Bioinspired CMOS Photosensor Adaptation using Local Luminance Feedback

R. Carmona, J. Cuadri, C. Domínguez-Matas, G. Liñán, E. Roca and A. Rodríguez-Vázquez
Instituto de Microelectrónica de Sevilla-CNM-CSIC
Campus de la Universidad de Sevilla. Avda. Reina Mercedes s/n, 41012-Sevilla, Spain. E-mail: rcarmona@imse.cnm.es

ABSTRACT: Local adaptation of the photosensors gain permits the accommodation of a larger illumination range into the restricted dynamic range of the photodiodes, and, on the other side, contrast enhancement in a concurrent and inexpensive implementation. This paper presents a simple implementation for the local adaptive control of the integration time in CMOS photodiodes, based on the feedback of the already captured image.

1 Sensors gain control through integration time

The mammalian retina operates on the captured visual stimuli at early stages in the process of vision. Complex spatio-temporal processing encodes visual information into a reduced set of channels [1] to be delivered to the brain by the optic nerve. Contrarily to the spike-like coding of neural information found elsewhere, they are continuous-time analog waves [2]. The captured signals are promediated and the high-gain characteristics of the cones and the bipolar cells are relocated to adapt to light conditions. These operations have a local scope and depend on the recent history of the cells. Once adaptation is achieved, patterns of activity are formed dynamically. The equations governing the dynamic formation of these activity patterns has been succesfully mapped to a CNN Universal Machine architecture [3] and also implemented in CMOS technology [4].

Theoretical studies [5] point that, based on concurrent processing of local information, the capture of light intensity can be enhanced in two senses: one is the adaptation to global illumination conditions by shifting the characteristics of the sensors, and the other is the histogram equalization by nonlinear light-to-voltage transformation. Both effects can be obtained by wisely controlling the integration time. Capturing light in CMOS technology counts on the separation of photogenerated electron-hole pairs by effect of an electric field, normally provided by a reverse-biased pn junction. The generated photocurrent is proportional to the incident average light power density over the sensor area. Driven through a linear resistive load, $I_{ph}$ generates an instantaneous voltage, proportional to the incident light power. Noise can be filtered out by integrating this current in the diode’s parasitic capacitor $C_p$. Let us consider the photodiode in Fig. 1(a). This sensor is first reset to $V_{max}$ by means of transistor $M_{rst}$ and signal $\phi_{rst}$. After that, $I_{ph}$ is allowed to discharge the sensing capacitance, thus lowering $V_{ph}$ during time $t$, rendering:

$$V_{ph}(t) = V_{max} - \frac{I_{ph}}{C_p}t$$  \hspace{1cm} (1)
For a fixed $t_{\text{int}}$, the relation between $V_{ph}$ and the incident light power, through $I_{ph}$, is linear. Fig. 1(b) represents the voltage difference $V_{\text{max}} - V_{ph}(t_{\text{int}})$ vs. $I_{ph}$ for different values of the integration time. For a large $t_{\text{int}}$ the pixel output saturates even for rather low photocurrents. With a short integration time, only the largest values of $I_{ph}$ can make the pixel output reach saturation. If $I_{ph}$ is represented in a logarithmic scale, it can be seen that the same photosensor characteristic is shifted to operate at different illumination ranges by changing the integration time. As depicted by Eq. (1), the control of $t_{\text{int}}$ is a mean for photosensors gain control.

![Figure 1: (a) Integrating photosensor. (b) $\Delta V$ vs. $I_{ph}$ for different integration times.](image)

2 Adaptive opto-electronic interface

While CMOS photosensors have a maximum dynamic range of 5 to 7 decades, light intensity on different natural scenes can vary over up to 14 decades (commonly 5-7 decades of intra-frame variations, and the rest corresponding to inter-frame variations on illumination conditions). Thus, linear sensors usually produce images with over-exposed and under-exposed regions. To accommodate a larger light intensity range within the photodiode dynamic range, some voltage compression is required. In addition, the perceptual quality of an image is closely related with the ability of separating the irradiance, and the reflectance — where most of the relevant information is —, both contained in the luminance signal. Thus, modifying the perception gain according to the local average luminance, that is an estimation of the irradiance, results in a contrast enhanced picture [6].

A circuit for local adaptation of the photosensors operation is proposed, based on the computational features of the CNN Universal Machine architecture. The convenience for this real-time local adaptation has been pointed out in [7]. Several rules are proposed for local adaptation of the integration time, that map the already stored voltages — some of them can be the result of rather involved CNN algorithms — to the local $t_{\text{int}}$. In this paper we are establishing the connection between the local controlling voltage and the local integration time. Fig. 2(a) depicts the schematic of the local adaptation circuit. Photogenerated currents are integrated in $C_p$. The switch $M_{\text{rst}}$ works as the electronic shutter. While the local stored control voltage, $V_{\text{ctrl}}$, remains below a global reference, $V_{\text{ref}}$, the photosensor voltage is shorted to $V_{\text{max}}$. Once $\phi_{\text{rst}}$ falls, $I_{ph}$ start discharging the sense capacitor. This stops when $\phi_{\text{rst}}$ rises again.
The reference signal, which is monotonic, starts at the largest value for $V_{ph}$, which is $V_{\text{max}}$, and ends in a programmed $V_{\text{min}}$. Let us consider, as an example, that $V_{\text{ref}}$ has the form depicted in Fig. 2(b):

$$V_{\text{ref}}(t) = \begin{cases} -(\frac{V_{\text{max}}-V_{\text{min}}}{t_0})t + V_{\text{max}} & \text{if } t \leq t_0 \\ V_{\text{min}} & \text{if } t > t_0 \end{cases}$$

(2)

defined over a period of $t \in [0,t_{\text{per}}]$. This particular selection limits the dynamic range improvement to only 6dB, but different shapes can be employed for $V_{\text{ref}}$ in order to implement some brightness compression for a wider DR extension, as long as this transformation is monotonic. When the reference signal crosses the value of the locally stored control voltage, $V_{\text{ctrl}}$, the photodiode starts discharging $C_p$, stopping at $t_{\text{per}}$. The larger $V_{\text{ctrl}}$ the sooner the inverse ramp crosses it, and the longer the integration time will be. In particular:

$$t_{\text{int}} = t_{\text{per}} - \left(\frac{V_{\text{max}}-V_{\text{ctrl}}}{V_{\text{max}}-V_{\text{min}}}\right) t_0$$

(3)

In order to accommodate a larger illumination range and, at the same time, move the pixel gray levels towards the center of the photosensor range, the local adaptation mechanism should assign larger values of $V_{\text{ctrl}}$ for the darker image areas, and smaller $V_{\text{ctrl}}$ for the brighter pixels. In this way, the image areas with a weaker average illumination will have a longer exposure, while strongly illuminated areas will have a shorter $t_{\text{int}}$.

### 3 Local pixels’ average as $V_{\text{ctrl}}$

A first approach to the calculation of the local control voltage can be using the local average pixels voltage, $\bar{V}_{ph}$, as an estimation of the local average illumination. Highly illuminated areas render low voltages, while darker areas output larger pixel voltages. If the local average is the middle of the range, $\frac{1}{2}(V_{\text{min}}+V_{\text{max}})$, the corresponding integration time should be $t_{\text{per}} - \frac{t_0}{2}$. Using Eq. (3), a pixel with a $\bar{V}_{ph}$ above the midrange will have an integration time between $t_{\text{per}}$ and $t_{\text{per}} - \frac{t_0}{2}$. Those with averages below it will have integration times between $t_{\text{per}} - \frac{t_0}{2}$ and $t_{\text{per}} - t_0$. Based on the intra- and inter-frame correlation found in natural scenes, the local average voltage of a previous frame, $\bar{V}_{ph}(n-1)$, can be employed to control the integration time of the next frame, so as to establish a shorter integration time for strongly illuminated areas and a larger one for
the areas lying in the dark. This results in the following additive correction of the local integration time:

\[ t_{\text{int}}(n) = \left( t_{\text{per}} - \frac{t_0}{2} \right) + t_0 \left[ \frac{1}{2} - \frac{V_{\text{max}} - \tilde{V}_{\text{ph}}(n - 1)}{V_{\text{max}} - V_{\text{min}}} \right] \]  

(4)

The required local information can provided by a CNN core circuit in tens of nanoseconds. A high dynamic range image has been built with 11 snapshots of the same scene taken at different integration times. Fig. 3(a) shows a 256-level representation of this HDR image. Information is truncated by the limited DR. If it is re-captured with a simulated pixel array with local adaptation to the average illumination, some information is recovered (Fig. 3(b)). This result can be compared with logarithmic compression (Fig. 3(c)). The average brightness employed here is obtained by linear diffusion, but it can be the result of anisotropic diffusion, diffusion with controlled contours, etc [8].

![Figure 3: Simulated capture of a still picture with local adaptation (center).](image)

Stability concerns arise if \( t_0 \), that is the difference between the largest and the shortest integration time, is relatively large compared to \( t_{\text{per}} \), leading to non-convergence of the series defined by Eq. (4). This can be avoid by breaking the adaptation loop and performing a double capture at the expense of a reduction in the frame rate. Or, by making \( t_{\text{per}} - \frac{t_0}{2} \) dependent on \( V_{\text{ph}} \). Then, this series converges for a larger range of intra-frame integration times. For instance, forcing the average integration time of the previous frame to be the center of the interval of the integration times for the new frame:

\[ t_{\text{per}} - \frac{t_0}{2} = \bar{t}_{\text{int}}(n - 1) \]  

(5)

makes Eq. (4) converge for a larger set of \( t_0 \) values. When the image remains still \( \bar{t}_{\text{int}}(n) \) converges to an optimum for the particular illumination conditions. Fig. 4 displays how tracking a fixed gray level evolves to a stable integration time, and, in all cases, to the same medium gray. In graph Fig. 4(a), the sensor is presented to different plain gray pictures. For every capture, the starting average integration time is 30ms. The final average \( t_{\text{int}} \) depends on the input gray level. In picture Fig. 4(b), it can be seen how starting at different gray levels, the final average gray level (in a 256-level scale) is always 128. The sensor array adapts its average integration time to render a final output at a medium gray level. Finally, Fig. 5 shows several frames of an artificially distorted sequence. After a few frames, \( \bar{t}_{\text{int}}(n) \) converges and a locally adapted capture takes place.
the evolution of $t_{\text{int}}(n)$, it dynamically adapts to the average brightness, increasing when dark elements enter the scene and decreasing if the image gains in brightness. The local adaptation algorithm helps also to globally adapt the average exposure according to global illumination conditions. In this case, each step towards the optimum average $t_{\text{int}}$ is bounded, thus avoiding the flickering observed in algorithms based in direct calculation of the necessary integration time to render an average pixel voltage equal to $\frac{1}{2}(V_{\min} + V_{\max})$.

4 Conclusions

A simple but versatile adaptive opto-electronic interface for the CNN Universal Machine is depicted. This interface employs local control of the photosensors integration time to enlarge the captured illumination ranges and enhance the perceptual quality of the image, in terms of contrast detail. A locally stored voltage that is the result of parallel and concurrent computation is used to control the photosensors gain via the integration time adaptation.

Acknowledgements

This work has been partially supported by projects IST2001-38097 (LOCUST), TIC2003-09817-C02-01 (VISTA), and ONR-NICOP N000140210884.

References


Figure 5: Simulated adaptation in a sequence taken at 25fps, 1 out of each 3 frames shown.