CEIG - Spanish Computer Graphics Conference (2012)
Isabel Navazo and Gustavo Patow (Editors)

Computational Photography

Diego Gutierrez†, Belen Masia‡ and Adrian Jarabo§

Graphics and Imaging Lab, Universidad de Zaragoza, Spain

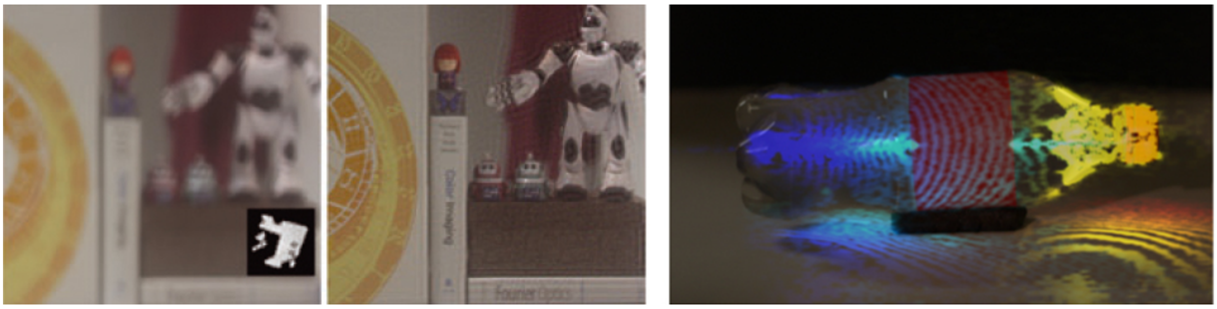

Figure 1: *Two examples of enhanced imaging achieved through computational photography techniques.* ***Left:*** *defocus deblurring by means of perceptually optimized coded apertures [MPCG12].* ***Right:*** *color-coded time-lapse visualization of light at picosecond resolution [Ras12].*

Abstract

Computational photography is an emergent multidisciplinary field that combines topics in optics, computer graphics, image and signal processing. The objective is usually the development of new algorithms or systems to capture images overcoming the limitations of conventional techniques. Images are not simply captured as 2D projections of a scene, they are rather coded before reaching the sensor in a way that can later be decoded and interpreted. This coding-decoding process allows to recover information lost in the traditional capture process. In the present text we will go through the fundamentals and will then offer some examples of recent research projects, a representative cross-section covering some of the most relevant results produced. This cross-section will include plenoptic imaging, coded apertures, novel analysis of light transport and recent advances in ultra-fast imaging, as well as an introduction to computational displays.

1. Introduction

Photography has been around for over 150 years. The relatively recent advent and definite establishment of digital photography has marked one of the biggest revolutions in the field. Nowadays, every single decent cell phone comes equipped with a digital camera, which means that there are over a thousand million cameras out there. However, the basic ways in which a photograph is taken has remained the same: some optics converge light onto a sensor, and that produces the image. We have replaced photons on a chemical sensor for electrons on a digital sensor, but that is that. For having moved into the digital realm, where ones and zeroes can be coded, manipulated and decoded at will, it seems obvious that we have not yet unleashed all the potential latent in digital photography.

Actually, any digital camera already performs *a lot* of computations every time a picture is taken. After the analog to digital conversion (electrons to bytes), the camera will

† diegog@unizar.es
‡ bmasia@unizar.es
§ ajarabo@unizar.es

most likely perform all of the following steps (unless you are shooting in raw format): demosaicing, denoising, color space conversion (from sensor to sRGB), white balancing, color enhancement, gamma encoding and final quantization before saving the image as a jpeg file. But there is actually no reason to stop there, or to not revisit any of these steps.

Computational photography is an emerging field of research that aims at exploiting all the exciting capabilities of digital photography. It is commonly described as lying at the intersection of the fields of optics, computer vision, graphics and signal processing. As such, it is clearly a multidisciplinary field. Its goal is to overcome the shortcomings of traditional photography by i) sampling the plenoptic functions in smarter ways, ii) coding the information arriving at the sensor and then iii) decoding it in order to extract more meaningful information from the scene being photographed. Quoting Ramesh Raskar, leader of the Camera Culture Group at MIT Media Lab: "Photographs will no longer be *taken*; they will be *computed*".

In this brief tutorial we will first introduce a few basic aspects on which computational photography is founded. In particular, we will define the plenoptic function as the complete description of the light present in a scene. Next, Section 3 is devoted to plenoptic imaging, that is, the techniques aimed at sampling the plenoptic function. The next sections will dwell on some of the most representative areas of computational photography: coded photography in Section 4, techniques devoted to analyzing light transport in Section 6 and we will also cover computational displays in Section 5. Finally, Section 7 will gloss over recent advances in ultra-fast imaging, which has offered impressive results in the last year.

2. Fundamentals: introducing the plenoptic function

Computational photography requires knowledge from many different backgrounds which cannot be summarized in a tutorial such as this one. Also, determining which are the fundamentals upon which it builds is a daunting task. Having at least a basic knowledge of geometric optics, ray matrix operations, Fourier optics; mastering photography concepts (e.g. depth of field); being familiar with signal processing and work in the frequency domain or dealing with acquisition/display hardware can be considered desirable knowledge for researchers in the field. Covering basic concepts from the mentioned disciplines, however, lies outside the scope of this tutorial.

Here, for the sake of brevity, we will limit ourselves to introducing the plenoptic function. In the end, computational photography deals with the capture, processing and display of light in ways aside of conventional photography, and the plenoptic function is a complete model of the light in a scene [AB91]. Adelson and Bergen define the plenoptic function as a complete representation of the visual world; it gives the light in a scene as a function of position (x,y,z), direction (θ,ϕ), wavelength λ and time t:

$$L = L(x,y,z,\theta,\phi,\lambda,t) \quad (1)$$

A conventional photograph samples just two dimensions of the plenoptic function (x,y), integrating over a certain range of all the other dimensions. The shutter is responsible for the integration over t, while the lens and aperture determine the angular integration over (θ,ϕ). The sensor, and specially the Bayer array laid over it, are responsible for the sampling and integration in λ. When we speak about computational photography, we commonly refer to how its goal is the *enhancement* of the abilities of conventional digital (or analog) photography. As such, a great amount of work in the field has been done in plenoptic imaging, that is, sampling of the dimensions of the plenoptic function (see Section 3).

Another key concept in computational photography, which is that of *light field*, is closely related to the concept of plenoptic function. A light field represents the radiance in a point in space at a given direction. Wavelength and the temporal dimension are typically not taken into account, leaving a 5D function. A key observation is that in free space, this is 5D function is really 4D [LH96,GGSC96], since the radiance along a ray does not change. The term was first used long ago [Ger39], but the concept the way we understand and use it today was defined by Levoy and Hanrahan in their seminal work [LH96], where they propose a parametrization which is well-suited for acquisition and rendering. This well-known parametrization consists in defining a ray by its intersection with two defined planes, the radiance of a ray being $L(u,v,s,t)$. The parameters u,v and s,t are the coordinates of the intersections with the planes, typically the camera and focal plane, respectively. A year earlier McMillan and Bishop [MB95] had already proposed a way of representing, sampling and reconstructing the plenoptic function for image-based rendering applications; and concurrently with Levoy and Hanrahan, Gortler et al. [GGSC96] presented, with the same 5D to 4D observation and analogous parametrization for light fields, a way to capture, represent and render a light field, which they term *Lumigraph*.

Methods to capture light fields –that is, capturing angular and spatial information of a scene as opposed to only spatial information such as a conventional camera does– have ranged from involved gantries with a moving camera, to camera arrays (see Figure 2, left), to commercial light field cameras (see Figure 2, right). A survey on computational cameras was published by Zhou and Nayar [ZN11].

3. Plenoptic Imaging

This section briefly introduces some concepts and the current state of the art in plenoptic imaging. For a more in-depth discussion, we refer the reader to the excellent recent survey by Wetzstein and colleagues [WILH11].

The main challenge in plenoptic imaging is trying to avoid

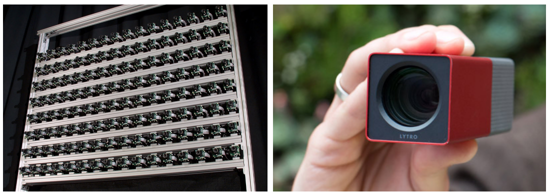

Figure 2: ***Left:*** *Stanford Multi-Camera Array, for capturing light fields [Sta12].* ***Right:*** *Lytro's light field hand-held camera, now commercially available [Lyt12].*

the integration along multiple domains that occurs at each sensor pixel. The time at which photons arrived, the angular information, wavelengthÉ all that information is lost after the integration step which bakes it all into one triplet: the R,G,B values that, combined, yield the final color of the pixel. The reason why cameras were designed that way is to mimic how we see the world around us: as a trichromatic, two-dimensional image.

But it does not have to be that way, necessarily. Nature provides us with some amazing examples of how evolution can take different routes. For instance, the mantis shrimp (see Figure 3) can see 12 color channels, perceives circular polarization and resolves depth using trinocular vision. It is probably the most complex visual system known for a living species, and it remains a complete mystery why it evolved that way.

Figure 3: *The mantis shrimp possesses the most sophisticated visual system known in Earth for a living creature (photo by Roy Campbell).*

So-called plenoptic cameras, for instance, use arrays of micro lenses in front of the sensor to keep angular information of the incoming light, at the cost of spatial resolution. Narrow band filters allow us to separate incoming wavelengths. Coded apertures retain depth information, high dynamic range imaging extends the limited contrast that traditional sensors capture while other techniques like the flutter shutter sample the time domain in smart ways. All these approaches have something in common: through the combination of novel hardware and smart computation, they manage to capture more information from the plenoptic function, which in turn yields new images that would not be possible to acquire with traditional techniques. Figure 4 shows an overview of the different plenoptic image acquisition approaches, divided according to the methodology used: i) single camera, single shot, ii) single camera, multiple shots and iii) multiple cameras multiple shots. In the following, we offer a brief summary of the main existing techniques, grouped according to the plenoptic dimension they focus on. Again, we refer the reader to the recent survey by Wetzstein et al. from a more comprehensive description [WILH11].

High dynamic range Perhaps the best known, it aims at extending the dynamic range (contrast) that can be captured. While the human visual system can adapt to a wide range of luminances through the process of dynamic adaptation, a camera's sensor can only capture about two orders of magnitude in luminance, in effect flattening the look of the scene. By taking a series of pictures with different exposures, the original dynamic range of a scene can be recovered [DM97]. Other techniques include assorted pixels, gradient cameras or split aperture imaging. The book by Reinhard et al. [RWP*10] is probably the publication of reference in this field.

Spectral imaging The most widely used technology to capture spectral information (color) is based on spatially varying filters, such as the Bayer filter. The trade-off is that in order to gain spectral information, one needs to sacrifice spatial resolution. Many other techniques are being explored, though, and some have even found their way into commercial hardware. This is the case for instance of the three-layered Foveon sensor filters, which allow for full spatial resolution capture. Other approaches include optical splitting trees or assorted pixels [NN05].

Angular information Extending the angular (directional) information of incoming light is equivalent in computer graphics terminology to *light field acquisition*. The introduction of light fields to computer graphics occurred simultaneously with two seminal papers in 1996 [LH96, GGSC96]. Big multi camera arrays and gantries are been progressively replaced by cameras with smart combinations of lenses, mirrors or masks. The works by Ng [Ng05] and Veeraraghavan et al. [VRA*07] deserve special attention: the former made integral photography actually manageable, without the need for huge camera arrays (in effect setting the grounds for what became later *Lytro*, the first company commercializing a plenoptic camera). The latter introduced frequency multiplexing to achieve single-sensor light field acquisition.

Space New techniques have evolved to overcome the limited spatial resolution of photographies, which is bounded by the physical size of elements such as the sensor or the

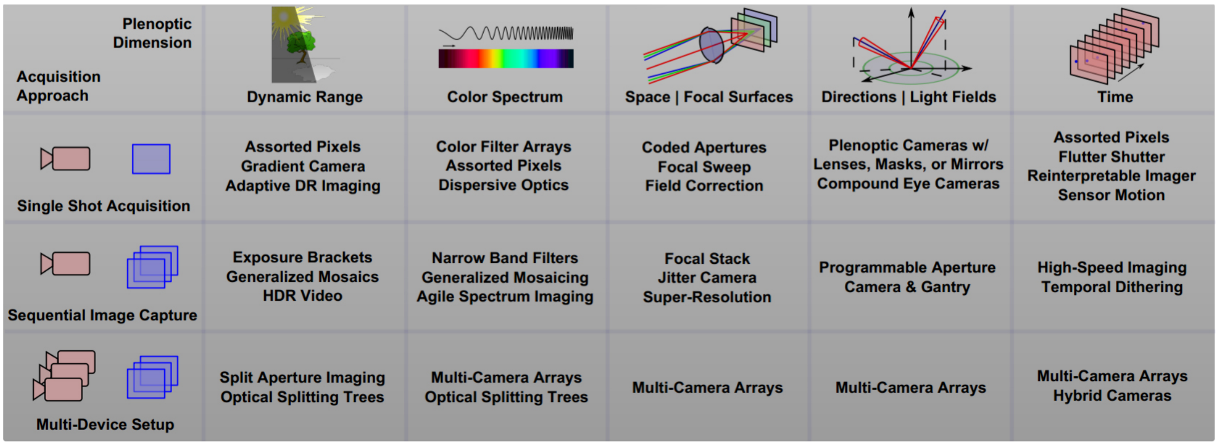

Figure 4: *Classification of the different approaches in plenoptic image acquisition (from [WILH11])*

diffraction limit. This is known as super-resolution photography (see for instance [CMN11]). Giga-pixel imaging refers to techniques that seamlessly stitch together a series of smaller images [KUDC07].

Motion and time Time-lapse photography or high-speed imaging are two of the main techniques to capture motion. Sometimes, the latter is helped by high-speed illumination, as in the case of the seminal work by Harold Edgerton or the recent advances in ultra-fast imaging that allow to see light in motion as it traverses a static scene [Ras12] (see also Section 7). Highly related to this is the field of motion deblurring, for instance by means of a coded temporal sampling [RAT06].

4. Coded Photography

Coded photography can be considered a subfield within computational photography. It comprises the series of techniques in which the information is somehow coded during the capture process. This information can be later decoded to obtain richer representations of the scene than those available in traditional photography. The coding can take place in any of the elements which are part of the imaging process, such as the light (coded illumination), the aperture (coded apertures or heterodyning) or the exposure (coded exposure), in which the present section will focus.

Coded illumination has long been used in the field of computer vision for tasks such as 3D reconstruction [SPB04, AGD07]. Applications in computational photography include the well-known flash/no-flash techniques [PSA*04, ED04, ARNL05], which aim at removing flash artifacts, or obtaining denoised images, by combining photographs taken with and without flash; dark flash photography, which builds on ideas from the former and exploits infra-red and ultraviolet light to capture images in low-light conditions [KF09]; or depth edges recovery via a multi-flash device such as that proposed by Raskar et al. [RTF*04]. In addition to the aforementioned techniques, which modify the flash light in a conventional camera to enhance its capabilities, there is a whole series of works which rely on coded illumination and use one or more projectors (light emitters) to capture and/or analyze the light transport matrix of a scene. These are explored in Section 6. Finally, the polarization of light has been widely exploited for capturing different aspects, such as reflectance [GCP*10].

Coded exposure imaging consists in coding the incoming light in the temporal dimension, modifying the way light is integrated in time within each exposure (i.e. photograph) taken by the camera. Perhaps the key work in this area is the *flutter shutter* by Raskar et al. [RAT06], in which they modify the shutter of a camera, making it open and close with a particular binary pattern during the exposure time of a photograph. This coding in time allows them to eliminate motion blur in the captured image, as long as movement in the captured scene is linear in time (i.e. constant velocity). An extension of their work to allow the system to deal with more general motions was performed by Ding et al. [DMY10]. Related to these techniques, and worth mentioning, are the works of Ben-Ezra and Nayar [BEN03] and Tai et al. [TDBL08], in which they use the trade-off between spatial and temporal resolution of an imaging system to obtain motion blur-free imagery.

Coded apertures find its origins in the field of astronomy, where they were used to code the direction of incoming rays as an alternative to focusing imaging techniques which rely on lenses [ItZ92]. From the family of different patterns that arised, the MURA patterns (Modified Uniformly Redundant Array) [GF89], an evolution of the URA patterns [FC78], were probably among the most popular ones.

Figure 5: *Left: Disassembled Canon EOS 50mm f/1.8 lens. Right: The lens with a coded aperture inserted. Adapted from [MPCG12].*

A coded aperture is essentially an attenuating mask which blocks part of the incoming light, and can be placed at various points in the optical path (see Figure 5). Placed near the sensor, it allows the reconstruction of a 4D light field from the information recorded in the (2D) sensor [VRA*07]. This, done by appropriately modulating the incoming light field in the frequency domain, is known as spatial heterodyning. Placing a mask of spatially varying transmittance next to the sensor has also been used for capturing images of higher dynamic range [NM00, NB03].

Another option is to place the mask at the lens, allowing refocusing of images at full resolution (assuming the scene being captured contains only Lambertian objects). Defocus deblurring and depth extraction are two of the main applications of coded apertures in computational photography, and complementary to each other. When deblurring is sought, coded apertures are designed so that they have a favorable response in the frequency domain, i.e. do not attenuate high frequencies and avoid zero-crossings. If, on the contrary, depth recovery is the objective, symmetric apertures containing zero-crossings in the frequency domain are preferred [LFDF07]. Hiura and Matsuyama proposed a four-pinhole aperture to approximate the depth of a scene and obtain a deblurred image of it, using multiple images [HM98]. Liang et al. also necessitated multiple images, they used Hadamard-based patterns, and achieved post-capture refocusing and scene depth estimation [LLW*08]. Levin et al. [LFDF07] try to achieve depth estimation and an all-in-focus image form a single photograph, by means of an optimal aperture pattern and a deconvolution method based on a sparsity prior of image derivatives. Pursuing the same objective, and inspired by the depth from defocus literature, Zhou et al. use pairs of coded apertures to recover both a fairly accurate depth map of the scene and focused images of it [ZLN09]. Concurrently, a framework was presented to evaluate coded apertures for defocus deblurring and obtain near-optimal ones by means of a genetic algorithm [ZN09]. Masia and colleagues aimed at extending this work, and analyzed the case of non-binary apertures [MCPG11]. Recently, perceptual metrics were introduced in the optimization process leading to the obtention of coded apertures, yielding patterns which behaved better for defocus deblurring [MPCG12]. An example of these apertures, together with a deblurred captured image and the corresponding recovered image, is shown in Figure 1 (left).

5. Computational Displays

Given its obvious relation to the field, we have also included here computational displays, which explore the co-design of optical elements and computational processing while taking into account the characteristics of human perception (for a thorougher review we refer the reader to [WLGH12]). These displays enable viewing no longer just 2D images or video; they also enable the perception of depth and the visualization of 4D light fields. As a side note, high dynamic range displays, pioneered by the work of Seetzen et al. [SHS*04], can also be considered part of this category.

As 3D becomes more and more common in both movies and videogames, displays capable of displaying this type of content have proliferated. Typically, these displays require the viewer to wear some kind of glasses or similar equipment, encumbering the viewing experience. More recently, devices which allow to see 3D content are becoming more and more common, specially when it comes to mobile devices. The technology behind most of these glasses-free displays is more than a century old, since it relies on parallax barriers [Ive03] or lenslet arrays [Lip08]. Only very recently some other approaches have been presented that enable glasses-free 3D displays and rely on a different principle, based on tomography [WILH11, LWH*11]. Volumetric displays are a different type of 3D displays which have existed for a while, without ever reaching the mass market [Fav05]. Different technologies have been presented, though, both inspiring and promising. Another type of displays which, as of today, have been unable to reach the consumer market are holographic displays. The theory behind the generation of holograms is quite old [Gab49], yet the required hardware remains too sophisticated, expensive, and with significant limitations; nevertheless, research in the field continues producing valuable advances [BBV*10].

The most basic characteristic of a 3D display is being able to present a different image to each eye, thus creating *binocular disparity*, one of the cues responsible for our perception of depth. Although the classification of depth cues varies from one author to another, a fairly thorough list is that of Cutting and Vishton [CV95]. From all the cues that the human visual system combines to perceive depth, only some of them (e.g. occlusion, relative size, aerial perspective or height in the visual field) are present in 2D imagery, and a subset of them, namely binocular disparity, motion parallax, accommodation and convergence, cannot be provided by conventional displays. Classic 3D (stereoscopic) displays that employ glasses achieve binocular disparity (i.e. presenting a different image to each eye) either by spatially multiplexing the images (e.g. anaglyphs, which use color filters, or polarized glasses) or with temporal multiplexing (e.g. shutter glasses). These are referred to as *stereoscopic* displays. The term *automultiscopic* displays is applied to those displays which enable the perception of depth without glasses or other similar equipment. These displays not only offer the depth cue of binocular disparity, but also of motion parallax, enhancing the viewing experience. Finally, volumetric and holographic displays –which we will leave out of this discussion from now on due to their practical restrictions– are capable of providing the depth cues of accommodation and convergence too.

The working principle of both parallax barriers and integral imaging is quite similar. In the first, a layer of occluders is placed in front of the actual screen. Occluders (which can be only vertical stripes, offering only horizontal parallax, or a grid, offering vertical and horizontal parallax) are devised in such a way that they select the rays coming from the screen that are visible to each of the viewers' eyes. In order for the system to work, the image shown in the screen needs to be processed adequately. In the same spirit, in integral imaging devices an array of tiny lenses is set in front of the screen, effectively redirecting rays coming from the screen to the viewers' eyes in an appropriate manner. Again, the images shown in the screen are composed of multiple images adequately interleaved. These two techniques have co-existed for decades as the only practical approaches for automultiscopic displays. In 2010, Lanman et al. [DLR10] built on the concept of parallax barriers and stacked two modified LCD panels to create content-adaptive parallax barriers, increasing brightness with respect to their traditional counterparts. Even more recently, a new type of displays were presented that relied on tomographic techniques for the display of light fields on multi-layer architectures. Both prototypes based on attenuating layers of transparencies [WILH11] and dynamic prototypes relying on the polarization state of stacked LCD panels [LWH*11] exist (see Figure 6). Finally, multilayer architectures have been combined with directional backlighting to enable wider fields of view and larger separation between viewers, while preserving a thing form factor similar to that of a conventional LCD display, creating the so-called tensor displays [WLHR12]. Although still in an early state of development, worth mentioning are also a new generation of computational displays which can be tailored for the specific viewing capabilities of the observer by moving the light field to fit the focus range of the viewer [VFPR12].

6. Capturing and Analyzing Light Transport

For the last decade, there is an increasing interest on capturing and analysing the light transport in a scene in both the fields of computer graphics and vision. These techniques base on reconstructing the linear transformation, usually called light transport matrix, that models the dependence between the incident light field in the scene and the resulting

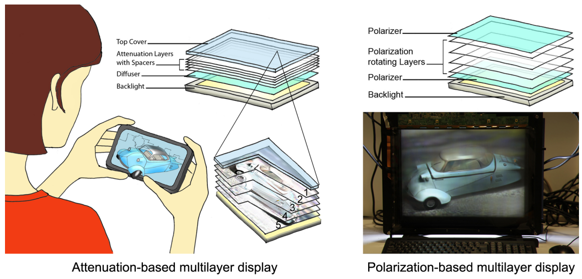

Figure 6: ***Left:*** *Conceptual sketch showing the structure of an attenuation-based multilayer display (from [WILH11]).* ***Right:*** *Structure of a polarization-based multilayer display, and photograph of a prototype (from [LWH*11]).*

image. Thus, the photograph on any scene, no matter how complex, can be expressed as [NRH03]:

$$p = T * I \tag{2}$$

where T is the transport matrix, and p and I are the resulting photograph and the input light field respectively, both of them in vector form. T takes into account all the light paths from the light source to the camera sensor occurring in the scene, including complex paths such as interreflections, caustics or scattering through participating media.

Computing T have been approached for both synthetic and real scenes. In this document we focus only on the latter. For synthetic scenes, we refer to the existing literature on the field of *Precomputed Radiance Transfer* [Ram09]. To capture the transport matrix T the common approach consists of illuminating the scene with coded illumination, using one or more projectors as light sources. This light sources project patterns on the scene, to reconstruct the transfer matrix from each pixel of the projector to the image.

Knowing the light transport matrix T gives an effective tool for relighting real-world captured scenes, allowing to re-render the scene under any incident light field I, accounting for complex light paths, materials and geometry [DHT*00, MPDW03], as shown in the examples in Figure 7. However, capturing T from real-world scenes presents two some important challenges: first, a full reconstruction of the matrix might be extremely large. Additionally, to reconstruct accurately T hundreds to thousands of photographs are required. Most research efforts have been dedicated to efficiently compute and model T, based on sampling adaptively the incoming light space [FBLS07], using low-rank matrix reconstruction [WDT*09], compressive sensing [PML*09] or optical computing methods [OK10].

Additionally, Sen et al. [SCG*05] demonstrated that, by exploiting Helmholtz reciprocity, the transport matrix T can be used also to obtain the view of the scene from the projector's point of view, or even to reconstruct occluded objects

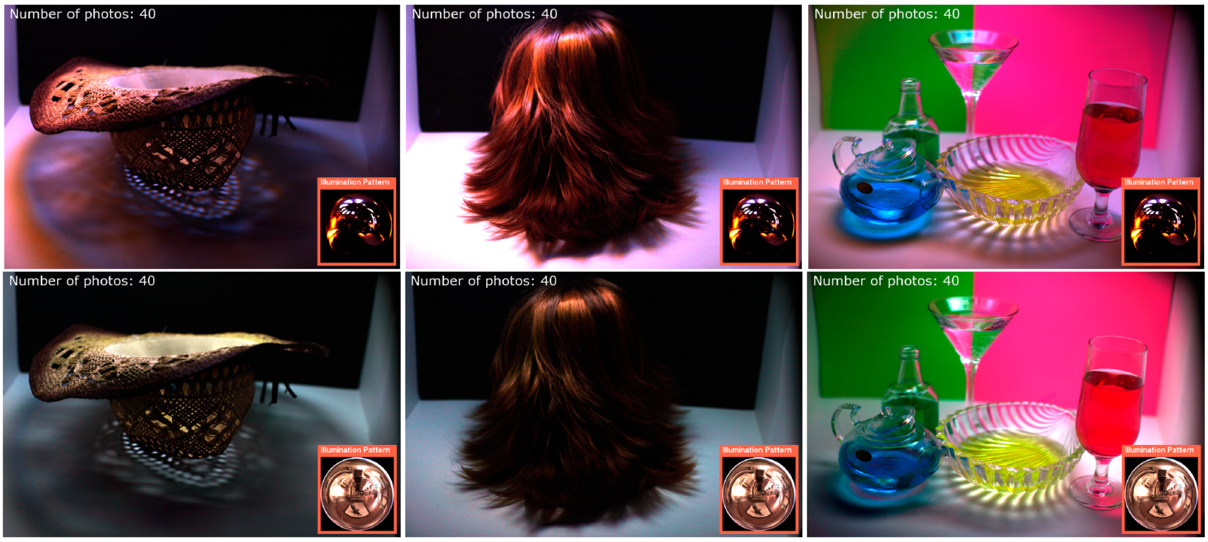

Figure 7: *Relighting results of a set of scenes with complex geometry and reflectance under different natural illumination. Note how high-frequency shadows or caustics are captured accurately. Image from [OK10].*

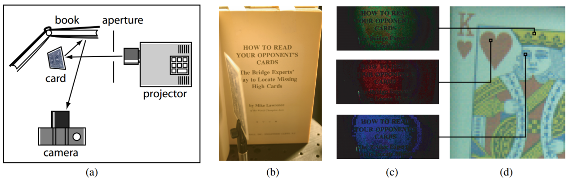

Figure 8: *Example of using* dual photography *[SCG*05] with indirect illumination to reconstruct the non-visible side of the card (d). The layout of the scene is shown in (a), where the camera sees only the back of the card and the book (b). Image from [SCG*05].*

using the indirect illumination to reconstruct T, as shown in Figure 8.

Finally, projecting illumination patterns in the scene has been also demonstrated as a useful approach to separate illumination components in a scene. Nayar et al. [NKGR06] project a set of high-frequency patterns to disambiguate the illumination between direct and indirect components in a scene. Following work by Mukaigawa et al. [MYR10] uses a similar approach to separate single and multiple scattering, further separating the latter into multiple bounce components.

7. Femtophotography: capturing a trillion frames per second

High-speed photography, including seminar work by Harold Edgerton, allows capturing the motion of objects moving at high velocities, allowing very impressing photographs (e.g. the impact of a bullet with an apple). However, the time resolution of these photographs is insufficient to capture the light moving through the scene. Femtophotography [Ras12] is a recent technique that captures videos where each frame has an effective exposure time of roughly two picoseconds ($2 * 10^{-12}$ seconds). This allows capturing how the light propagates through the scene and interacts with the matter. In comparison with high-speed photography, femtophotography captures events occurring around 300000 times faster. Figure 1 (right) shows an example of the data captured with the technique, where a pulse travels through a bottle interacting with the media (diluted milk). The duration of the recorded events in the video is around one nanosecond.

To capture light in motion, the system built consists of ultra-fast laser and detector (femtosecond- and picosecond-accurate respectively), and mathematical reconstruction techniques. The scene is illuminated with the laser, which emits light pulses on the order of femtoseconds. Then, the photons reflected by the scene are captured by a streak tube [Ham]. This is an imaging device capable of capturing photons at an effective framerate of around a trillion (10^{12}) frames per second. The streak camera captures 1D movies of the scene (i.e. *scanlines* in the x-axis), with an effective exposure time of 1.71 picoseconds.

Capturing light at such small time resolution makes recording light with sufficient brightness nearly impossible. To solve it, it is taken into account the statistical repeatability of the light-matter interactions in the scene: millions of repeated measurements are performed using an indirect *stroboscopic* illumination in order to reduce significantly the signal-to-noise ratio of the capture. Finally, to reconstruct a two-dimensional video from the data recorded by the streak camera, several 1D videos are captured varying the view direction in the y-axis using a system of mirrors. This 1D movies are used as the rows of the final 2D movie.

Capturing the time-of-flight of the light provides more information than regular photographs or videos, because the integration-time of the light in the sensor is very small. Using this data have been demonstrated useful to reconstruct geometry out of the line-of-sight (*look around corners*) [VWG*12], to capture BRDFs from a single point of view by using indirect reflections through the scene [NZV*11] or to disambiguate between different components of the illumination, such as direct, indirect and subsurface light transport [WOV*12]. These are three recent examples from the multiple possible applications and research avenues that might be opened by femtophotography.

8. Conclusions

The intention of this tutorial has not been to provide an exhaustive study of computational photography techniques and applications, but rather to offer a representative cross-section of the field. At the intersection of several other fields (computer graphics, vision, optics, ...), computational photography provides new and exciting avenues of novel research, and we hope this brief document helps inspire other groups.

We have already seen a transition from the so-called *epsilon photography* focusing of expanding the capabilities of traditional photography by combining conventional photographs taken with varying parameters, to *coded photography*, where the concept of *taking* a photograph is abandoned in favor of *computing* it in a coding-decoding process. Conferences specifically devoted to the field have arised and consolidated over several years (IEEE International Conference on Computational Photography, ICCP), and we have seen research in the field reaching the mass market in the form of a new type of camera [Lyt12]. We are also seeing novel related fields that keep pushing the boundaries of computational photography, such as computational displays or computational models of perception. More and more recent research works combine these concepts yielding vibrant, thought-provoking papers. And this is one of the most fun parts of this research: everything is up for grabs; the future is wide open.

Acknowledgements

This research has been funded by the European Commission, Seventh Framework Programme, through the projects GOLEM (Marie Curie IAPP, grant agreement no.: 251415) and VERVE (Information and Communication Technologies, grant agreement no.: 288914), and by the Spanish Ministry of Science and Technology (TIN2010- 21543). Belen Masia is supported by an FPU grant from the Spanish Ministry of Education and by a NVIDIA Graduate Fellowship.

References

[AB91] ADELSON E. H., BERGEN J. R.: The plenoptic function and the elements of early vision. In *Computational Models of Visual Processing* (1991), MIT Press, pp. 3–20. 2

[AGD07] ALBITAR I., GRAEBLING P., DOIGNON C.: Robust structured light coding for 3d reconstruction. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on* (oct. 2007), pp. 1 –6. 4

[ARNL05] AGRAWAL A., RASKAR R., NAYAR S. K., LI Y.: Removing photography artifacts using gradient projection and flash-exposure sampling. *ACM Trans. Graph. 24*, 3 (July 2005), 828–835. 4

[BBV*10] BLANCHE P. A., BABLUMIAN A., VOORAKARANAM R., CHRISTENSON C., LIN W. E. A.: Holographic three-dimensional telepresence using large-area photorefractive polymer. *Nature 468*, 7320 (Nov. 2010), 80–83. 5

[BEN03] BEN-EZRA M., NAYAR S.: Motion deblurring using hybrid imaging. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on* (june 2003), vol. 1, pp. I–657 – I–664. 4

[CMN11] COSSAIRT O., MIAU D., NAYAR S. K.: Gigapixel Computational Imaging. In *IEEE International Conference on Computational Photography (ICCP)* (Mar 2011). 4

[CV95] CUTTING J. E., VISHTON P. M.: *Perception of Space and Motion*. Academic Press, 1995, ch. Perceiving Layout and Knowing Distances: The integration, relative potency, and contextual use of different information about depth. 5

[DHT*00] DEBEVEC P., HAWKINS T., TCHOU C., DUIKER H.-P., SAROKIN W., SAGAR M.: Acquiring the reflectance field of a human face. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques* (New York, NY, USA, 2000), SIGGRAPH '00, ACM Press/Addison-Wesley Publishing Co., pp. 145–156. 6

[DLR10] DOUGLAS LANMAN MATTHEW HIRSCH Y. K., RASKAR R.: Content-adaptive parallax barriers: optimizing dual-layer 3D displays using low-rank light field factorization. *ACM Trans. Graph. 29*, 6 (2010). 6

[DM97] DEBEVEC P. E., MALIK J.: Recovering high dynamic range radiance maps from photographs. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques* (New York, NY, USA, 1997), SIGGRAPH '97, ACM Press/Addison-Wesley Publishing Co., pp. 369–378. 3

[DMY10] DING Y., MCCLOSKEY S., YU J.: Analysis of motion blur with a flutter shutter camera for non-linear motion. In *Proceedings of the 11th European conference on Computer vision: Part I* (Berlin, Heidelberg, 2010), ECCV'10, Springer-Verlag, pp. 15–30. 4

[ED04] EISEMANN E., DURAND F.: Flash photography enhancement via intrinsic relighting. *ACM Trans. Graph. 23*, 3 (Aug. 2004), 673–678. 4

[Fav05] FAVALORA G. E.: Volumetric 3D displays and application infrastructure. *IEEE Computer 38* (2005), 37–44. 5

[FBLS07] FUCHS M., BLANZ V., LENSCH H. P., SEIDEL H.-P.: Adaptive sampling of reflectance fields. *ACM Trans. Graph. 26*, 2 (June 2007). 6

[FC78] FENIMORE E. E., CANNON T. M.: Coded aperture imaging with uniformly redundant arrays. *Appl. Opt. 17*, 3 (Feb 1978), 337–347. 4

[Gab49] GABOR D.: Microscopy by Reconstructed Wave-Fronts. *Royal Society of London Proceedings Series A 197* (July 1949), 454–487. 5

[GCP*10] GHOSH A., CHEN T., PEERS P., WILSON C. A., DEBEVEC P.: Circularly polarized spherical illumination reflectometry. *ACM Trans. Graph. 29*, 6 (Dec. 2010), 162:1–162:12. 4

[Ger39] GERSHUN A.: The light field, moscow 1936. *Journal of Mathematics and Physics XVIII* (1939), 51–151. Translated by P. Moon and G. Timoshenko. 2

[GF89] GOTTESMAN S., FENIMORE E.: New family of binary arrays for coded aperture imaging. *Applied Optics*, 20 (1989), 4344–4352. 4

[GGSC96] GORTLER S. J., GRZESZCZUK R., SZELISKI R., COHEN M. F.: The lumigraph. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques* (New York, NY, USA, 1996), SIGGRAPH '96, ACM, pp. 43–54. 2, 3

[Ham] HAMAMATSU: Guide to streak cameras. http://sales.hamamatsu.com/assets/pdf/hpspdf/Guidetostreak.pdf. [Online, last accessed 01-June-2012]. 7

[HM98] HIURA S., MATSUYAMA T.: Depth measurement by the multi-focus camera. In *IEEE Conference on Computer Vision and Pattern Recognition* (Washington DC, USA, 1998), IEEE Computer Society. 5

[ItZ92] IN 'T ZAND J.: *Coded Aperture Imaging in High-Energy Astronomy*. PhD thesis, University of Utrecht, 1992. 4

[Ive03] IVES F. E.: Parallax stereogram and process of making same. U.S. Patent 725,567, 1903. 5

[KF09] KRISHNAN D., FERGUS R.: Dark flash photography. *ACM Trans. Graph. 28*, 3 (July 2009), 96:1–96:11. 4

[KUDC07] KOPF J., UYTTENDAELE M., DEUSSEN O., COHEN M. F.: Capturing and viewing gigapixel images. *ACM Trans. Graph. 26*, 3 (July 2007). 4

[LFDF07] LEVIN A., FERGUS R., DURAND F., FREEMAN W.: Image and depth from a conventional camera with a coded aperture. *ACM Transactions on Graphics 26*, 3 (2007). 5

[LH96] LEVOY M., HANRAHAN P.: Light field rendering. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques* (New York, NY, USA, 1996), SIGGRAPH '96, ACM, pp. 31–42. 2, 3

[Lip08] LIPPMANN G.: Épreuves réversibles donnant la sensation du relief. *Journal of Physics 7*, 4 (1908), 821–825. 5

[LLW*08] LIANG C., LIN T., WONG B., LIU C., , CHEN H.: Programmable aperture photography: multiplexed light field acquisition. *ACM Transactions on Graphics 27*, 3 (2008). 5

[LWH*11] LANMAN D., WETZSTEIN G., HIRSCH M., HEIDRICH W., RASKAR R.: Polarization fields: Dynamic light field display using multi-layer LCDs. *ACM Trans. Graph. 30*, 6 (2011). 5, 6

[Lyt12] Lytro Inc., July 2012. URL: `http://www.lytro.com`. 3, 8

[MB95] MCMILLAN L., BISHOP G.: Plenoptic modeling: an image-based rendering system. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques* (New York, NY, USA, 1995), SIGGRAPH '95, ACM, pp. 39–46. 2

[MCPG11] MASIA B., CORRALES A., PRESA L., GUTIERREZ D.: Coded apertures for defocus deblurring. In *Symposium Iberoamericano de Computacion Grafica* (Faro, Portugal, 2011). 5

[MPCG12] MASIA B., PRESA L., CORRALES A., GUTIERREZ D.: Perceptually-optimized coded apertures for defocus deblurring. *Computer Graphics Forum* (To appear, 2012). 1, 5

[MPDW03] MASSELUS V., PEERS P., DUTRÉ P., WILLEMS Y. D.: Relighting with 4d incident light fields. *ACM Trans. Graph. 22*, 3 (July 2003), 613–620. 6

[MYR10] MUKAIGAWA Y., YAGI Y., RASKAR R.: Analysis of light transport in scattering media. In *Computer Vision and Pattern Recognition, 2010* (2010). 7

[NB03] NAYAR S., BRANZOI V.: Adaptive Dynamic Range Imaging: Optical Control of Pixel Exposures over Space and Time. In *IEEE International Conference on Computer Vision (ICCV)* (Oct 2003), vol. 2, pp. 1168–1175. 5

[Ng05] NG R.: Fourier slice photography. *ACM Trans. Graph. 24*, 3 (July 2005), 735–744. 3

[NKGR06] NAYAR S. K., KRISHNAN G., GROSSBERG M. D., RASKAR R.: Fast separation of direct and global components of a scene using high frequency illumination. *ACM Trans. Graph. 25*, 3 (July 2006), 935–944. 7

[NM00] NAYAR S., MITSUNAGA T.: High dynamic range imaging: spatially varying pixel exposures. In *IEEE Conference on Computer Vision and Pattern Recognition* (Hilton Head, SC, USA, 2000), IEEE Computer Society, pp. 472–479. 5

[NN05] NARASIMHAN S., NAYAR S.: Enhancing Resolution along Multiple Imaging Dimensions using Assorted Pixels. *IEEE Transactions on Pattern Analysis and Machine Intelligence 27*, 4 (Apr 2005), 518–530. 3

[NRH03] NG R., RAMAMOORTHI R., HANRAHAN P.: All-frequency shadows using non-linear wavelet lighting approximation. *ACM Trans. Graph. 22*, 3 (July 2003), 376–381. 6

[NZV*11] NAIK N., ZHAO S., VELTEN A., RASKAR R., BALA K.: Single view reflectance capture using multiplexed scattering and time-of-flight imaging. *ACM Trans. Graph. 30*, 6 (Dec. 2011), 171:1–171:10. 8

[OK10] O'TOOLE M., KUTULAKOS K. N.: Optical computing for fast light transport analysis. *ACM Trans. Graph. 29*, 6 (Dec. 2010), 164:1–164:12. 6, 7

[PML*09] PEERS P., MAHAJAN D. K., LAMOND B., GHOSH A., MATUSIK W., RAMAMOORTHI R., DEBEVEC P.: Compressive light transport sensing. *ACM Trans. Graph. 28*, 1 (Feb. 2009), 3:1–3:18. 6

[PSA*04] PETSCHNIGG G., SZELISKI R., AGRAWALA M., COHEN M., HOPPE H., TOYAMA K.: Digital photography with flash and no-flash image pairs. *ACM Trans. Graph. 23*, 3 (Aug. 2004), 664–672. 4

[Ram09] RAMAMOORTHI R.: Precomputation-based rendering. *Foundations and Trends in Computer Graphics and Vision 3*, 4 (2009), 281–369. 6

[Ras12] Femto-photography: Visualizing photons in motion at a trillion frames per second, July 2012. URL: `http://web.media.mit.edu/~raskar/trillionfps/`. 1, 4, 7

[RAT06] RASKAR R., AGRAWAL A., TUMBLIN J.: Coded exposure photography: motion deblurring using fluttered shutter. *ACM Trans. Graph. 25*, 3 (July 2006), 795–804. 4

[RTF*04] RASKAR R., TAN K.-H., FERIS R., YU J., TURK M.: Non-photorealistic camera: depth edge detection and stylized rendering using multi-flash imaging. *ACM Trans. Graph. 23*, 3 (Aug. 2004), 679–688. 4

[RWP*10] REINHARD E., WARD G., PATTANAIK S., DEBEVEC P., HEIDRICH W., MYSZKOWSKI K.: *High Dynamic Range Imaging: Acquisition, Display, and Image-based Lighting*, 2nd ed. The Morgan Kaufmann series in Computer Graphics. Elsevier (Morgan Kaufmann), Burlington, MA, 2010. 3

[SCG*05] SEN P., CHEN B., GARG G., MARSCHNER S. R., HOROWITZ M., LEVOY M., LENSCH H. P. A.: Dual photography. *ACM Trans. Graph. 24*, 3 (July 2005), 745–755. 6, 7

[SHS*04] SEETZEN H., HEIDRICH W., STUERZLINGER W., WARD G., WHITEHEAD L., TRENTACOSTE M., GHOSH A., VOROZCOVS A.: High dynamic range display systems. *ACM Trans. Graph. 23*, 3 (Aug. 2004), 760–768. 5

[SPB04] SALVI J., PAGÃĹS J., BATLLE J.: Pattern codification strategies in structured light systems. *Pattern Recognition 37*, 4 (2004), 827 – 849. <ce:title>Agent Based Computer Vision</ce:title>. 4

[Sta12] The stanford multi-camera array, July 2012. URL: `http://graphics.stanford.edu/projects/array/`. 3

[TDBL08] TAI Y.-W., DU H., BROWN M., LIN S.: Image/video deblurring using a hybrid camera. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on* (june 2008), pp. 1 –8. 4

[VFPR12] VITOR F. PAMPLONA MANUEL M. OLIVEIRA D. G. A., RASKAR R.: Tailored Displays to Compensate for Visual Aberrations. *ACM Trans. Graph. (Proc. SIGGRAPH) 31*, 4 (2012), 87:1–87:12. 6

[VRA*07] VEERARAGHAVAN A., RASKAR R., AGRAWAL A., MOHAN A., TUMBLIN J.: Dappled photography: mask enhanced cameras for heterodyned light fields and coded aperture refocusing. *ACM Transactions on Graphics 26* (July 2007). 3, 5

[VWG*12] VELTEN A., WILLWACHER T., GUPTA O., VEERARAGHAVAN A., BAWENDI M., , RASKAR R.: Recovering ThreeDimensional Shape around a Corner using Ultra-Fast Time-of-Flight Imaging. *Nature Communications 3* (March 2012). 8

[WDT*09] WANG J., DONG Y., TONG X., LIN Z., GUO B.: Kernel nyström method for light transport. *ACM Trans. Graph. 28*, 3 (July 2009), 29:1–29:10. 6

[WILH11] WETZSTEIN G., IHRKE I., LANMAN D., HEIDRICH W.: Computational Plenoptic Imaging. In *Computer Graphics Forum* (2011), vol. 30, pp. 2397–2426. 2, 3, 4, 5, 6

[WLGH12] WETZSTEIN G., LANMAN D., GUTIERREZ D., HIRSCH M.: Computational displays: Combining optical fabrication, computational processing, and perceptual tricks to build the displays of the future. SIGGRAPH Courses, 2012. 5

[WLHR12] WETZSTEIN G., LANMAN D., HIRSCH M., RASKAR R.: Tensor Displays: Compressive Light Field Synthesis using Multilayer Displays with Directional Backlighting. *ACM Trans. Graph. (Proc. SIGGRAPH) 31*, 4 (2012), 1–11. 6

[WOV*12] WU D., O'TOOLE M., VELTEN A., AGRAWAL A., RASKAR R.: Decomposing global light transport using time of flight imaging. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Jun 2012), pp. 1–8. 8

[ZLN09] ZHOU C., LIN S., NAYAR S.: Coded aperture pairs for depth from defocus. In *IEEE International Conference on Computer Vision (ICCV)* (Kyoto, Japan, 2009). 5

[ZN09] ZHOU C., NAYAR S. K.: What are Good Apertures for Defocus Deblurring? In *IEEE International Conference on Computational Photography* (San Francisco, CA, USA, 2009). 5

[ZN11] ZHOU C., NAYAR S.: Computational cameras: Convergence of optics and processing. *Image Processing, IEEE Transactions on 20*, 12 (December 2011), 3322 –3340. 2