

SIMULTANEOUS ACQUISITION OF MULTIPLE IMAGES WITH HIGHER DYNAMIC RANGE

René M A Teixeira & Kiyoharu Aizawa

The University of Tokyo

ABSTRACT

Computational photography has redefined the possibilities of conventional photography. New effects can be obtained by mixing modified hardware with software. In this paper we present the basis of a theoretical framework for a computational camera. By using coded apertures, the framework is able to multiplex images of a scene and simultaneously acquire them. As an application, we show a system to capture images with different exposure values, a feature that is often useful in high dynamic range imaging.

This framework is not limited to the acquisition of images with varying exposure values. It is also adaptable to different types of filters, e.g.: color filters. The advantages of this system are the low cost of implementation and ease of adaption to different conditions. However, the need and difficulty of the decoding process can add extra image artifacts.

Index Terms— Computational photography, transmultiplexing, HDR

1. INTRODUCTION

Computational photography is a set of techniques capable of capturing images with enhanced characteristics, which would be difficult to produce with conventional cameras. A computational camera is able to record images with depth information, wider field of view, or larger dynamic range [1]. This is possible by changing the elements of the camera (e.g. lenses and apertures), and adding extra software [2][3].

In this paper we propose a theoretical framework for a transmultiplexer [4], to be used in a computational camera. This camera is composed of a special type of aperture (coded aperture), and an ordinary image sensor. Apart from the coded aperture, no other special hardware is required. The purpose is to record multiple images simultaneously, each under different filtered characteristics.

Coded apertures have a long history in astronomy and medical imaging. Its need came from the necessity of improved signal-to-noise ratio (SNR) in an environment where lenses were not possible. The traditional technique places an aperture mask in the path of light. Unlike traditional coded aperture that uses binary masks, the aperture mask of our proposal is multi-symbol (ternary). In our example, each sym-

bol represents either transparent, semi-transparent or opaque elements. The position of each element within the mask is determined by a code sequence. The nature of the filtering performed by the aperture depends on the intended characteristic to be extracted from the scene: amplitude, wavelength, or phase. In this work we show the use of amplitude filters and their respective transparencies.

In the proposed computational camera, the scene is imaged through the lenses and the aperture mask. Each channel (see section 4 for definition) of the mask registers the same scene, but with a small modification introduced by the channel filter. The channel images are optically combined and recorded by sensor. The expectation is that the recorded image can be decoded and the images of the channels can be developed. We show an example of how the framework can be applied to capture multiple images, with different exposure values, plus an added dynamic range enhancement due to light spreading. This type of image is usually of interest for high dynamic range techniques known as bracketing and camera response estimation [5][6].

2. PREVIOUS WORKS

A large variety of works exists in the literature related to computational cameras. A detailed introduction to the topic was given by Fortunato [2], Zhou, and Nayar [3].

Computational cameras have been largely used for light-field capture. Veeraraghavan et al [7] used masks in the cameras to record heterodyed light-field and perform refocusing. Ihrke et al [8] developed a theory of multiplexed acquisition of the plenoptic function. In a similar way, Liang et al [9] used programmable apertures for the acquisition of the light-field.

Coded masks have been used in other types of computational cameras. Mohan et al [10] used coded masks in a super-resolution camera. Instead of moving the camera by sub-pixel distances, coded apertures were used for sampling. Grosse et al [11] used coded apertures to increase the depth-of-field of a projector. A similar technique is used for depth measurement from defocus, as done by Zhou et al [12].

Perhaps one of the most successful uses of computational cameras is related to high dynamic range imaging. Robertson [5] and Debevec [6] developed methods to increase the dynamic range of photographs by using multiple exposures of

the same scene (bracketing). Yasuma et al [13] used attenuators on top of the pixels of the sensor for increased dynamic range effects (assorted pixels). Rouf et al [14] employed star filters in order to spread the light of a scene, thus enhancing the dynamic range of the images. These works are well related to ours as we will obtain multiple images (like Yasuma et al), albeit simultaneously, and use spreading of light in order to enlarge the sensor’s dynamic range, as Rouf et al, but with coded masks. In our case, the mask will cause each channel image to overlap on the sensor, effectively making all pixels of the sensor record the light of all the channels.

3. CODED APERTURES

Coded aperture is a technique used in photography in substitution of the traditional single aperture/pinhole. An aperture is an opening in an optical system that primarily limits the amount of light through the system. It also limits the field of view, depth of field and resolution of the camera.

Fenimore [15] showed that coded apertures can be used to achieve a better signal to noise ratio in the imaging process. In the presence of a coding mask with multiple openings, the light that reaches the sensor is actually a combination of the images projected by each opening of the mask. This effect is modeled by Equation 1.

$$x = (s * h) + b = \sum_m s[m] \cdot h[n - m] + b \quad (1)$$

Where s is the scene being captured, h is the aperture mask and x is the image projected on the sensor. The symbol $*$ means convolution. The original s can be recovered by the use of a decoding pattern g . The ideal decoding pattern is the one that satisfies $h * g = N\delta$. The symbol δ is the Kronecker delta. N is the number of open elements in the mask. A set of binary patterns that satisfies these requirements was developed by Gottesman and Fenimore [16] and it is called Modified Uniformly Redundant Array (MURA). By using the decoding pattern, s can be estimated using the Equation 2.

$$\hat{s} = x * g = \sum_m x[m] \cdot g[n - m] \quad (2)$$

This process correctly recovers a single image of the scene, with enhanced signal-to-noise ratio characteristics. Further modifications must be done in order to recover images that represent the scene under different characteristics.

4. TRANSMULTIPLEXING

In this section, the theory of coded apertures is extended to accommodate the recovery of images with different properties, like varying exposure value.

The imaging pipeline is shown in Figure 1. Similarly to the traditional coded aperture, the imaging model can be described by: $x_M = s * h$, where x_M is the multiplexed image

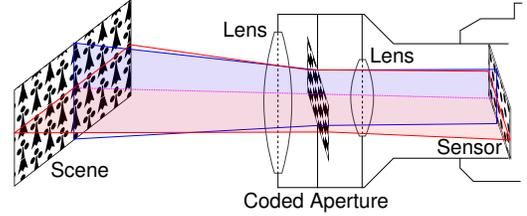


Fig. 1. Imaging pipeline.

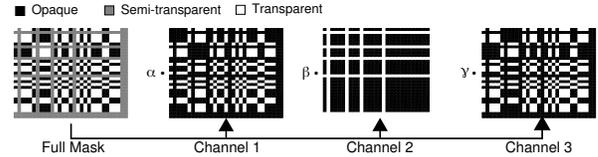


Fig. 2. Mask decomposition.

captured by the sensor. The decoding process is described by: $\hat{s} = x_M * g_i$, where g_i is the decoding filter of each channel.

In order to capture different images, light must be split into different versions of the original scene. The idea behind this methodology is inspired on the theory of transmultiplexers [4]. A transmultiplexer is a device that multiplexes and demultiplexes signals. It is important to emphasize that, in this case, information is spread over space, and not over time or frequency bands.

According to the elements used in the code of the mask, different channels can be defined. Strictly, a channel is defined by a sub-mask h_i of h . Any sub-mask defines a valid channel, with better or worse efficiency. We define the sub-masks as proposed by Lüke and Busboom [17], where transparent elements of the mask are considered as one channel while all the semi-transparent elements are considered as belonging to a different sub-mask. A physical implementation of the system in Figure 1 would have only one physical mask, but the coding mask is logically split into sub-masks (channels). As the mask h is the result of the linear combination of all its sub-masks, the coding model changes to:

$$x_M = s * h = s * (h_1 + h_2 + \dots + h_n) \quad (3)$$

Figure 2 shows an example of how a coding mask can be split into binary sub-masks. Each channel of the mask has an associated filtering property e_i , in this case transparency levels α , β , and γ . Because of the difference in transparency, and therefore, attenuation, it is possible to encode images with different exposures. The code sequence used for this example was derived from a sequence with perfect correlation properties [18] (see section 6).

Suppose a transmultiplexer with N channels, as shown in Figure 3. The scene s will be coded by the N channels h_i of the mask, and at the same time the transformation e_i will be applied, which is equivalent to the transmittance of each

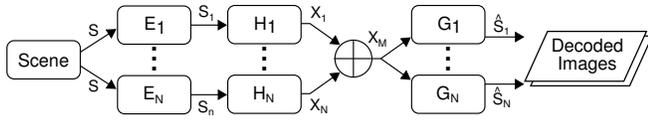


Fig. 3. System model with N channels.

channel. This process results in N images x_i that are spatially modulated and combined on the sensor (equation (3)).

The image recorded by the sensor is not intelligible, because of the superposition of the channel images. A user would have no practical use for it in its raw state. Decoding is necessary to produce usable images.

The coding is an optical process. The decoding of x_m , on the other hand, is a digital process. For that, we define N filters $G_i = \frac{1}{H_i}$, where $H_i = \mathfrak{F}\{h_i\}$.

The decoding filters are defined in the Fourier domain. Variables displayed in capital letters are the Fourier transformed version of the variables in small letters. For many cases, G_i will be unstable due to the presence of small amplitudes in the frequency spectrum of h_i . For this reason, G_i is defined as:

$$G_i = \begin{cases} \frac{1}{H_i} & \text{if } |H_i| > T \\ 0 & \text{if } |H_i| \leq T \end{cases} \quad (4)$$

Where T is a threshold.

By means of the inverse filters, the images can be decoded out of X_M by the operation in equation (5), which is similar to a deconvolution.

$$\hat{S}_i = X_M \cdot G_i \quad (5)$$

Finally, the complete transfer function of the system in Figure 3 can be defined:

$$\frac{\hat{S}_i}{S} = E_i H_i G_i + G_i \sum_{\substack{n=1 \\ n \neq i}}^N (E_n H_n) \quad (6)$$

The limits of the summation mean that all i from the lower limit to the upper limit are included, except for $n = i$. The equation (6) has two well-defined terms. The first term ($E_i H_i G_i$) is the system point spread function (SPSF) related to the transcoding of the channel h_i , i.e. the decoding of channel h_i by its decoding filter g_i . The second term is a cross-talk element. It is the result of images coded by h_i being decoded by g_j , where $i \neq j$.

In a transmultiplexer with perfect reconstruction (PR) [4], the following conditions would be satisfied:

$$PR \Rightarrow \begin{cases} E_i H_i G_i = E_i \\ G_i \sum_{\substack{n=1 \\ n \neq i}}^N (E_n H_n) = 0 \end{cases} \quad (7)$$

From the transfer function it is easier to understand that the natural decoding filter G_i is the inverse of H_i . This way,

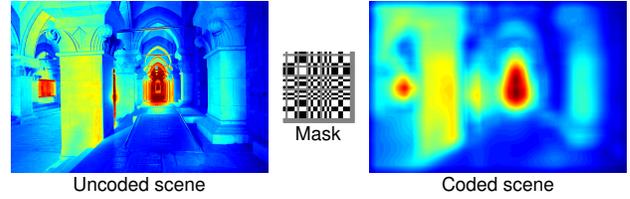


Fig. 4. Light spreading effect on a scene (fake colors).

the first term of the transfer function reduces to E_i , while the second term reduces to zero.

When a mask code is designed, these conditions should be kept in mind. A variety of codes exist in the literature, with different levels of compatibility and trade-off. If the mask channels are not uncorrelated one to another, further processing is required to eliminate the added artifacts.

5. SATURATION AND LIGHT SPREADING

It is worth noting that sensor saturation is an undesirable effect that should be avoided. Its effects in the Fourier domain are explained by Wetzstein et al [19]. For small defocus values, error propagation due to saturation is constrained to small regions. The area where noise spreads over in the decoded images enlarges as the defocus grows.

In a sensor, two parameters constrain the dynamic range of a photograph, the saturation level and the noise level. Previous works dealing with dynamic range enhancement have tried to circumvent the limitations imposed by the saturation level, often considered the main problem in high dynamic range imaging.

As we try to avoid saturation as much as possible, the extra enhancement in dynamic range of the proposed framework comes from the lower limit. Due to a light spreading effect this camera is able to image scene details that otherwise would be too dark. When a coded mask is used, the energy of a point of light is spread over an area proportional to the amount of defocus, following the code sequence of the mask. Because of this effect, for a given exposure time, the pixels of the sensor will take longer to saturate. This fact is exemplified in Figure 4. The red regions denote nearly saturated pixels and the blue areas show under-exposed pixels, for the same exposure time. Note that details that are saturated in the uncoded version does not saturate as quickly in the coded version.

Light spreading was first observed by Trentacoste et al [20]. A tangible gain from the use of the technique could not be observed by them because of decoding artifacts. Later, Rouf et al [14] used star filters for similar purposes in a more complex method.

The spreading of light caused by a mask is a function of not only the element spacing of the mask but also a function of the scene. For this reason, the gain provided by one channel

of the mask does not necessarily add up linearly to the gain provided by the other channels.

6. APPLICATION SIMULATION

In high dynamic range imaging the acquisition of photos with different exposure values is usually a desirable feature. Multiple exposures are used to recover the characteristic “camera curve” and from the curve create a radiance map of the scene [5][6]. Although two exposures are usually enough to determine the maximum dynamic range of the scene, the recovery of the camera curve requires more exposures for better fitting.

A simulation was created in order to exemplify this framework. We intended to show how different versions of a scene could be obtained by using coded apertures. For this purpose we set the filters E_i as light attenuators. These attenuators can be physically implemented by means of neutral density filters. For the simulations, we used the mask pattern from Figure 2. The pattern is a 2D array, converted from a 1D sequence with perfect periodic autocorrelation, of size 31×31 [18]. The transparency of the elements of the channels e_i were set as $\alpha = 1.0$, $\beta = 0.5$, and $\gamma = 0.1$.

As input scene, we used images in HDR formats, like Radiance HDR and OpenEXR. The input files were processed using the MATLAB[®] environment. Coding and decoding were performed by the methods explained by equations (3) and (5). The image sensor was simulated as a 14-bit sensor, which is comparable to commercial cameras. The decoded images were saved as 8-bit JPEG files. As explained in section 5, all scenes were considered as not saturated, what can be obtained by shorter exposure periods. A ternary sequence with 31 elements provided by Lüke et al [18] was implemented as the mask code. The sequence is a one dimensional vector. Let it be called p . The 2D array, necessary to build the coding mask was created by doing $p' = p^T \times p$. The symbol T is the transpose operator and \times is the matrix multiplication operator. The resultant matrix has elements with values in the set $\{-1, 0, +1\}$. The method introduced in [17] was used to identify the channels of the mask. All -1 elements were allocated to one channel. The same was done to all the zero and $+1$ elements. Coding was done by using the code p' . Decoding was performed by using equation (4). Whenever a ternary mask is used, at most four images can be retrieved from the sensor image: one for each channel plus one for the full mask. Because the sub-masks may not be perfectly uncorrelated, cross-talk will introduce noise in the decoded images. Depending on the scene, filtering the decoded image of the cross-talk term $H_i G_n$, which is a known variable, can be helpful.

The simulation results can be seen in figures 5, and 6. These images show the transcoding result of an HDR scene, converted from 14 bits to an 8 bit JPEG file, hence the presence of saturation. Observe that the results show images with different exposure values, being “full mask” the reference and

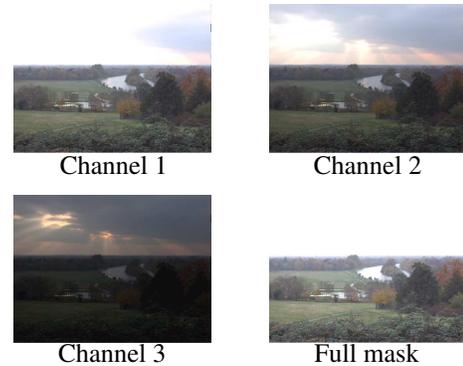


Fig. 5. Different exposures of a scene. Image source: [21]

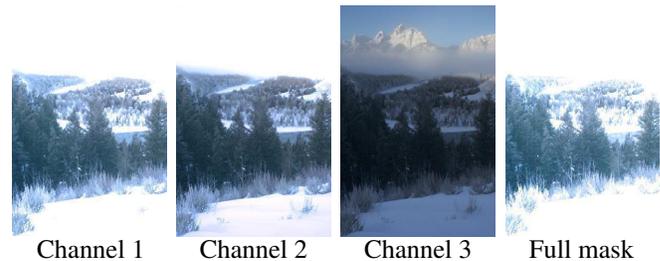


Fig. 6. Different exposures of a mountain. Image source: [22]

also the brightest picture. Full mask decoding means that no channel information was used for this image (i.e. traditional coded aperture). The exposure level of each channel is lower than the exposure level of the full mask because their effective area and transparencies are smaller. Observe that in the brighter channels, details like the grass and the leaves are clearly visible. On lower exposure channels, these details are not visible, while elements like clouds and the mountain are now visible.

7. CONCLUSIONS

We have proposed a theoretical framework for simultaneous acquisition of images in a photographic camera, where the acquired images have different properties. This framework makes use of a coded aperture. The images that are coded by the mask channels are combined on the sensor and recorded as one single image. Further decoding by software is necessary in order to decode the different images from the data recorded by the sensor. We demonstrated the use of the framework in a system where the acquisition of images with different exposure values are desired. This type of problem is often found in high dynamic range imaging. The simulation results showed that images with varying exposure values can be obtained if intensity filters are used as elements of the coding mask. A physical implementation of the system would have to consider sources of errors such as decoding artifacts, diffraction, mask alignment, sensor noise, and resolution.

8. REFERENCES

- [1] S. K Nayar, "Computational cameras: Redefining the image," *Computer*, vol. 39, no. 8, pp. 30–38, Aug. 2006.
- [2] H.E. Fortunato and M.M. Oliveira, "A gentle introduction to coded computational photography," in *24th SIB-GRAPI*, Aug. 2011, pp. 39–55.
- [3] Changyin Zhou and S. K Nayar, "Computational cameras: Convergence of optics and processing," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3322–3340, Dec. 2011.
- [4] A. N Akansu, P. Duhamel, Xueming Lin, and M. de Courville, "Orthogonal transmultiplexers in communication: a review," *IEEE Transactions on Signal Processing*, vol. 46, no. 4, pp. 979–995, Apr. 1998.
- [5] Mark A. Robertson, Sean Borman, and Robert L. Stevenson, "Estimation-theoretic approach to dynamic range enhancement using multiple exposures," *Journal of Electronic Imaging*, vol. 12, no. 2, pp. 219, 2003.
- [6] Paul E Debevec and Jitendra Malik, "Recovering high dynamic range radiance maps from photographs," in *ACM SIGGRAPH 2008 classes*, New York, NY, USA, 2008, p. 31:131:10, ACM.
- [7] Ashok Veeraraghavan, Ramesh Raskar, Amit Agrawal, Ankit Mohan, and Jack Tumblin, "Dappled photography: mask enhanced cameras for heterodyned light fields and coded aperture refocusing," *ACM Trans. Graph.*, vol. 26, no. 3, July 2007, ACM ID: 1276463.
- [8] I. Ihrke, G. Wetzstein, and W. Heidrich, "A theory of plenoptic multiplexing," in *CVPR*. June 2010, pp. 483–490, IEEE.
- [9] Chia-Kai Liang, Tai-Hsu Lin, Bing-Yi Wong, Chi Liu, and Homer H Chen, "Programmable aperture photography: multiplexed light field acquisition," in *ACM Transactions on Graphics (TOG)*, New York, NY, USA, 2008, p. 55:155:10, ACM.
- [10] A. Mohan, Xiang Huang, J. Tumblin, and R. Raskar, "Sensing increased image resolution using aperture masks," in *CVPR*. June 2008, pp. 1–8, IEEE.
- [11] Max Grosse, Gordon Wetzstein, Anselm Grundhfer, and Oliver Bimber, "Coded aperture projection," *ACM Transactions on Graphics (TOG)*, vol. 29, no. 3, pp. 22:122:12, July 2010.
- [12] Changyin Zhou, Stephen Lin, and Shree K. Nayar, "Coded aperture pairs for depth from defocus and defocus deblurring," *International Journal of Computer Vision*, vol. 93, no. 1, pp. 53–72, Dec. 2010.
- [13] F. Yasuma, T. Mitsunaga, D. Iso, and S.K. Nayar, "Generalized assorted pixel camera: Postcapture control of resolution, dynamic range, and spectrum," *Image Processing, IEEE Transactions on*, vol. 19, no. 9, pp. 2241–2253, Sept. 2010.
- [14] M. Rouf, R. Mantiuk, W. Heidrich, M. Trentacoste, and C. Lau, "Glare encoding of high dynamic range images," in *CVPR*. June 2011, pp. 289–296, IEEE.
- [15] E. E. Fenimore, "Coded aperture imaging: predicted performance of uniformly redundant arrays," *Applied Optics*, vol. 17, no. 22, pp. 3562–3570, Nov. 1978.
- [16] Stephen R. Gottesman and E. E. Fenimore, "New family of binary arrays for coded aperture imaging," *Applied Optics*, vol. 28, no. 20, pp. 4344–4352, Oct. 1989.
- [17] Hans Dieter Lüke and Axel Busboom, "Binary arrays with perfect odd-periodic autocorrelation," *Applied Optics*, vol. 36, no. 26, pp. 6612–6619, 1997.
- [18] H. D Lüke, "Sequences and arrays with perfect periodic correlation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 24, no. 3, pp. 287–294, May 1988.
- [19] G. Wetzstein, I. Ihrke, and W. Heidrich, "Sensor saturation in fourier multiplexed imaging," in *CVPR*. June 2010, pp. 545–552, IEEE.
- [20] Matthew Trentacoste, Cheryl Lau, Mushfiqur Rouf, Rafal Mantiuk, and Wolfgang Heidrich, "Defocus techniques for camera dynamic range expansion," *Proceedings of SPIE*, vol. 7537, no. 1, pp. 75370H–75370H–11, Jan. 2010.
- [21] JALOXÁ, "WebHDR," <http://www.jaloxa.eu/web-hdr/example.shtml>, 2013.
- [22] "HDR tools," http://ttic.uchicago.edu/~cotter/projects/hdr_tools/, 2011.