THE GIGAVISION CAMERA

Luciano Sbaiz [1], *Feng Yang* [1], *Edoardo Charbon* [2], *Sabine Süsstrunk* [1] *and Martin Vetterli* [1,3]

[1] LCAV - School of Computer and Communication Sciences
Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland
[2] AQUA - School of Computer and Communication Sciences
Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland
[1,3] Department of Electrical Engineering and Computer Science
UC Berkeley, Berkeley CA94720, USA

ABSTRACT

We propose a new image device called *gigavision camera*. The main differences between a conventional and a gigavision camera are that the pixels of the gigavision camera are binary and orders of magnitude smaller. A gigavision camera can be built using standard memory chip technology, where each memory bit is designed to be light sensitive. A conventional gray level image can be obtained from the binary gigavision image by low-pass filtering and sampling.

The main advantage of the gigavision camera is that its response is non-linear and similar to a logarithmic function, which makes it suitable for acquiring high dynamic range scenes. The larger the number of binary pixels considered, the higher the dynamic range of the gigavision camera will be. In addition, the binary sensor of the gigavision camera can be combined with a lens array in order to realize an extremely thin camera. Due to the small size of the pixels, this design does not require deconvolution techniques typical of similar systems based on conventional sensors.

Index Terms— high dynamic range imaging, computational photography, thin camera, logarithmic sensor response

1. INTRODUCTION

In a conventional camera, an array of light-sensitive pixels located in the focal plane of a lens evaluates the irradiance over a given detection area. The result is an analog signal that is subsequently amplified and A/D converted. The digital equivalent of the different light intensities are different gray-levels. The main drawback of this architecture is that the dynamic range, i.e. the ratio between the maximum and minimum detectable irradiance values, is limited by the sensor technology and smaller than that of the human eye [1]. To overcome this problem, researchers proposed to combine pictures taken with different exposure times/apertures [2, 3, 4] or to use sensors with logarithmic response [5]. In this paper, we propose a novel image sensor called gigavision that is based on pixels with binary output. A more general architecture using few-ary pixels that can produce 4, 8, or 16 gray levels is also possible, but not addressed in this paper. This architecture can be implemented using standard memory chip technology, where each bit is designed to be sensitive to visible light. The current level of integration exceeds 10^9 bits per chip; therefore, we can process the binary image to obtain a conventional gray level image with high spatial resolution. Such a conversion is realized by low-pass filtering and sampling, in a way similar to the well-known oversampling techniques applied in A/D converters [6]. We call N the oversampling factor, i.e. the ratio between the number of pixels of the binary image and that of the desired gray level image. Suppose the resolution of the binary image is M, then the resolution of the desired gray level image is $\frac{M}{N}$.

We show that, since each pixel value reconstructed by the gigapixel camera is obtained by combining at least N binary pixels, the response function of the camera is non-linear and similar to a logarithm without the disadvantages of logarithmic sensors. This makes it suitable to acquire scenes with high dynamic range. Moreover, the estimation noise decreases with the factor N; therefore, by increasing the oversampling factor, it is possible to reduce noise. The non uniform pixel response in conventional sensor is responsible for fixed pattern noise. In the case of the proposed architecture, such non uniformity is averaged by the low-pass filter and attenuated.

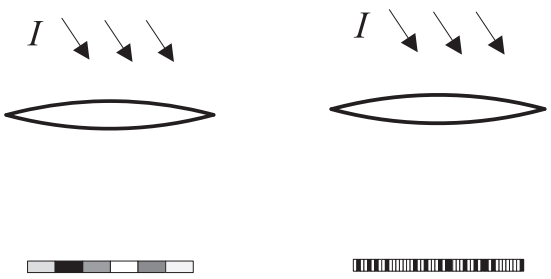
Recently, an architecture based on a lens array and a conventional image sensor has been proposed to realize an ultra thin camera [7]. Unfortunately, the large pixel pitch size of the sensor corresponds to a low-pass effect on the resulting image. This effect increases with the number of lenses composing the array, i.e. with the reduction in thickness of the camera. To compensate this image degradation, deconvolution algorithms are used, but they increase the noise. Instead, the gigavision sensor presents an extremely small pixel pitch size and deconvolution is not needed. A conventional image is reconstructed by combining the binary pixels at corresponding positions of the lens array, making the sensor very interesting for this application.

This paper is organized as follows. Section 2 shows the system architecture of a gigavision camera. Section 3 focuses on the theoretical analysis of the gigavision camera, like response function, the influence of threshold and estimation error variance. Simulation results are in Section 4, and the conclusion is in Section 5.

2. CAMERA ARCHITECTURE

In Fig. 1(a) the architecture of a simplified conventional camera is represented. A single lens concentrates the incident light on the image sensor, which converts it into an electrical signal. A quantizer discretizes the amplitude in values corresponding to different gray levels. If this element is neglected, the pixel value is proportional to the light intensity *I*. Since the maximum voltage of the electrical signal is limited, the sensor saturates above a certain light intensity. The saturation level and the technological limit on sensor noise makes it difficult to arbitrarily increase the dynamic range. With the current technology this is too small for many applications [1].

The work presented in this paper was supported by the Swiss National Science Foundation under grant number 200021-116742.



(a) a conventional camera (b) a gigavision Camera

Fig. 1. Simplified architecture of a conventional camera and a gigavision Camera. The incident light is focused by the lens and then impinges on the image sensors.

In a gigavision camera the analog sensor is replaced by a binary sensor, as shown in Fig. 1(b). An image with a number of pixels equal to a conventional camera is obtained by low-pass filtering and sampling the binary image. In this paper, for simplicity, the low-pass filter corresponds to the sum of a block composed by N pixels.

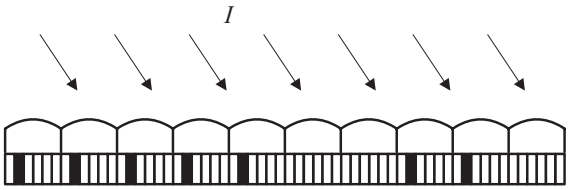


Fig. 2. Architecture of a thin camera based on a lens array and a gigavision sensor. Here parallax is neglected, and every element of the lens array generates the binary picture of the same scene. And the final picture is constructed by adding these binary pictures.

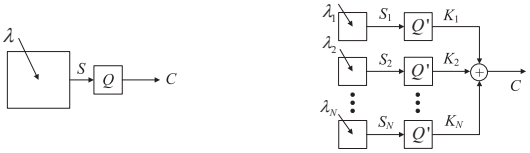
An alternative architecture is based on a lens array instead of a single lens, as shown in Fig. 2. We suppose that the lens array is composed of N lenses, i.e. as many as the oversampling factor. If parallax is neglected, the array projects N binary replicas of the original image on the sensor. Since each image is generated independently, as in the previous case, the binary images can be added to obtain a conventional image. That is, the averaging over the spatial domain is replaced by averaging over the elements of the array.

We remark that, with the current technology, for large values of N the pixel size can be much smaller than the diffraction-limited spot. However, diffraction is equivalent to a low-pass filter, which is combined to the reconstruction filter. Hence, diffraction limits only the maximum bandwidth of the reconstructed image and not the possibility to reduce the reconstruction noise by increasing N.

3. PERFORMANCE ANALYSIS

3.1. Pixel model

A simplified pixel model of a conventional sensor is shown in Fig. 3(a). The number of photons impinging on the pixel can be modeled as a Poisson process. If the area of the pixel is A, then the intensity of the process is IA and the average number of arrivals during the exposure time t is $\lambda = ItA$. The number of photons S impinging upon the pixels during the exposure time is a Poisson random variable with parameter λ. The photons are converted into an electrical signal, amplified, and A/D converted. If we neglect the noise introduced by the electronic circuits, the result is the process $C = Q(S)$.



(a) a conventional pixel (b) a gigavision pixel

Fig. 3. Simplified block diagram of a conventional pixel and a gigavision pixel. For the conventional camera, the incident photons first are converted to the electrical signal S, then quantized by a multilevel quantizer. For the gigavision camera, the electrical signal S_i is quantized by an one-bit quantizer with threshold T and the pixel value C is the sum of the N pixels.

The model of the gigavision sensor pixel is shown in Fig. 3(b). As explained above, a simple way to convert the binary image into a gray level image is to combine blocks of N consecutive pixels by addition. We remark that the same method can be applied to the architecture of Fig. 2, where, in this case, the added pixels are taken from the corresponding positions of the different elements of the lens array. Each pixel is obtained in a way similar to the conventional sensor, but the final quantizers Q' are binary with threshold T. In the following, we suppose that each pixel is exposed to the same intensity IA/N, i.e. the area of each pixel is N times smaller than that of the conventional pixel. The value of each pixel K_i, $i = 1, \ldots, N$ is obtained by comparing the number of arrivals S_i, i.e. the arrived photons, with the threshold T. In this case, the quantities K_i are binary random variables with parameter

$$p_\lambda = \mathbb{P}[S_i \geq T] = \sum_{k=T}^{\infty} e^{(-\lambda/N)} \frac{(\lambda/N)^k}{k!}.$$

The result after the addition of the N pixels is given by $C = \sum_{i=1}^{N} K_i = \sum_{i=1}^{N} Q'(S_i)$. The quantity C is a binomial random variable with parameters N and p_λ, since it is the sum of N binary independent random variables.

3.2. Response function

The response function is the relation between photon intensity and expectation of the pixel value. For a conventional sensor, when quantization is neglected, the response function is linear, i.e. $\mathbb{E}[C] \simeq \alpha\lambda$, where α is constant for all the pixels of the image. If quantization is not neglected, the measured values of C are rounded towards the closest quantized value and saturation occurs when the number of arrivals exceed a maximum value. Two examples, for $N = 256$, $\alpha = 1$, and $\alpha = 0.25$, are shown in Fig. 4. Note that the relation between $\mathbb{E}[C]$ and λ is not exactly a staircase function, due to the random distribution of photon arrivals for each value of intensity, which tends to round the steps of the staircase. For the gigavision sensor, $\mathbb{E}[C] = Np_\lambda$. The responses for $N = 256$, $N = 4096$, and thresholds $T = 1$, and $T = 4$, are shown in Fig. 4.

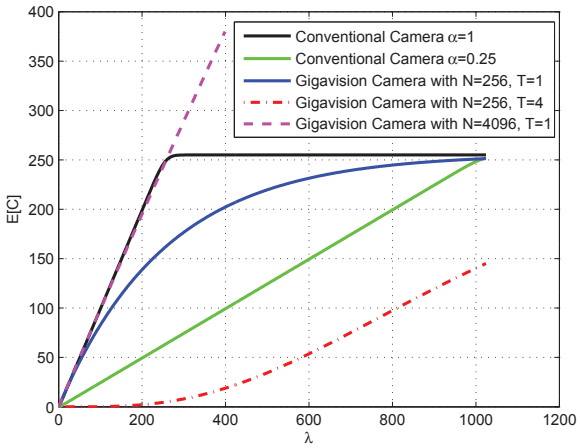


Fig. 4. Response function of a conventional pixel and a gigavision pixel. The quantizer of the conventional pixel has $N = 256$ levels and parameters $\alpha = 1$ and $\alpha = 0.25$. For the gigavision pixel the thresholds are $T = 1$, and $T = 4$, and the oversampling factors are $N = 256$ and $N = 4096$.

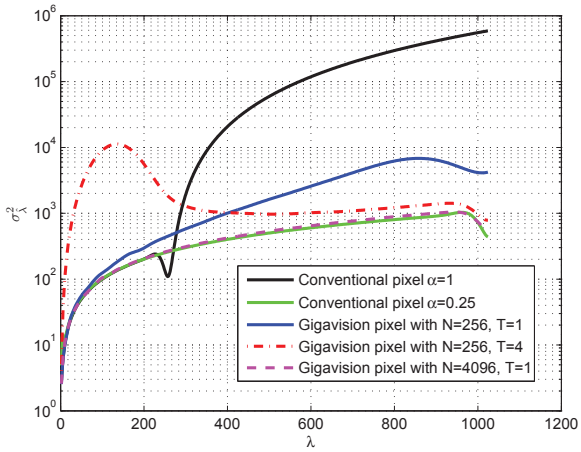


Fig. 5. Estimation error variance of the conventional pixel and the gigavision pixel. The quantizer of the conventional pixel has 256 levels and parameter $\alpha = 1$ and 0.25. For the gigavision pixel the thresholds are $T = 1$, and 4, and the oversampling factors are $N = 256$ and 4096.

We note that the gigavision pixel presents a non-linear response. When the threshold is set to $T = 1$, the response function of gigavision camera is

$$\mathbb{E}[C] = N\left(1 - e^{-\frac{\lambda}{N}}\right),$$

then,

$$\lambda = -Nlog\left(1 - \frac{\mathbb{E}[C]}{N}\right),$$

is a logarithmic function. Both the conventional pixel and the gigavision pixel have the same sensitivity at low light intensities. This happens because, for both pixels, each incoming photon increases by one the total photon count. However, when intensity is higher (above 256 for this example) the conventional pixel saturates. Instead, in the case of the gigavision pixel, an increasing number of photons hit pixels already in the "1" state and don't contribute to the final result. That corresponds to a reduction of sensitivity as intensity increases, and the pixel never saturates.

When $T = 4$, the sum C of active pixels remains close to zero for low light intensity. This happens because, in this case, it is unlikely that four photons arrive on the same pixel and switch it to the "1" state. After this "dead zone" the transfer function is still non-linear and similar to a logarithm, even if the sensitivity is reduced with respect to the previous case.

3.3. Estimation error variance

The sum c of active pixels is used to obtain the estimate $\hat{\lambda}$ of parameter λ, which corresponds to light intensity. This is obtained by using a maximum likelihood estimator,

$$\begin{aligned}\hat{\lambda} &= \arg\max_{\lambda} \mathbb{P}[C = c] \\ &= \arg\max_{\lambda} \binom{N}{c} p_\lambda^c (1 - p_\lambda)^{N-c}, \end{aligned} \quad (1)$$

i.e. $\hat{\lambda}$ is chosen so that, $p_{\hat{\lambda}} = c/N$.

It is interesting to compute the estimation variance, i.e. the quantity $\mathbb{E}[(\hat{\lambda} - \lambda)^2]$ and compare the conventional to the gigavision pixel variance.

Fig. 5 shows the estimation error variance of the conventional pixel and the gigavision pixel with threshold $T = 1$, $N = 256$, and $N = 4096$. For low light intensities, the conventional pixel with $\alpha = 1$ is better than the conventional pixel with $\alpha = 0.25$, since the steps of the quantizer are finer. As said above, the gigavision pixel with $T = 1$ behaves, in this regime, as the conventional pixel with $\alpha = 1$, i.e. it gives also a low estimation error variance. When intensity increases, the conventional pixel with $\alpha = 1$ saturates and the estimation error variance rapidly increases. Instead, the conventional pixel with $\alpha = 0.25$ still gives a low estimation error variance. Due to the non-linear behavior, the gigavision pixel, with $T = 1$ and $N = 256$ does not saturate and the estimation error variance is only slightly higher than the conventional pixel with $\alpha = 0.25$. We can expect that the difference in estimation error variance can be reduced by choosing a higher value of N. In fact, by taking $N = 4096$ the estimation error variance is very close to the conventional pixel with $\alpha = 0.25$, but having at the same time a lower estimation error variance at low light intensity, as the conventional pixel with $\alpha = 1$.

Fig. 6 shows the estimation error variance of the conventional pixel and the gigavision pixel when λ is in the initial region $[0, 50]$. It can be noted that at low light intensities, the performance of the conventional pixel is worse than the performance of a gigavision pixel.

In a practical case, one may have to consider a value T larger than 1. In Fig. 5, the case $T = 4$ and $N = 256$ is shown. The dead zone of the response function is responsible for the relatively high estimation error variance at low light intensity. However, at higher light intensity the pixel gives a low estimation variance, similar to the case $T = 1$.

4. EXPERIMENTS

In this section, we simulate the acquisition of an image using the gigavision sensor and we compare the result with a conventional sensor. For this purpose, we use a high dynamic range image 'col-

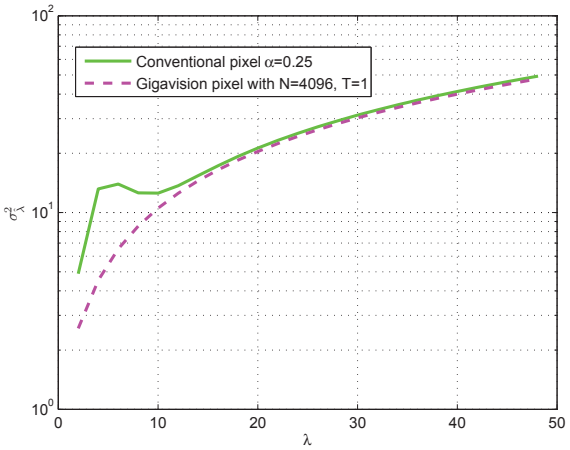
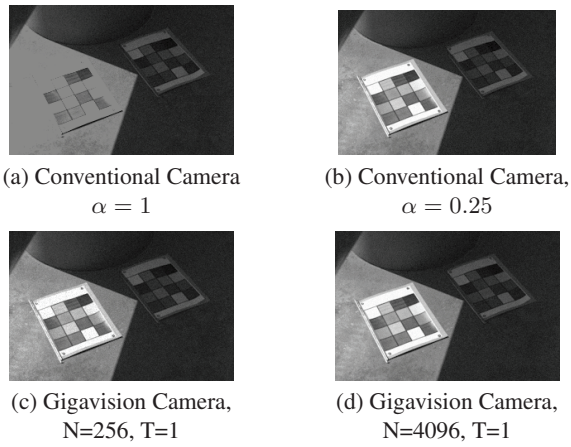


Fig. 6. Estimation error variance of the conventional pixel and the gigavision pixel when λ is in region $[0, 50]$. The quantizer of the conventional pixel has 256 levels and parameter $\alpha = 0.25$. For the gigavision pixel the threshold is $T = 1$, and oversampling factor $N = 4096$.



(a) Conventional Camera $\alpha = 1$

(b) Conventional Camera, $\alpha = 0.25$

(c) Gigavision Camera, N=256, T=1

(d) Gigavision Camera, N=4096, T=1

Fig. 7. Acquired image with the conventional and the gigavision sensor. The image 'colorChecker.hdr' is used to simulate the number of photons at each pixel and each sensor leads to a different estimation of the observed image.

orChecker.hdr' [8]. Each image pixel has a gray level in the range [0, 1]. We assume that the gray level corresponds to the light intensity, i.e. we set λ equal to the level [0, 1024]. For each pixel, we generate the random number of detected photons according to the Poisson distribution of parameter λ and we simulate the behavior of the conventional and the gigavision sensor. For each sensor, the maximum likelihood estimator gives the measured image, i.e. an approximation of the intensity λ. The results are shown in Fig. 7. The top row corresponds to the conventional sensor for $\alpha = 1$ and $\alpha = 0.25$. As expected from the analysis of the previous section, the sensor with $\alpha = 1$ gives a smaller error at low light intensity, but it saturates when the intensity is higher. For the gigavision sensor, we set $T = 1$. The results for $N = 256$ and $N = 4096$ are shown in Fig. 7(c) and Fig. 7(d) respectively. The case $N = 256$ gives a similar result to the conventional pixel with $\alpha = 1$ at low light intensities (this is hard to see in Fig. 7, due to the different scaling of the images) but it does not saturate as the conventional pixel does. On the smooth regions, the higher noise level is clearly visible. This corresponds to the higher estimation error variance determined in the previous section. However, if $N = 4096$, the obtained image, shown in Fig. 7(d), has a noise level comparable with that of Fig. 7(b).

5. CONCLUSION

In this paper, a new kind of image sensor has been proposed based on deep subwavelength binary pixels. We found that, when a group of binary pixels is used to estimate light intensity, the response function is non-linear. The sensor has high sensitivity at low light intensity, and small sensitivity at high light intensity, which makes the sensor suitable for high dynamic range imaging. Since the pixels are binary, they can be very small and the total number is much larger than that of a conventional camera. In this way, noise can be reduced. Also, a thin camera can be achieved easily by placing a lens array on top of the binary sensor. The small pixel pitch size allows to reconstruct a conventional image without the need of decovolution, and reduce the reconstruction noise. The reconstruction algorithm proposed in this paper was based on a simple low-pass filter. Several extensions are possible and under investigation. An interesting one considers space varying filters, in order to obtain the optimal trade-off between noise and bandwidth for different image regions. This allows, for example, to obtain a lower noise level in smooth regions, where vision is more sensitive to artifacts.

6. REFERENCES

[1] Bernd Hoefflinger, *High-Dynamic-Range Vision — Microelectronics, Image Processing, Computer Graphics*, Springer, Berlin Heidelberg, 1998.

[2] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proc. of SIGGRAPH*, 1997, pp. 369–378.

[3] M. D. Grossberg and S. K. Nayar, "High dynamic range from multiple images: Which exposures to combine?," in *ICCV Workshop on Color and Photometric Methods in Computer Vision*, 2003.

[4] Wen-Chung Kao, "High dynamic range imaging by fusing multiple raw images and tone reproduction," *IEEE Transactions on Consumer Electronics*, vol. 54, no. 5, pp. 10–15, February 2008.

[5] S. Kavadias, B. Dierickx, D. Scheffer, A. Alaerts, D. Uwaerts, and J. Bogaerts, "A logarithmic response CMOS image sensor with on-chip calibration," *IEEE Journal of Solid-State Circuits*, vol. 35, no. 8, pp. 1146–1152, August 2000.

[6] James C. Candy and Gabor C. Temes, *Oversampling Delta-Sigma Data Converters — Theory, Design and Simulation*, IEEE Press, New York, 1992.

[7] J. Duparre, P. Dannberg, P. Schreiber, and A. Brauer, "Thin compound-eye camera," *Applied Optics*, vol. 44, no. 15, pp. 2949–2956, May 2005.

[8] L. Meylan and S. Süsstrunk, "High dynamic range image rendering using a retinex-based adaptive filter," *IEEE Transactions on Image Processing*, vol. 15, no. 9, pp. 2820–2830, September 2006.