.

La gamme dynamique d'intégration (100 dB) est définie par les temps d'intégration minimal et maximal que le capteur peut mesurer. Le temps d'intégration minimal est limité par le délai de réponse du comparateur des pixels, alors que le temps d'intégration maximal est limité par le courant noir des photodiodes qui limite le rapport signal sur bruit pour les faibles illuminations. La dynamique de calcul du contraste, estimée à 30 dB, découle directement de la mesure de contraste avec une précision de 2%, qui ne dépend pas de l'illumination. La gamme dynamique résultante totale atteint ainsi 130 dB à température ambiante. A température plus élevée, elle est dégradée par l'augmentation du courant noir (bruit thermique) des photodiodes.

La figure 5 illustre les cartes de contraste (a) et de mouvement (b) extraites en temps réel par le capteur lors de l'observation d'une scène d'intérieur. La comparaison des deux représentations montre que l'information de mouvement est extraite efficacement de l'arrière-plan immobile.

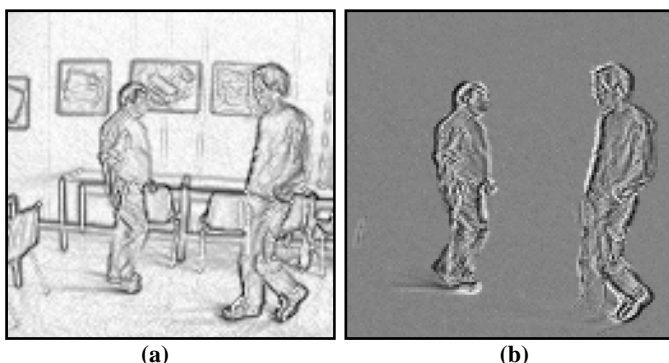


Figure 5 : représentations de l'amplitude du contraste (a) et de l'information de mouvement (b).

5 Conclusion

Cette communication décrit un capteur de vision qui calcule directement dans le plan focal le contraste local avec une précision de 2%, l'orientation avec une précision de $\pm 3^\circ$, ainsi que l'information du mouvement local avec une précision de 3%.

Les représentations du contraste et du mouvement sont indépendantes du niveau d'illumination et le capteur présente une gamme dynamique totale de 130 dB, ce qui est particulièrement intéressant dans le cas du traitement de scènes dont les conditions d'illumination ne sont pas contrôlées, inhomogènes et changeantes (applications automobiles et domestiques notamment).

En outre, la sortie des données par ordre d'amplitude décroissante permet de limiter la quantité de données et donc de restreindre les besoins du système en puissance de calcul nécessaire au post-traitement.

Références

- [1] G. Andreou, R. C. Meitzler, K. Strohhahn, and K. A. Boahen, "Analog VLSI neuromorphic image acquisition and pre-processing systems," *Neural Network*, vol. 8, pp. 1323–1347, 1995.
- [2] B. E. Shi and T. Choi, "A Michelson contrast sensitive silicon retina", in *Proc. 8th Int. Conference on Neural Information Processing*, 2001.
- [3] B. E. Shi, "A low power orientation selective vision sensor," *IEEE Trans. Circuits and Systems-II: Analog and Digital Signal Processing*, vol. 47, no. 5, pp. 435–440, 2000.
- [4] A. Moini, *Vision chips*. Norwell, MA: Kluwer, 1999.
- [5] R. Etienne-Cummings, J. van der Spiegel, and P. Mueller, "A foveated silicon retina for two-dimensional tracking," *IEEE Trans. Circuits Syst. II*, vol. 47, pp. 504–517, 2000.
- [6] M. Barbaro, P.-Y. Burgi, A. Mortara, P. Nussbaum and F. Heitger: "A 100 x 100 pixel silicon retina for gradient extraction with steering filter capabilities and temporal output coding," *IEEE J. Solid-State Circuits*, vol. 37-2, pp. 160-172, 2002.
- [7] P.-F. Rüedi, P. Heim, F. Kaess, E. Grenet, F. Heitger, P.-Y. Burgi, S. Gyger, P. Nussbaum, "A 128 \times 128 pixels 120 dB dynamic range vision sensor chip for image contrast and orientation extraction," *IEEE J. Solid-State Circuits*, Vol. 38-12, pp. 2306-2317, Dec. 2003.
- [8] P. Lichtsteiner, C. Posch and T. Delbruck, "A 128 \times 128 120dB 15 μs Latency Asynchronous Temporal Contrast Vision Sensor," *IEEE J. Solid State Circuits*, Vol.43-2, pp. 566-576, Feb 2008.
- [9] M. Gottardi, N. Massari and S. A. Jawed, "A 100 μw 128 \times 64 Pixels Contrast-Based Asynchronous Binary Vision Sensor for Sensor Networks Applications," *IEEE J. Solid State Circuits*, Vol.44-5, pp. 1582-1592, May 2009.
- [10] K. Bult, "Analog CMOS square-law circuits," Ph.D. dissertation, Univ. Twente, Twente, The Netherlands, 1988.
- [11] A. Rodriguez-Vasquez, R. Dominguez-Castro, F. Medeiro, and M. Delgado-Restituto, "High-resolution current comparators: design and application to current-mode function generation," *Analog Integrated Circuits and Signal Processing*, vol. 7, pp. 149-164, 1995.