MAS.963 Special Topics: Computational Camera and Photography
Fall 2008

MAS.963: Computational Camera and Photography **Fall 2008**

Light fields and Geometric Optics

Prof. Ramesh Raskar September 12, 2008
Notes by Ahmed Kirmani Lecture 2

What is this course all about?

Computational Photography (CP) is an emerging multi-disciplinary field that is at the intersection of optics, signal processing, computer graphics and vision, electronics, art, and online sharing in social networks. In this course we will be studying the 3 major phases of this evolving field [1]. Our approach will be a blend of theoretical understanding of concepts as well as hands on implementation.

[1] summarizes the three phases of CP. The first phase which comprised of about building a super-camera that has enhanced performance in terms of the traditional parameters, such as dynamic range, field of view or depth of field. This was called *Epsilon Photography.* It corresponds to the low-level vision: estimating pixels and pixel features. The second phase called *Coded Photography* is building tools that go beyond capabilities of this super-camera. The goal here is to reversibly encode information about the scene in a single photograph (or a very few photographs) so that the corresponding decoding allows powerful decomposition of the image into light fields, motion deblurred images, global/direct illumination components or distinction between geometric versus material discontinuities. This corresponds to the mid-level vision. The third phase will be about going beyond the radiometric quantities and challenging the notion that a camera should mimic a single-chambered human eye. Instead of recovering physical parameters, the goal will be to capture the visual essence of the scene and analyze the perceptually critical components. This phase is called *Essence Photography* and it may loosely resemble depiction of the world after high level vision processing. It will spawn new forms of visual artistic expression and communication.

Computation vs. Convention

How is CP different from Digital and Film photography? Again, heavily borrowing from [2], Computational photography combines plentiful computing, digital sensors, modern optics, actuators, and smart lights to escape the limitations of traditional film cameras and enables novel imaging applications. Unbounded dynamic range, variable focus, resolution, and depth of field, hints about shape, reflectance, and lighting, and new interactive forms of photos that are partly snapshots and partly videos are just some of the new applications found in Computational Photography. Pixels versus Rays In traditional film-like digital photography, camera images represent a view of the scene via a 2D array of pixels. Computational Photography attempts to understand and analyze a ray-based representation of the scene. The camera optics encode the scene by bending the rays, the sensor samples the rays over time, and the final 'picture' is decoded from these encoded samples. The lighting (scene illumination) follows a similar path from the source to the scene via optional spatio-temporal modulators and optics. In addition, the processing may adaptively control the parameters of the optics, sensor and illumination.

The encoding and decoding process differentiates Computational Photography from traditional *film-like digital photography.* With film-like photography, the captured image is a 2D projection of the scene. Due to limited capabilities of the camera, the recorded image is a partial representation of the view. Nevertheless, the captured image is ready for human consumption: what you see is what you almost get in the photo. In Computational Photography, the goal is to achieve a potentially richer representation of the scene during the encoding process.

What's the future of cameras and photography?

One of the central aims of studying Computational photography is to understand the limitations of current designs and make fundamental changes to the way we image the world. During this course we will learn to think about the future camera and attempt to find answers to the following questions.

1. What will a camera look like in 10,20 years? What will be the form factor of the future camera? Will it continue to look like a thick rectangular block with a lens, a sensor and view finder? [1]

2. How will the next billion cameras change the social culture? The current billion people have been exposed to all kinds of technology and there exists a huge differential with regard to what kind of technology do people at different levels have access to. With all probability, the next billion people will surely be walking around will a cell phone camera which will be perpetually ON capturing and sending video continuous video streams. How will that impact the social culture?

3. How can we augment the camera to support best *image search*? How can we augment the camera so that image search becomes as easy as text. Can we think of camera design that can add enough metadata to the scene at capture time so that images can be indexed and searched efficiently in a way similar to text?

4. What are the opportunities in pervasive recording? What will happen when Google earth goes live? What kind of tasks can you expect to solve collaboratively and what kind of information would you want to share when you can zoom in to and see LIVE any part of the world? Can we crowd source in a way that is similar to ReCAPTCHA to solve problems that heavily task computers? Or in other words can we have a CAMCHA?

---

[1] Talk about the camera of the future! It may just be a thin blackbox with a button, inertial sensors, weather sensors, a GPS receiver, a mobile transreceiver and a display. When you point and shoot at a scene the camera will accurately estimate your location and pose, browse the web for photographs and return a photo of the scene take roughly at the same time of the day under similar weather conditions!

5. How will ultra-high-speed[2]/resolution imaging change us? As hardware becomes cheaper, high speed imaging will be brought to your pocket camera. What can you do with a high speed camera that takes you not just better looking pictures but tells you about the properties of the objects in the scene? Like can we sense the number of calories in a food object by taking its photograph? Can we manufacture portable, marker less motion trackers based on CT technology?

6. How should we change cameras for movie-making, news reporting? Will movies be shot and directed in studios in the future? Who will be the director? Can we send a team of enthusiasts who take random photos of whatever they like, post it on the youtube and a team of artistic people collaboratively spins a story around the seemingly random videos captured by the enthusiasts!

So what's the approach?

Not just USE but CHANGE camera. We can do this at several levels: Optics, illumination, sensor, movement Probes, actuators etc. Imagine using cameras in tandem with other devices like projectors and force feedback. Computer Vision has already squeezed most of the information out pixel bits. Even with that, it is still very challenging to robustly solve many typical vision problems. We have exhausted bits in pixels Scene understanding in challenging. Feature specific imaging or feature revealing cameras can be used solve this problem. The basic ingredients to the CP recipe is: Think in Higher dimensions (beyond 2D: in 3D, 4D, 6D, 8D!). Think about the nature of Light. Play around with illumination, modify the optical path with new elements like crazy optics, crystals, coded masks. Learn from other fields of imaging like IR, X-Ray, Astronomy, CT etc. Process photons and not pixels. The next level of information is contained in the incoming photons.

We live in an age of Digital Renaissance

The sensor market is huge. The projected market for image sensors is 2000000 sensors sold in 2011. Optical mouse uses sensors which are just 32×32 pixels occupy a curiously large share of the sensor market. However the clear winner is the Mobile phone camera sensors, which is much larger than the share of digital cameras. Predictions show that more and more people will replace their digital cameras with mobile phones. Infact most of the users who have bought the new Nokia N95 have completely discarded their digital cameras. This is the best time to study and change the camera, whatever good we can contribute changes the course of future. Task specific cameras compared to generic, feature revealing cameras, high speed cameras, all improvements will shape the way the next billion mobile cameras will be used.

---

[2] Time of flight cameras: Light travels 1foot/ns

Assignment 1: Tips

Its all about using your imagination to add color channels. For example, use illumination from one scene to relight another scene. Or use illumination from a well lit scene and use it to relight a low light picture of the same scene (Gradient domain). For example: Kendall square night picture doesn't convey how tall buildings are. If we had a millions dollars like Paul Debevec (ICT USC) we could construct a light dome than can capture all the reflectance field of a actor and insert her into a new scene using a completely new illumination. Create webpage for your project. Use of any s/w allowed, please always reference it. Of course DON'T use classmate's code.

How do we see the world?

What are the different ways to image the world? One simple way is to just put a sensor in front of object, the image we get is a mush, every point in the world contributes to every point on the sensor (BAD idea!). What else? holes! Each point in the world contributes (ideally) one ray and maps to a unique point on the sensor. We get an inverted image but which is much sharper and keeps on getting sharper till the hole becomes a point called a Pinhole. But the pinhole has other effects such as diffraction[3]. We get the beginnings of a camera now. Camera literally means a chamber or a room[4].

So why pinholes? Because they produce sharp images. But they also throw away a lot of light which reduces the SNR (Signal-to-Noise ratio). To capture more light and yet produce sharp images we use lenses. A lens image can be formed by just using 2 rays one the principal ray and a parallel to focal plane, both emanating from a point on the scene/object. For lens unlike a pinhole, some points that are in focus map to single points on the sensor, other points which are Out-of-focus (OOF) map to Circle-of-Confusion (CoC)[5].

The emphasis of this lecture is to think of Light and analyze it using concepts of Light Field (LF) [6], [7]. Fundamental questions are asked that can be explained using concepts of Light field:

1. Does and OOF image get dark? Based on observation answer is NO. Explained later.

2. Does a zoomed in image get dark? Based on observation answer is YES. Explained later.

3. Why CCD camera sensor (and Cat's eyes) behaves like retroreflectors?
To explain what a retroreflector, we define first 3 kinds of objects: Mirrors, Glossy with Specular highlights and Diffused (Lambertian) as shown in figure 1. If we consider a corner mirror arrangement in a popular jewelery store (see figure 2), it behaves as a retroreflector which essentially means it reflects light back in the same direction that it is coming from. Retroreflective material is made of very small corner cubes, though there is still a limit on the viewing angle.

---

[3] Like a water jet through a nozzle, at some point if the nozzle is too small the water spreads out rather than focus

[4] In several languages like Hindi and Italian, translations of the word *chamber* are strikingly similar to Ka-me-ra

[5] It's not really a circle, it has different shapes depending on the camera impulse response or Point spread Function (PSF). However it is usually approximated by a circle

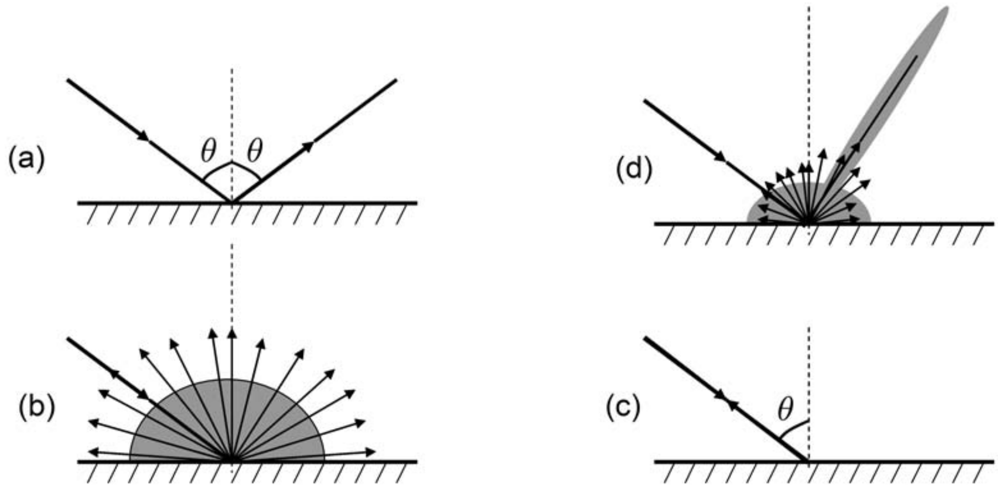

Figure 1: Four kinds of surfaces based on reflectance properties: (a) Mirrored (Specular). (b) Glossy. (c) Lambertian (Diffused) (d) Retroreflectors

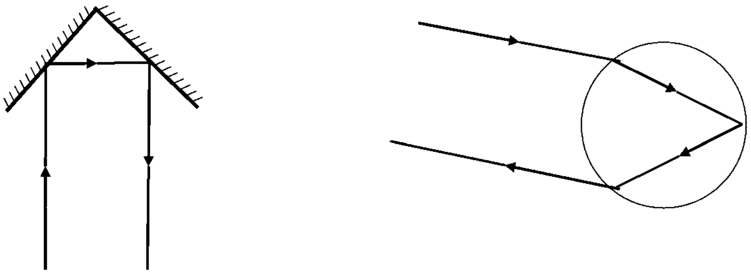

Figure 2: Retroreflector: (Left) The corner cube mirror arrangement (Right) Sphere acts as a retroreflector based on Total Internal reflection. A camera's CCD sensor is covered with a reflecting filter. The overall lens-mirror combination is a retroreflector, although only within a certain cone.

A sphere is a retro reflector (see figure 2) based on Total internal reflection (TIR), though it requires the Index of refraction to be around two[6]. Moon is retroreflective, a cat's eye is retroreflective like a CCD sensor. The latter two are retroreflective because it's a lens mirror combination, both cats eye and CCD camera (thin filter sheet over sensor that is highly reflective). A ray of light incident is going to travel back in the same direction due to duality of light. The CCD retroreflector reflects back in a cone though (see figure 3).

4. How does auto-focus work?
In cheaper cameras autofocusing involves changing of lens setting till contrast is maximized. Also the same principle for a range camera (which also measures time like TOF camera) for depth adjustment, ultrasound ranging, active focusing. Fancier cameras involve Passive methods like Stereo which involves capturing 2 images. But how do we get depth from single lens? Solution is simple: Block one part of the lens at a time and and form 2 images.

---

[6] A rainbow is formed on similar principle of TIR. That's why the rainbow is formed in the same direction as sun

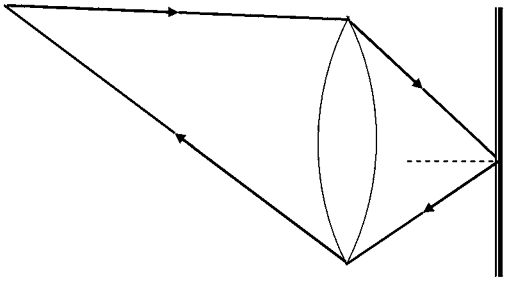

Figure 3: A camera's CCD sensor is covered with a reflecting filter. The overall lens-mirror combination acts as a retroreflector, although only within a certain cone.

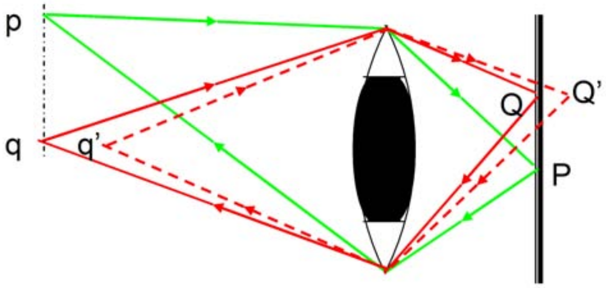

Figure 4: Obtaining 2 images using only one lens for focusing: cover the middle portion. The image of an OOF point (q) exhibits a parallax in the 2 images, while the position of focused point p does not change.

Image formed on the sensor will not change if the scene is in focus. For OOF scenes, the image formed in the 2 cases will be different, and exhibits a shift (see figure 4). Also based on shift, we can infer depth, and change focal length so as to remove the shift and focus the image. In manual SLR cameras, Parallax reference is used and features should align in 2 images if the scene is in focus[7].

Also to note that what you see from camera viewfinder is not what you get on your sensor[8], for example cameras employ optical elements like such as beam splitters, diffusing screens. Also in modern cameras, auto-focusing is done using 2 CCD line sensors which have separate optics. Also, Depth of field[9] (DoF) in a view finder is much larger than the camera. View finders' aperture may be whole main lens and viewfinder's lens but actual photo is taken from the main lens which is a smaller patch and hence will not match (see figure 5). Thus the Field of View (FoV) is matched but the aperture and hence the DoF is

---

[7] Focus on the wall and also look at your finger, the image of finger shifts back on forth. This is the same analogy to lens covering as lens covering. The Left and right eyes form the 2 images

[8] As an experiment to prove this, capture an image od a Point light source. When the source is OOF the the image is a disk. Vary the distance of the source from the camera from closer to further such that the image formed changes from a point to disk. Now take 2 photos, one at decreased aperture and other at increased the aperture. We will observe that we get 2 different images while from the view finder image is the same. This experiment also shows the effect on aperture on Depth of Field (DoF)

[9] the distance within which blur remains constant. Effectively If blur size = 1 pixel the image is in focus

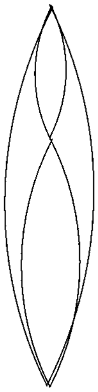

Figure 5: Combined aperture of the View-finder camera system.

not matched[10]. Point shoot cameras are essentially video cameras, they continuously focus before capture.

Geometric Optics

A thin lens is a central element in Geometric optics. It is essentially made of an arrangement of a small angle prism, truncated prisms, a light slab and inverted versions of these elements. This breakdown is called Lens discretization (see figure 6). Thin prims deflect uniformly, that is every ray arriving at the same angle to the normal gets deflected by the same angle. For truncated there is lesser and lesser deflection as they approach the light slab. This deflection is proportional to the distance (t) from the central axis. Finally a Light slab only shifts light rays and does not deflect them.

$$Deflection(t) = k \times t$$

An interesting observation is that if a lens is as wide as the distance between 2 pinholes then the lens image can be obtained by shifting and abutting the pinhole images at the point where the 2 rays meet behind the lens. No matter where the image point is, the rays will always meet at some behind the lens. This view leads to the powerful notion that a lens is an array of pinholes with their own prisms bending the light to meet at one point behind the lens (see figure 7). We can subdivide the lens into segments and each pair has its own deflection. The resulting image can be obtained by shifting, deflecting and adding. The sensor image is thus a superimposed version of pinhole images, if we are lucky we get the same image superimposed and hence a sharp image.

If we use a pinhole array mask to separate the rays then rays coming from diff points do not overlap. The non overlapping of images can be ensured by an arrangement (choosing a pinhole separation and distance from the sensor) such that it abuts the pinhole images and not overlaps them. The pinhole separation and distance of the pinhole array from

---

[10] It is possible to have a depth ordering of scene in which P and Q appear in different orders than they are in the real world and hence it is possible to confuse and cheat the auto focusing mechanism of the camera. Same thing possible by using Shifting patterns

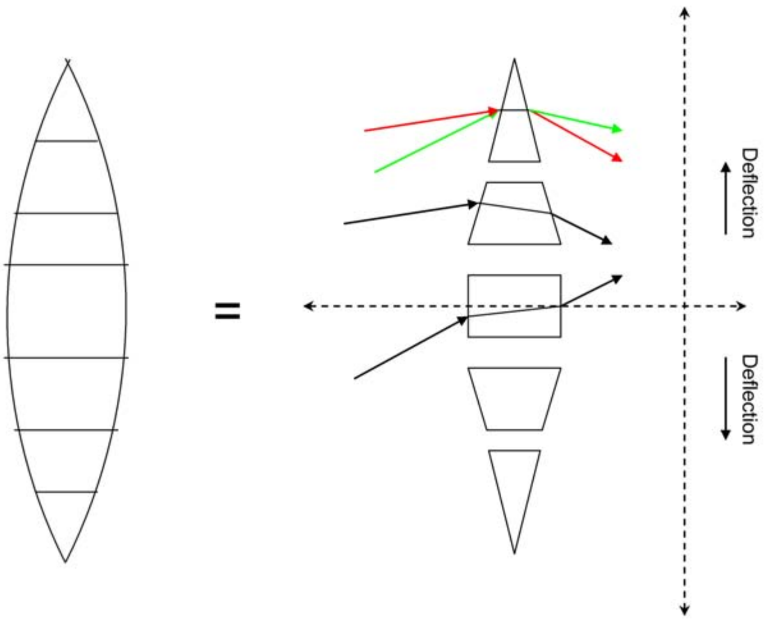

Figure 6: Lens discretization: a Thin prism deflects light rays equally, a truncated prism deflects light less strongly than the prism and a Rectangular slab only shifts the light ray and does not deflect (bend) it.

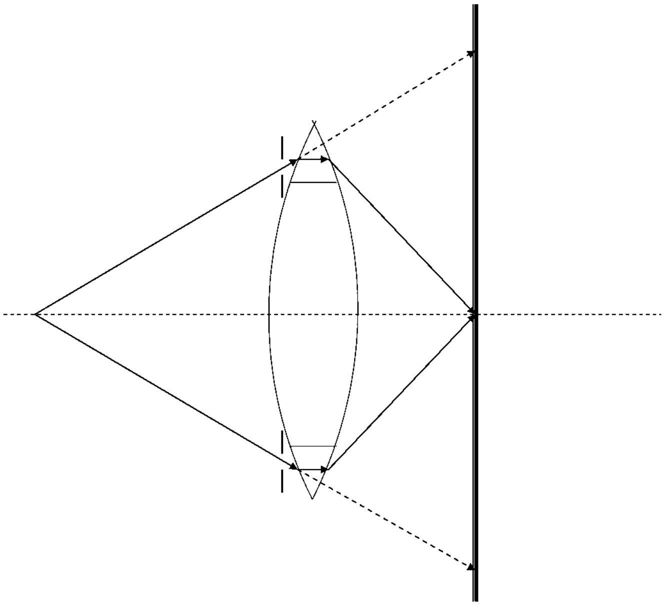

Figure 7: A lens is equivalent to a set of pinholes, each with its own prism bending the light and forcing it to converge at one point behind the camera.

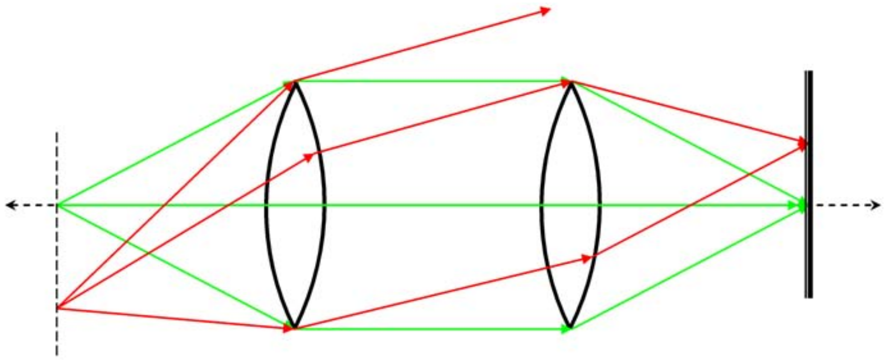

Figure 8: Rays of points which are off axis contributes only a part, coaxial rays contribute fully. Hence resulting in a spatially varying camera PSF.

the sensor can be easily computed based on similar triangles. If we can sample each ray independently, not just the sum of the rays, then it is a powerful technique as shown later. Using a Lenslet array vs. pinholes has the same advantages as using a lens over a pinhole. However in this case the scene is conjugate to lenslet array and each of the micro lenses are in sharp focus on sensor. These are however strong conditions. But this way we can limit diffraction effects and collect more light.

Effects of varying aperture size and Vignetting

In optics, the f-number[11] (sometimes called focal ratio, f-ratio, or relative aperture) of an optical system expresses the diameter of the entrance pupil in terms of the focal length of the lens; in simpler terms, the f-number is the focal length divided by the *effective* aperture diameter [3]. Smaller apertures (pinholes and bigger) allow capturing of all-in-focus images but are impractical due to limited exposures required by lighting or motion. For larger apertures, depth of field effects are observed, including spatially-varying blur depending on depth, and vignetting [8]. Lanman et al. [8] have analyzed the effects of aperture size and particularly analyzed the causes of Vignetting. The camera's PSF is a function of vignetting. At the lens center we get an actual CoC but at the periphery we get an intersection of 2 circles of confusion. This is because of superposition of CoCs of multiple lenses in the camera optical system. Rays of points which are off axis contributes only a part, coaxial rays contribute fully (see figure 8). In OOF objects, disks of corner points gets chopped. It is interesting to note that each point behaves like a digit as shown in Vignetting Synthesis: Superposition Principle [8][12]. Also if we put a coded (mask) at the aperture, the we notice that for OOF parts of the scene, the mask get copied on the sensor while the parts in focus get no pattern and only their intensities are reduced to attenuation of light by the coded aperture (see figure 9).

---

[11] Stay away from photographer's terminology! Too complicated and unnecessary and sometimes even wrong. Just vary aperture size and observe effects

[12] We can use Vignetting Synthesis to create an image of "Happy birthday" using OOF candles!

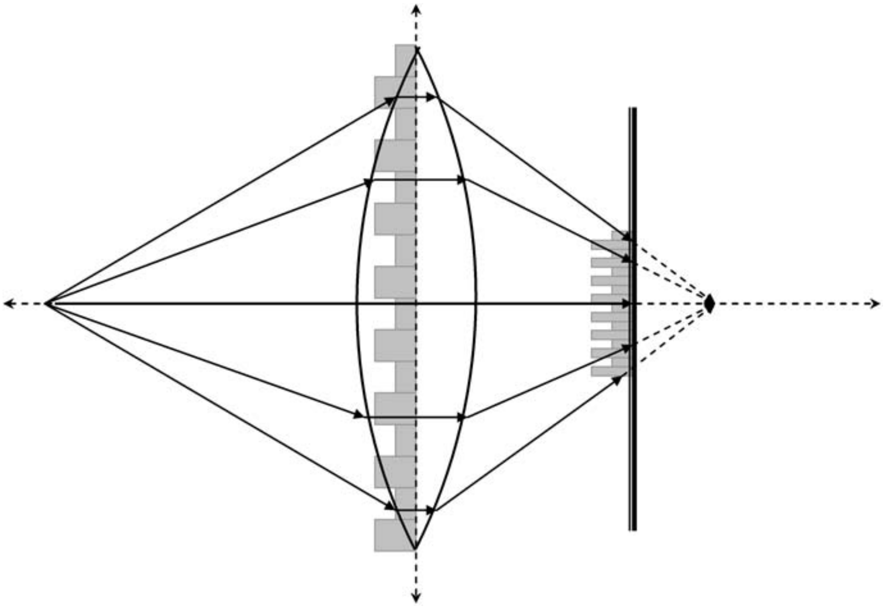

Figure 9: Coded Aperture imaging

Parameterization of rays

In geometric optics, the fundamental carrier of light is a ray. The measure for the amount of light traveling along a ray is radiance, usually denoted by L and measured in watts (W) per steradian (sr) per meter squared (m^2). Several equivalent representation of rays in free space exist [4] (Also see figure 10 and 11). The radiance along all such rays in a region of three-dimensional space illuminated by an unchanging arrangement of lights is called the *plenoptic* function (Adelson 1991). Since rays in space can be parameterized by three coordinates, x, y, and z and two angles[13] θ and ϕ, it is a five-dimensional function (One can consider time, wavelength, and polarization angle as additional variables, yielding higher-dimensional functions). If the region of interest does not contain occluders, the radiance (reflectance) along a ray remains constant and hence we need only four dimensions to represent the ray. In such a case 2 plane parameterizations as well as one point and 2 angle parameterizations suffice for most cases[14]. Also for example, a diamond's appearance can be captured in 4D not 3D. However it is always a good idea to start with 1D and generalize to 2D. Some things still don't generalize though. We choose to mostly use the (x, θ) parameterizations [5]. We alternatively also use the 2 plane parametrization (q, p) representation as well as $(x, theta)$ parametrization (see figure 12).

Transformation of Light fields

We analyze in Flatland (1D). If we limit our analysis to small deflection angles, then $tan(a) \approx a$ and we can represent Light propagation and ray bending by lenses by 2×2

---

[13] Only 2 angles and not 3 because we don't care about the rotation of the ray

[14] No vertical rays representable, degeneracy observed

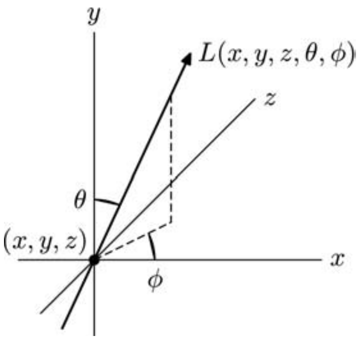

Figure 10: The 5-dimensional plenoptic function. 3 dimensions for the coordinate and 2 dimensions for the angle (*Image source Wikipedia [4]*).

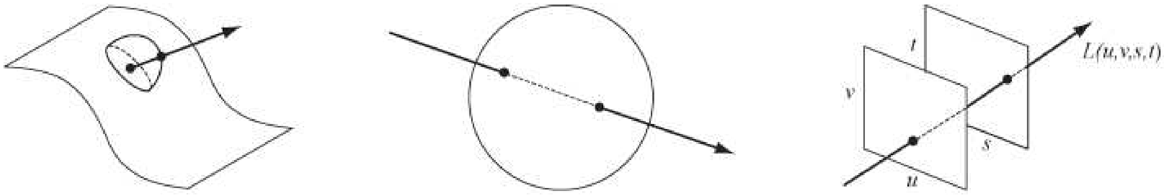

Figure 11: Alternate parameterizations of the Light field including the 2-plane parametrization (right). (*Image source Wikipedia [4]*)

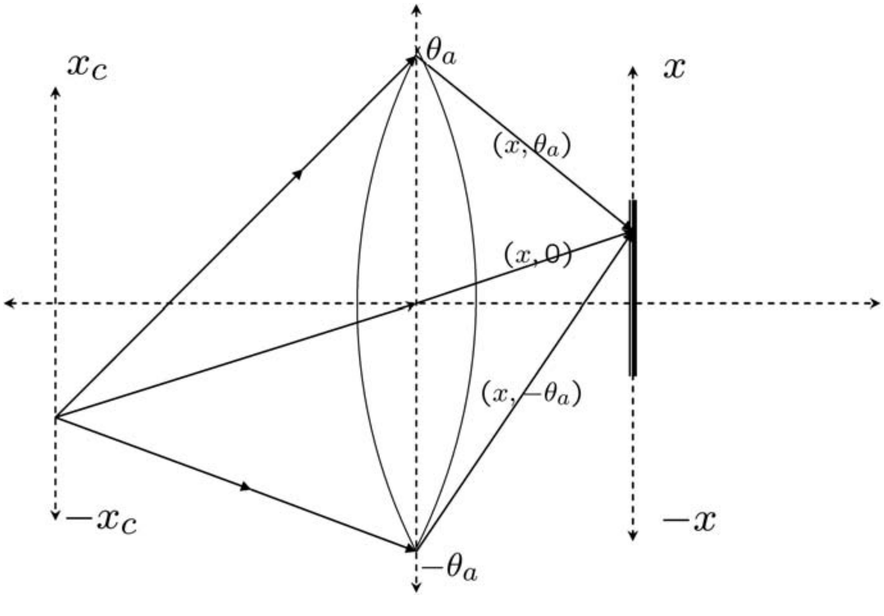

Figure 12: The $(x-\theta)$ parametrization of 2D ray space.

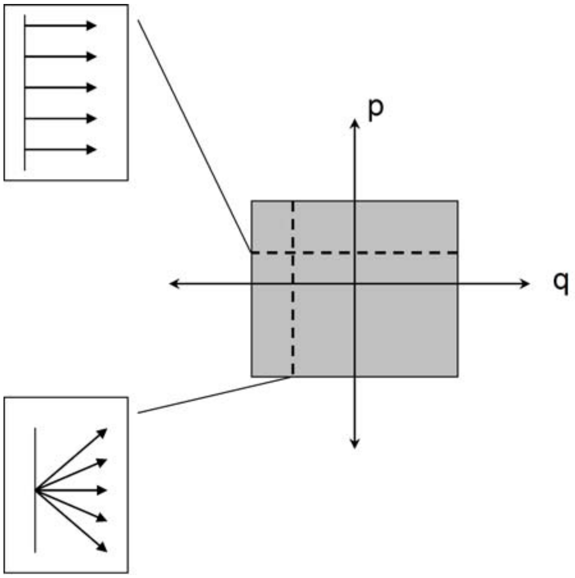

Figure 13: The $q-p$ two plane parametrization of Ray space.Parallel rays from different points have same p but diff q. Rays at different angles from the same point will have different p but same q

transformations. Also any ray gets same transformation so both the transformations are Linear Shift Invariant. We now study some basic transformations.

Propagation in free space

A key observation is that parallel rays from different points have same p but diff q. Rays at different angles from the same point will have different p but same q (see 13). The Propagation transformation results in *shearing* of the Light field. A vertical line in the $q-p$ plane maps to a slanted line after the transformation, though it still passes through the same point on the q-axis because the central ray ($p = 0$ or the q-axis) even after propagation remains unchanged. Higher the value of p (larger angle) the larger the change in q and hence the shear (see 14). A similar explanation can be obtained for the negative q values. Also note that the propagation transform matrix corresponds to a shear transform. Thus propagation of Light field (LF) through free space introduces shear along the q-axis.

Lens transform

For the lens case, we would like to adopt the (x, θ) parameterizations. θ corresponds to the plane of lens. This is not natural since the 2D coordinates x and θ) are not orthogonal in the real world. If a ray hits a thin lens then its position doesn't change, only the angle changes due to deflection and also the deflection is proportional to the distance from the center axis. Its easy to show that the constant of proportionality is $k = -1/f$ since a ray with $p = 0$ deflects to pass through the lens focus. It is easy to show that the thin lens transform results in the *shearing* of LF along the p-axis (see 15). Lenses with small

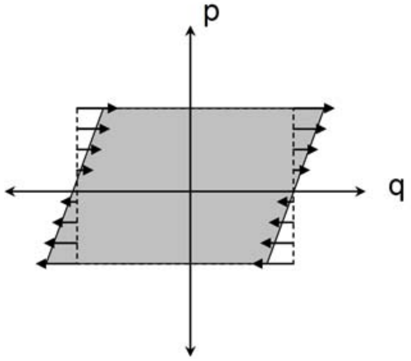

Figure 14: Propagation of Light field through free space introduces shear along the q-axis

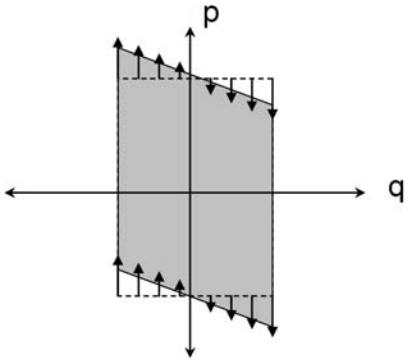

Figure 15: Propagation of Light results in the *shearing* of LF along the p-axis

focal length induce smaller deflection (more shear) while ones with large focal length induce larger deflection (less shear).

Owing to the above 2 transformations, at any given point in time, either the position of the propagating light ray changes or its angle changes. Now we do some Light field (LF) analysis. We fix notation by stating that the x and sensor planes are conjugate to each other. Both x and θ are positive or negative on either side of the central axis. We also want $+x$ and $-x$ in sensor rather than the object so we invert the signs of the point in the object plane. If all rays emanating from a point map to a line (in 2D feature space which has the same x then the point is in focus. OOF points will lie on a slanted line in the 2D parameter space. Defined a *Projection* as (integration or summing up along a certain line). Then a sensor is a line purely along x (see 16). The image is vertical projection obtained by summing up all the intensities along the θ dimension. For an OOF point only a tiny contribution (depending on the degree of out-of-focus) from the original ray is included in the vertical projection and contributions from all other rays are also included (see 17). Also note that a OOF point contributes to array of pixels determined by the spread or blur size. Now we have mechanism in place to analyze rays in 2D and understand OOF. The process of Light field capture is to sample and store the radiance along each ray by sampling in theta and recombining this angular information in software to recover novel views.

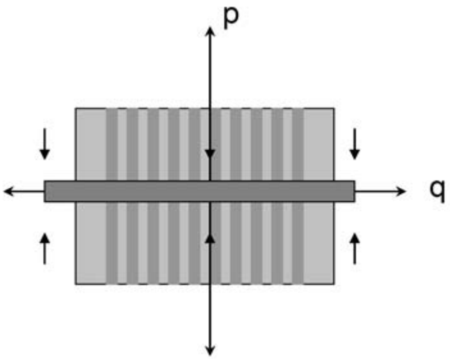

Figure 16: The sensor image is a slice of the LF along the spatial dimension. This slice is obtained by integrating the LF along the angular dimension.

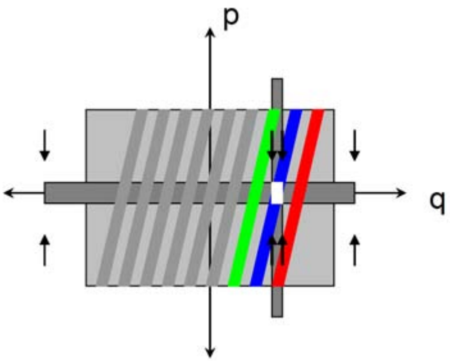

Figure 17: Understanding the effect of blur using higher dimensional LF analysis

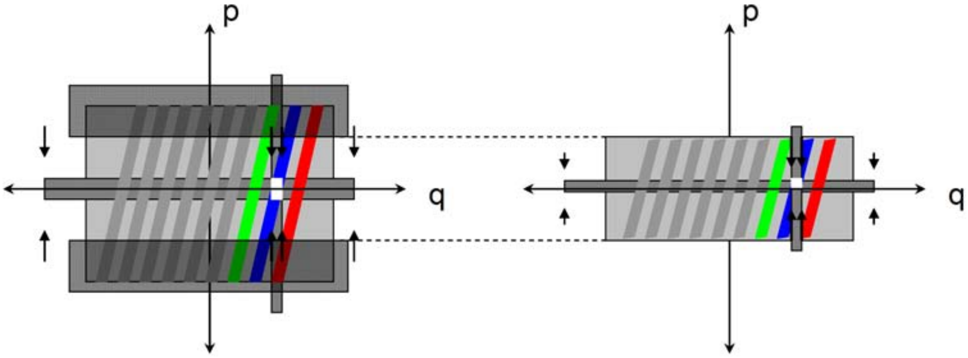

Figure 18: Understanding the reduction in aperture size to reduce size blur (or equivalently increasing the DoF) using higher dimensional LF analysis

Revisiting the 2 questions

We now use our newly learnt Light field analysis to answer the 2 questions we posed at the beginning. Does and OOF image get darker? Based on observation. The answer is NO and this is because we are still capturing the same amount of light using the camera. Although we get a blurred image since each point now contributes to Vertical projections of other points, the net contribution of each object point to intensities in the image remains the same. Hence the image cannot get dark.

Now, Does a zoomed in image get dark? Based on observation answer is YES. The analysis of the answer by thinking in higher dimensions is easy. Reduction of aperture increase the Depth of Field thereby bringing the object in focus. Reduction in aperture effectively reduces the blur size by chopping the radiance (in 2D) by half and in 4D by a quarter (see 18). So by using an aperture size half the size of the original, we increase DOF by roughly twice but we get an image that is half as bright (in 2D) and only one-fourth as bright (in 4D). Thus zoom darkens image. The DoF increases because the blur size decreases and that happens because we sum over fewer values. We get an image half (or one-fourth) as bright, but blur size also roughly halved. In Lanman et al. [8] the center of the lens was blocked by 101 code which is the opposite of aperture size reduction. Now the blur also split into 2 parts and the image also 101 code. We can explain the effect of change in focal length etc. all by thinking in 2D (or 4D).

Light field capture results in reflectance values being stored in lookup tables. Now Digital Refocusing is to project (summing or integration) of angular samples along slanted projection lines. Also all in focus images by summing along different angles and combining the resulting images using a depth map. We could also achieve the effects like imaging a scene the upper half of which is OOF and the lower half is in focus by projections along slanted lines (////) and then vertical (||||) integration. We have limited sensor pixels so we need to sample and rebinning along x and θ (space and angle) resulting in Intermittent columns. That is exactly what lenslets does. When we jump from one microlens center to the adjacent microlens center we are subsampling the space (in x dimension). There are other designs for Ray sampling but the central ideas that all designs distribute rays.

References

[1] Ramesh Raskar *http://web.media.mit.edu/~raskar/*

[2] Ramesh Raskar *http://web.media.mit.edu/~raskar/photo/*

[3] http://en.wikipedia.org/wiki/F-number

[4] http://en.wikipedia.org/wiki/4D_light_field#The_4D_light_field

[5] Veeraraghavan, Raskar, Agrawal, Mohan, and Tumblin, *Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing*, ACM Trans. on Graphics, July 2007

[6] Gortler, S.J., Grzeszczuk, R., Szeliski, R., Cohen, M., *The Lumigraph*, Proc. ACM SIGGRAPH, ACM Press, pp. 43-54. 1996

[7] Levoy, M., Hanrahan, P., *Light Field Rendering*, Proc. ACM SIGGRAPH, ACM Press, pp. 31-42. 1996

[8] Douglas Lanman, Ramesh Raskar, and Gabriel Taubin, *Modeling and Synthesis of Aperture Effects in Cameras* , Int'l Symposium on Computational Aesthetics in Graphics, Visualization, and Imaging. 2008