Multi-resolution low-power Gaussian filtering by reconfigurable focal-plane binning

J. Fernández-Berni [a], R. Carmona-Galán[a], F. Pozas-Flores [a], Á. Zarándy[b] and
Á. Rodríguez-Vázquez[b]

[a]Institute of Microelectronics of Seville (IMSE-CNM)
CSIC-Universidad de Sevilla, Spain.
[b]Computer and Automation Research Institute (MTA-SZTAKI)
Hungarian Academy of Sciencies, Budapest, Hungary.

ABSTRACT

Gaussian filtering is a basic tool for image processing. Noise reduction, scale-space generation or edge detection are examples of tasks where different Gaussian filters can be successfully utilized. However, their implementation in a conventional digital processor by applying a convolution kernel throughout the image is quite inefficient. Not only the value of every single pixel is taken into consideration sucessively, but also contributions from their neighbors need to be taken into account. Processing of the frame is serialized and memory access is intensive and recurrent. The result is a low operation speed or, alternatively, a high power consumption. This inefficiency is specially remarkable for filters with large variance, as the kernel size increases significantly. In this paper, a different approach to achieve Gaussian filtering is proposed. It is oriented to applications with very low power budgets. The key point is a reconfigurable focal-plane binning. Pixels are grouped according to the targeted resolution by means of a division grid. Then, two consecutive shifts of this grid in opposite directions carry out the spread of information to the neighborhood of each pixel in parallel. The outcome is equivalent to the application of a 3×3 binomial filter kernel, which in turns is a good approximation of a Gaussian filter, on the original image. The variance of the closest Gaussian filter is around 0.5. By repeating the operation, Gaussian filters with larger variances can be achieved. A rough estimation of the necessary energy for each repetition until reaching the desired filter is below 20nJ for a QCIF-size array. Finally, experimental results of a QCIF proof-of-concept focal-plane array manufactured in 0.35μm CMOS technology are presented. A maximum RMSE of only 1.2% is obtained by the on-chip Gaussian filtering with respect to the corresponding equivalent ideal filter implemented off-chip.

Keywords: Focal-plane processing, Gaussian kernels, binomial filter mask, low-power smart image sensors

1. INTRODUCTION

Gaussian kernels are a fundamental component of a computational approach to visual perception motivated by physics and biological vision.[1] Convolution with Gaussian kernels and Gaussian derivatives constitute a canonical class of image operators for early vision. As a family, Gaussian kernels form a semi-group. One important property is that any coarser scale representation can be obtained from any representation at a finer level. Additionaly, Gaussian kernels have the property of preserving local extrema in the image, i. e. no minima nor maxima are accidentally introduced when a Gaussian blur is applied in order to supress finer scale details of the image.[2] Because of these properties, Gaussian filters are able to generate a scale space[3] and, consequently, a multi-scale image representation.[4] It is worth mentioning that scale-space operators have a similar form to the receptive fields observed in neurophysiological studies.[5] This type of image representation is certainly useful for image interpretation. As there is no a priori knowledge about the scale of the relevant elements in the scene, a multi-scale representation covers all the possible ranges. Image features can then be extracted at different scales and scale-invariant features can be highlighted as characteristic of whatever takes place in the visual field.[6] It is not strange that visual attention models based on saliency make extensive use of these operators.[7]

Further author information:
Jorge Fernández-Berni: E-mail: berni@imse-cnm.csic.es, Telephone: +34 954466666

The isotropic Gaussian kernel, centered at the origin, employed to generate a scale-space representation of a two-dimensional image, is defined as a parametrized function $g : \mathbb{R}^2 \times \mathbb{R}_+ \to \mathbb{R}$ where:

$$G\left(\mathbf{x}; \xi\right) = \frac{1}{2\pi\xi}e^{-|\mathbf{x}|^2/2\xi} \qquad \Leftrightarrow \qquad \hat{G}\left(\mathbf{k}; \xi\right) = e^{-2\pi^2|\mathbf{k}|^2\xi} \tag{1}$$

in which ξ is referred as the scale parameter and corresponds to the variance of the Gaussian kernel ($\xi = \sigma^2$), and $\hat{G}(\cdot)$ is the Fourier transform of $G(\cdot)$. One advantage from the point of view of the implementation is that the Gaussian kernel is separable into two orthogonal functions $G_1(\cdot)$ and $G_2(\cdot)$:

$$G\left(\mathbf{x}; \xi\right) = G_1(x_1; \xi) * G_2(x_2; \xi) = \frac{1}{2\pi\xi}\left(e^{-x_1^2/2\xi} * e^{-x_2^2/2\xi}\right) \tag{2}$$

Given that the image plane is discretized, the function $G(\cdot)$ is only evaluated at valid points of the grid. For a relatively large σ, i. e. higher scales, the number of elements of the kernel that cannot be neglected is prohibitively large, as can be seen below:

0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.04	0.00	0.00
0.00	0.04	1.00	0.04	0.00
0.00	0.00	0.04	0.00	0.00
0.00	0.00	0.00	0.00	0.00

0.00	0.00	0.00	0.00	0.00
0.00	0.03	0.11	0.03	0.00
0.00	0.11	0.44	0.11	0.00
0.00	0.03	0.11	0.03	0.00
0.00	0.00	0.00	0.00	0.00

0.00	0.01	0.02	0.01	0.00
0.01	0.06	0.10	0.06	0.01
0.02	0.10	0.16	0.10	0.02
0.01	0.06	0.10	0.06	0.01
0.00	0.01	0.02	0.01	0.00

$$\sigma = 0.4 \qquad\qquad \sigma = 0.6 \qquad\qquad \sigma = 1.0$$

In fact, a minimum size of 6σ has been estimated in order to avoid excessive ripple in the stop band introduced by truncation.[8] In terms of the required computing power and resources, the dynamic adaptation of the kernel size represents a significant drawback. An alternative approach will be time-multiplexing the smoothing operators. In other words, repeatedly applying smaller kernels in order to obtain a higher scale parameter, what directly derives from the semi-group characteristic of the Gaussian kernels:

$$G\left(\mathbf{x}; \xi_1 + \xi_2\right) = G\left(\mathbf{x}; \xi_1\right) * G\left(\mathbf{x}; \xi_2\right) \tag{3}$$

that can easily be understood in the Fourier domain:

$$\hat{G}\left(\mathbf{k}; \xi_1 + \xi_2\right) = e^{-2\pi^2|\mathbf{k}|^2(\xi_1+\xi_2)} = e^{-2\pi^2|\mathbf{k}|^2\xi_1} \cdot e^{-2\pi^2|\mathbf{k}|^2\xi_2} = \hat{G}\left(\mathbf{k}; \xi_1\right)\hat{G}\left(\mathbf{k}; \xi_2\right) \tag{4}$$

Therefore, we need to select an elementary Gaussian filter, or an approximation, that can be easily implemented, both in terms of the number of non-zero elements of the kernel and in terms of the relations between them. The 2-D binomial filter[4] is a good candidate:

$$\mathbf{B}^2 = B^2 * \left(B^2\right)^T = \frac{1}{4}\begin{bmatrix} 1 & 2 & 1 \end{bmatrix} * \frac{1}{4}\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} = \frac{1}{16}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} \tag{5}$$

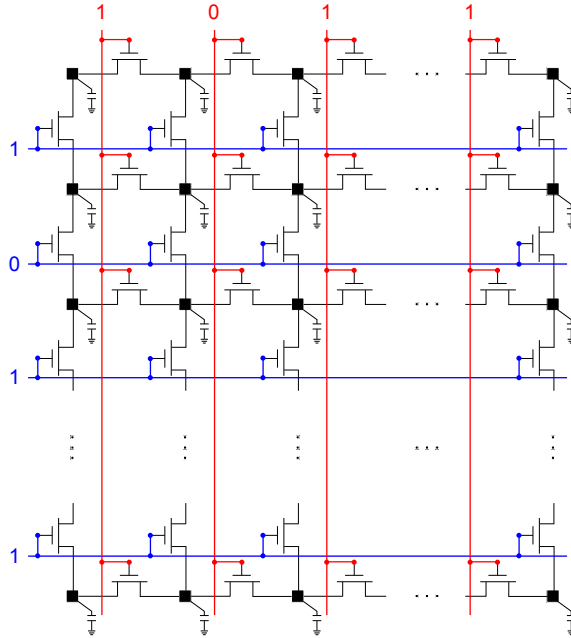Figure 1: Focal-plane capacitor grid for charge redistribution

which is the result of convolving a horizontal, B^2, and a vertical, $\left(B^2\right)^T$, 1-D binomial masks. Each of these 1-D filters are, in turn, the result of convolving twice the elementary averaging mask, B^1:

$$B^2 = B^1 * B^1 = \frac{1}{2} \left[\begin{array}{cc} 1 & 1 \end{array} \right] * \frac{1}{2} \left[\begin{array}{cc} 1 & 1 \end{array} \right] = \frac{1}{4} \left[\begin{array}{ccc} 1 & 2 & 1 \end{array} \right] \qquad (6)$$

Because of the central limit theorem, the transfer function and the mask of the binomial filter approximate the Gaussian filter with an equivalent variance. In the case of the kernel expressed in Eq. (5) the variance is 0.5, and the error committed in the approximation of the equivalent Gaussian filter is around 0.8%, depending on the input image.

The rest of the paper is dedicated to an efficient implementation of the binomial filter based on the use of focal-plane multi-resolution capabilities. It is organized as follows. First we will show how reconfigurable resolution is implemented by adding the possibility of binning pixels together and allowing for charge redistribution among them. Then we will demonstrate that the effect of repeatedly averaging the pixels in shifted divisions of the focal-plane grid is that of applying a binomial filter. Finally, some experimental results, obtained with a prototype chip fabricated in a 0.35μm CMOS technology, are displayed, confirming the validity of the approach.

2. CHARGE REDISTRIBUTION AND PIXEL BINNING

At the focal plane of a CMOS imager, the photogenerated current is directly sensed and (or) integrated.[9] In the latter case, the pixel value is a voltage at the sensing capacitor. This voltage is stored, at least temporarily, so it can be read out. If an electronic shutter is provided,[10] the pixel voltage is maintained until the next reset, within the accuracy permitted by leakages. Fully-parallel operations can be performed onto these voltages at the focal plane without using an external memory as these capacitors act as a distributed analog memory. If switches are provided between the capacitors, as can be seen in Fig. 1, the stored charge redistributes ending in the averaging of the initial voltages. Let us consider that, by setting the appropriate control pattern, a sub-image of size $m \times n$ is isolated. This is realized by turning on the $m - 1$ signals that control the connections between the m rows of pixels, and the $n - 1$ signals that control the connections between the n columns in Fig. 1. By enabling the

electrical paths between the $m \times n$ capacitors, the pixels whose original values are $p_{ij}^0, \ldots, p_{i+m-1,j+n-1}^0$ end in:

$$p_{i+k,j+l}\Big|_{\forall k \in \{0,\ldots,m-1\}, \forall l \in \{0,\ldots,n-1\}} = \frac{1}{mn} \sum_{k=0}^{m-1} \sum_{l=0}^{n-1} p_{i+k,j+l}^0 \qquad (7)$$

It is worth to mention that the result is exactly the same if the switches conforming the $m \times n$ region are set from the start, as charge redistributes in parallel with photocurrent integration. This is called pixel binning.[11]

Consider now a regular subdivision of the focal-plane grid. For instance, an alternate sequence of 1's and 0's is loaded into the row and column connection control registers of Fig. 1. It means that the full-resolution image of $M \times N$ pixels is divided into 2×2-pixel blocks. As the four pixels within each block are connected together, they will end up having the same pixel value:

$$p_{i,j}\Big|_{i \in \{1,3,5,\ldots,M-1\}, j \in \{1,3,5,\ldots,N-1\}} = \frac{1}{4}\left(p_{ij}^0 + p_{i,j+1}^0 + p_{i+1,j}^0 + p_{i+1,j+1}^0\right) \qquad (8)$$

that is the average of the original values of the four pixels contained in the block. We have assumed that M and N are even. The resulting image contains $M/2 \times N/2$ pixels, with the connection scheme depicted in Fig. 2(a). It will be the starting point for the processing we will explain later. Another relevant assumption is that any feature that we are interested in must be noticeable at this resolution. The following analysis applies to images divided in blocks of any size as long as their dimensions are even and the results to be expected are $M/2 \times N/2$-pixel or smaller images.

3. GAUSSIAN FILTERING BY GRID SHIFTING

Let us start with an image, of size $M \times N$-pixels, stored in a capacitor grid like that of Fig. 1. The grid has been divided into 2×2-pixel blocks, within which charge has been allowed to redistribute. It means that our initial image is of size $M/2 \times N/2$-pixels and has four capacitors storing the same voltage, i. e. the same pixel value (Fig. 2(a)). Let us concentrate on the transformation that is going to be suffered by the value p_{ij} stored at the position indicated by the arrow in Fig. 2(a). At a certain point in time, the alternate sequences of 1's and 0's at the row and column connection control registers are shifted one space down and to the right, respectively. The
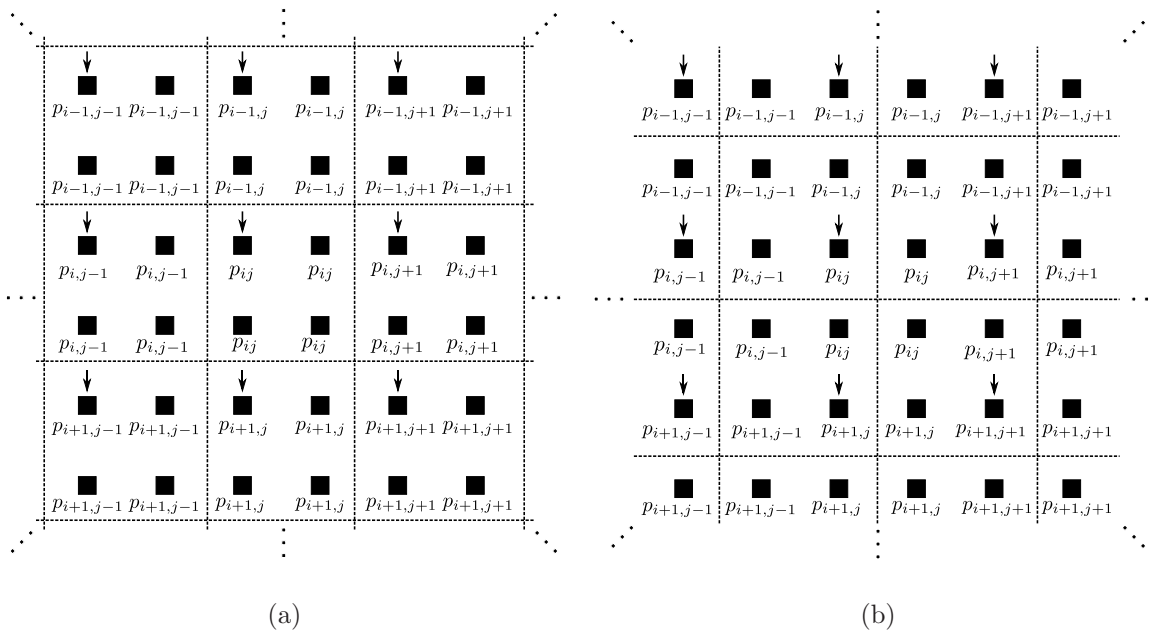

(a)

(b)

Figure 2: (a) Focal-plane division in 2×2-pixel blocks and (b) shifted grid.

pixel grouping scheme changes from that of Fig. 2(a) to the one depicted in Fig. 2(b). Consequently, because of a new redistribution of the charge in the newly formed blocks, the value of the marked node becomes:

$$p'_{ij} = \frac{1}{4} \left(p_{i-1,j-1} + p_{i-1,j} + p_{i,j-1} + p_{ij} \right) \tag{9}$$

The values at the neighboring nodes, that were originally p_{ij} as well, are now averaged in their new 2×2-pixel blocks, so they have been transformed into:

$$p'_{i,j+1} = \frac{1}{4} \left(p_{i-1,j} + p_{i-1,j+1} + p_{ij} + p_{i,j+1} \right) \tag{10}$$

$$p'_{i+1,j} = \frac{1}{4} \left(p_{i,j-1} + p_{ij} + p_{i+1,j-1} + p_{i+1,j} \right) \tag{11}$$

$$p'_{i+1,j+1} = \frac{1}{4} \left(p_{ij} + p_{i,j+1} + p_{i+1,j} + p_{i+1,j+1} \right) \tag{12}$$

If the control sequences are shifted back to the original position, one space up and to the left, then the new values expressed by Eqs. (9)-(12) and averaged once more, resulting in:

$$p''_{ij} = \frac{1}{16} \left(p_{i-1,j-1} + 2p_{i-1,j} + p_{i-1,j+1} + 2p_{i,j-1} + 4p_{ij} + 2p_{i,j+1} + p_{i+1,j-1} + 2p_{i+1,j} + p_{i+1,j+1} \right) \tag{13}$$

Notice that the $M \times N$-pixel image has undergone two shifts of the connection scheme followed by the averaging of the pixel values within the resulting 2×2-pixel blocks. Each combination of grid shifting and averaging has the same effect as applying the averaging mask:

$$\mathbf{B}^1 = B^1 * \left(B^1 \right)^T = \frac{1}{2} \left[\begin{array}{cc} 1 & 1 \end{array} \right] * \frac{1}{2} \left[\begin{array}{c} 1 \\ 1 \end{array} \right] = \frac{1}{4} \left[\begin{array}{cc} 1 & 1 \\ 1 & 1 \end{array} \right] \tag{14}$$

over a $M/2 \times N/2$-pixel image. By doing it twice, we are applying the 3×3 binomial filter mask of Eq. (5):

$$\mathbf{B}^1 * \mathbf{B}^1 = \frac{1}{4} \left[\begin{array}{cc} 1 & 1 \\ 1 & 1 \end{array} \right] * \frac{1}{4} \left[\begin{array}{cc} 1 & 1 \\ 1 & 1 \end{array} \right] = \frac{1}{16} \left[\begin{array}{ccc} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{array} \right] = \mathbf{B}^2 \tag{15}$$

that is precisely what is expressed in Eq. (13). This theoretical result has been checked by numerical simulation[*], yielding 0.16% RMSE for a 256×256-pixel image of Lena. This small error is associated to differences in the rounding error committed on following the different methods.

4. CHIP MEASUREMENTS

Although the above described procedure may theoretically render the same results as the direct convolution with the binomial filter mask, its physical implementation involves a number of switches to reconfigure and shift the connection grid. Switching error becomes more apparent when the storage capacitors are small. In this section we are showing the results obtained by implementing binomial filtering by shifted average grids in a prototype chip with focal-plane reconfigurability and multi-resolutional capabilities. The prototype chip (Fig. 3)[12] has been fabricated in a 0.35μm CMOS process with anti-reflective coating and reduced photodiode dark response. A summary of the chip characteristics and features is given at Table 1. This chip was not originally thought to operate following the already explained scheme, but it has a reconfigurable focal-plane connection grid, like that in Fig. 1, that provides multi-resolution capabilities.

The filtering procedure explained above has been programmed into the chip test environment. The results obtained on-chip render a 1.12% RMSE for the first application of the filter. This overall error is attributable to the accumulated switching errors and also to the noisy readout. Fig. 4 depicts the original 176×144-pixel

---

[*]Matlab$^{\circledR}$ files for comparing the results of realizing binomial filtering either directly or by grid shifting and averaging can be found at http://www.imse-cnm.csic.es/wivisnet/spie_files/
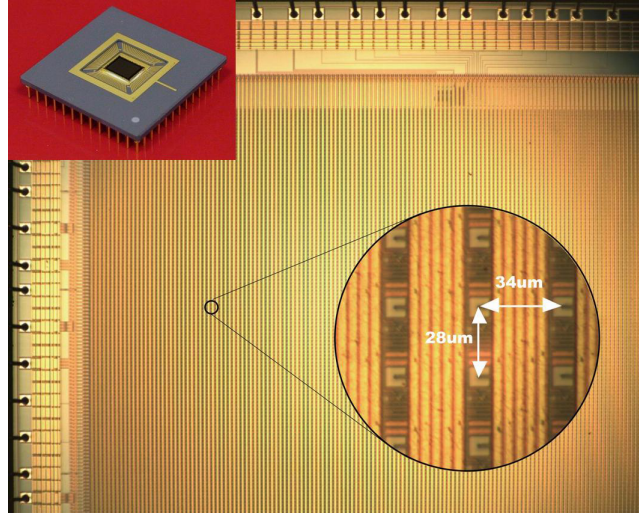
Figure 3: General view and microphotographs of the CMOS prototype chip

Technology	0.35μm CMOS 2P4M 3.3V
Vendor (Process)	Austria Microsystems (C35OPTO)
Die size (with pads)	7280.8μm $\times$ 5780.8μm
Cell size	34.07μm $\times$ 29.13μm
Fill factor	6.45%
Resolution	QCIF: 176$\times$144 px
Photodiode type	n-well/p-substrate
FPN	0.72%
PRNU (50% signal range)	2.42%
Sensitivity	0.15V/(lux·s)
Measured power consumption	5.6mW@12kSa/s
Maximum throughput	110kSa/s (9μs/Sa)

Table 1: Summary of the prototype chip features.

image captured by the chip, together with the downsampled, after pixel binning, 88×77-pixel version, that is the initial image for both the on-chip and the off-chip (ideal) filtering. Starting from this image, sucessive steps has been realized in order to generate a space scale. Each step implies the convolution with the binomial filter mask ($\mathbf{B}^2$), either by averaging and shifting the connection grid on-chip or directly applying the mask off-chip with Matlab®. This can be seen in Fig. 5. The first column represents the image filtered on-chip. The second the off-chip, ideal, version starting from the same input (Fig. 4(b)). The third column is the difference normalized to the value of the maximum individual pixel error detected at each step. This maximum deviation is 3.17%, 3.83%, 3.69%, 3.82%, 5.07%, 4.74%, 4.79%, 4.96% and 5.66%, respectively. For the complete image, the measured RMSE is 1.12%, 1.39%, 1.55%, 1.69%, 1.82%, 1.92%, 2.02%, 2.12% and 2.23%, respectively for each step. Notice that the ideal filtering has the effect of averaging the zero-mean noise introduced by readout at every step of the on-chip filtering. This noise is re-sampled each time a new image is delivered from the on-chip processing. The consequence is that the error tends to increase as we go up the scale space.

An important feature of this alternative method to compute the scale space is that the incidence on the power budget is far below the milliwatt. For each repetition, shifting the grid and averaging twice is estimated to require 20nJ. This estimation is obtained by simulation and represents switching the complete connection grid twice. Image capture and readout are excluded from this sum. At 30fps, it represents 0.6μW, what is certainly negligible and below the precision of our measurement setup.

5. CONCLUSIONS

Theoretical background for the implementation of an approximated Gaussian filter by using multi-resolution capabilities at the focal-plane is given. Ideally, the only difference with the direct application of the binomial filter convolution mask is rendered by the rounding error of the computing hardware. We have implemented this procedure in a prototype chip with all the necessary means to reconfigure the focal-plane connection scheme. The results evidence the validity of our assumption. The on-chip filtering approximates the ideal within a 1.2% error. The incidence of this processing in the total power budget of the smart imager operation is negligible.

ACKNOWLEDGMENTS

REFERENCES

[1] Romeny, B. t. H., [*Front-End Vision and Multi-Scale Image Analysis*], Springer (2003).

[2] Lindeberg, T. and Romeny, B. t. H., "Linear scale-space: (i) basic theory (ii) early visual operations," in [*Geometry-Driven Diffusion in Computer Vision*], ter Haar Romeny, B. t. H., ed., 1–77, Kluwer Academic Publishers (1994).

[3] Lindeberg, T., "Scale-space," in [*Encyclopedia of Computer Science and Engineering*], Wah, B., ed., **IV**, 2495–2504, John Wiley and Sons (2008).

[4] Jahne, B., "Multiresolutional signal representation," in [*Handbook of Computer Vision and Applications*], Jahne, B., Hauβecker, H., and Geiβler, P., eds., **2**, 67–90, Academic Press (1999).

[5] Soodak, R. E., "Two-dimensional modeling of visual receptive fields using Gaussian subunits," *Proceedings of the National Academy of Sciences* **20**, 9259–9263 (December 1986).

[6] Lowe, D. G., "Object recognition from local scale-invariant features," in [*Proc. of the IEEE Int. Conference on Computer Vision*], **2**, 1150–1157 (1999).

[7] Itti, L., Koch, C., and Niebur, E., "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20**, 1254–1259 (November 1998).

[8] Sotak, G. E. and Boyer, K. L., "The Laplacian-of-Gaussian kernel: a formal analysis and design procedure for fast, accurate convolution and full-frame output," *Comput. Vision Graph. Image Process.* **48**, 147–189 (November 1989).

Figure 4: Chip captured image (a) and downsampled version (b).

[9] Otha, J., [*Smart CMOS Image Sensors and Applications*], CRC Press (2007).

[10] Aw, C. H. and Wooley, B., "A 128x128-pixel standard-CMOS image sensor with electronic shutter," *Solid-State Circuits, IEEE Journal of* **31**, 1922 –1930 (December 1996).

[11] Zhou, Z., Pain, B., and Fossum, E., "Frame-transfer CMOS active pixel sensor with pixel binning," *Electron Devices, IEEE Transactions on* **44**, 1764 –1768 (October 1997).

[12] Fernández-Berni, J., Carmona-Galán, R., and Carranza-González, L., "FLIP-Q: A QCIF resolution focal-plane array for low-power image processing," *IEEE J. of Solid-State Circuits* **46**, 669–680 (March 2011).
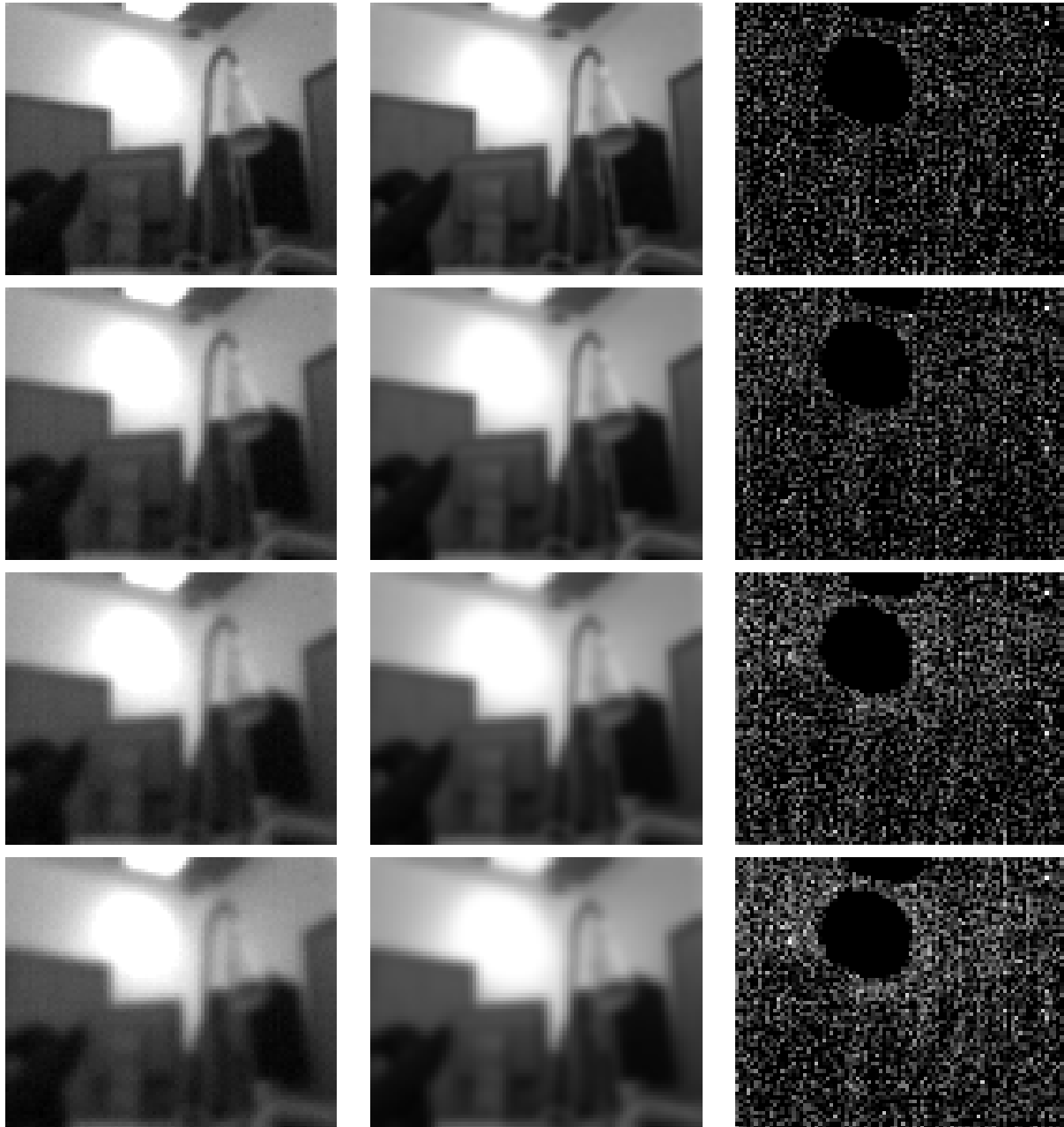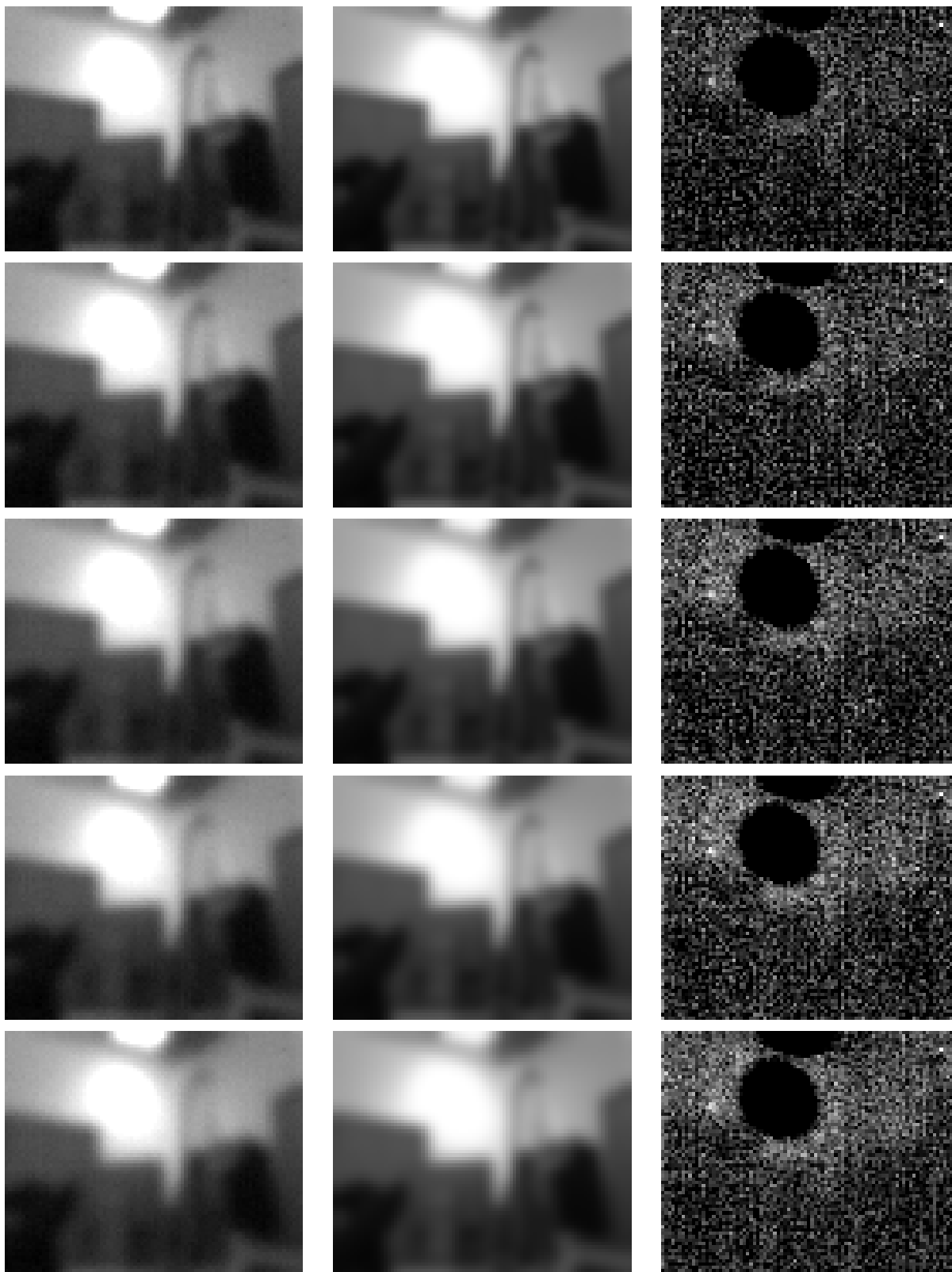
Figure 5: On-chip filtering, ideal and amplified difference.

Figure 5: (Cont.) On-chip filtering, ideal and amplified difference(c)