MAS.963: Computational Camera and Photography **Fall 2009**

Class 6: Cameras for HCI

Prof. Ramesh Raskar October 16, 2009
Scribe: Anonymous MIT student Class 6: Cameras for HCI

Exam:

Light Fields continued

A light field can be represented in 4 dimensions: The sensor with x and y, and the Lens with Φ and Θ. Light fields are a complete representation of the the light rays captured by the lens. Therefore, parameters like focus, zoom and aperture size can be changed after the photo is taken.

Light Field with Pinhole Mask

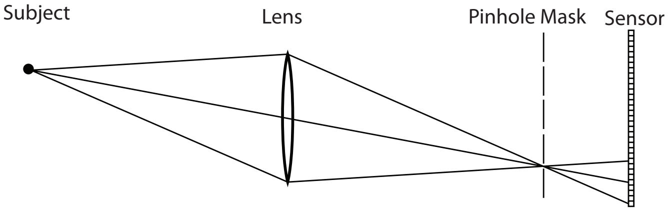

Figure 1: Light Field Camera with Pinhole Mask (not to scale)

Figure 1 shows an example of a light field camera with a pinhole mask in front of the sensor. If we look at the sensor in 1D and assume an total x-resolution of 900 pixels and a Θ resolution of 9, the local x resolution is 100. The disadvantages are therefore a loss in resolution and a loss of light, as most of the light gets blocked by the mask. Therefore, while this model is very clean on a theoretical level, it is inefficient to apply in the real world.

How does the spacing of the mask from the sensor effect the captured light field regions? Figure 2 depicts the effects of the spacing between the sensor and the mask. When the mask is correctly spaced, the blobs on the sensor barely touch each other. When it is spaced too far away, artifacts appear due to overlapping blobs. If the mask is too close, empty space between the blobs results in wasted pixels.

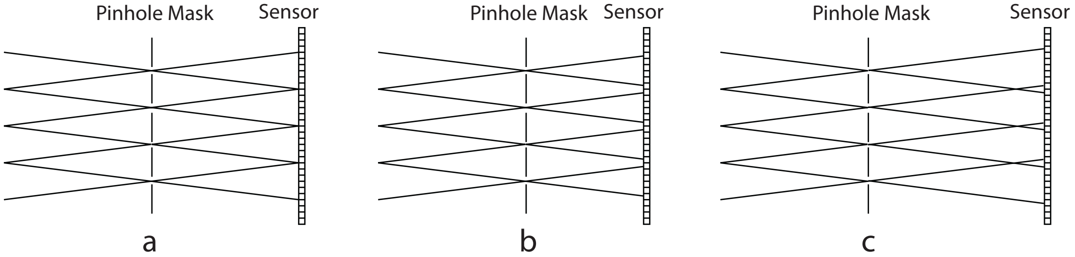

Figure 2: Effect of distance between mask and sensor: a) correct spacing between sensor and mask. b) mask is too close, sensor pixels are wasted. c) mask is too far away, blobs overlap.

f-number The f-number is the ratio of the focal length over the lens diameter. A larger lens diameter results in a smaller f-number. Example: If the lens has a diameter of 25 mm and the focal length is 50 mm, the F-number is 2. If the lens diameter is 12.5 mm, the f-number is 4

When the f-number is decreasing by factor 2, the area is decreasing by factor 4, therefore only 1/4 of the light reaches the sensor. Going down by one f-stop means going down from 2 to 2.8, after that 4.2 (3*square(2)) and 5.6 (4*square(2)) same angle for same f-number: cone of light for each pixel

We want a ratio of focal length to lens diameter, which is equal equal to the mask distance to blob-size. If that is matched, the blobs will barely touch each other, if not, there will be overlap or lost pixels (see 2).

Pinhole Photography Problems Because of the pinhole, very little light reaches the sensor, therefore long exposure. Also, the image is blurred because of diffraction: single point in the world maps to a blurred spot on the sensor. Analogy to water hose: when the size of the opening in the water hose becomes comparable to the size of the water molecule, it will start to spray.

Wavelength of light: 400 to 700 nm (visible light). Green light is 500 nm (.5 micrometers). If we assume a 1mm pinhole = 1000 micrometers, below 500 micrometers becomes too small = diffraction artifacts. This problem is also common with cellphone camera lenses. A 2 mm aperture results in the light spreading out about 20 nm. If the sensor pixel size is 5 microns, the blur is larger than the pixel size.

A light field camera with a pinhole mask has the same problems as a traditional pinhole camera.

0.1 Light Field with Lenslet Array

When using a lenslet array, the setup is similar to a pinhole camera, with an array of lenses instead of pinholes (see Figure 3). There are two special constraints we have to consider: Because it is a light field camera, we want to create a lens which forms an image of a lens on the sensor. This results in very tight tolerances for the distance of the lenslet array to

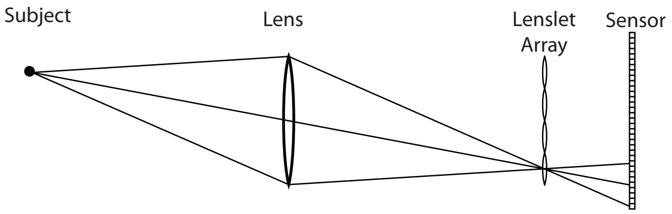

Figure 3: Light Field Camera with lenslet array (not to scale)

the sensor, as a wrong distance will result in a blurred image on the sensor. Due to this constraint, building such a camera is very challenging [10]. Another constraint is matching the main lenses f-number to the lenslet arrays f-numbers to get an optimal directional resolution. If the f-number is too low, the resulting images formed by the lenslet array will overlap, if it is too high, pixels on the sensor will be wasted.

0.2 Light field with aperture mask

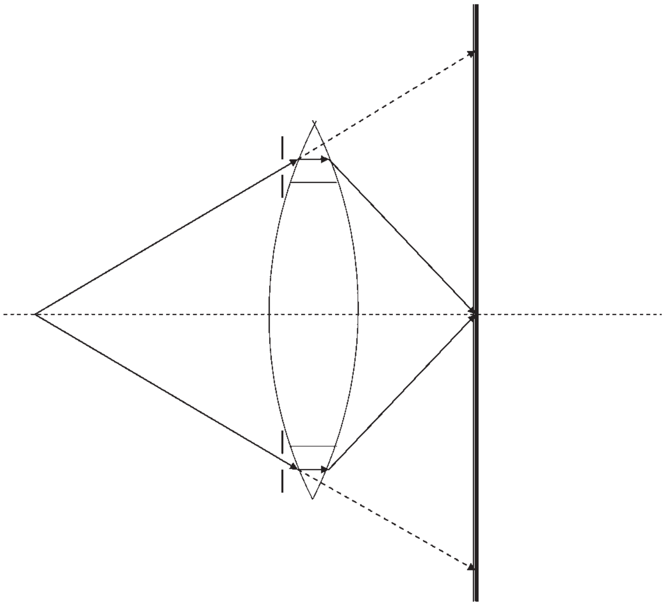

Figure 4: A lens is equivalent to a set of pinholes, each with its own prism bending the light and forcing it to converge at one point behind the camera. (Image courtesy by Ahmed Kirmani)

A lens is can be modeled as an array of pinholes with corresponding prisms (see Figure 4). So if we take a light field image with a camera array, it's the same setup without the prisms. By shifting the image mathematically, we create the set of corresponding prisms. Considering this, we can obtain a light field by taking multiple photos with parts of the lens blocked. [3]

0.3 Light field with mask placed between sensor and lens

When taking a 2d fourier transform of an image, most of the energy is in the low frequencies = center. If we replace the pinhole array in front of the sensor with a high frequency mask, a lot of the energy is in the high frequencies (similar to heterodyning radio signals by mixing a signal with a high frequency carrier). If we then rearrange the individual tiles into 4D planes and compute the inverse Fourier transform, we get the light field. The benefit over a pinhole array is more light hitting the sensor: almost 50% of the light goes through, therefore less diffraction artifacts. Output is almost the same as a lenslet array, after computation.

Aliasing happens in the frequency domain, unless pre-filtering.

0.4 Hadamard multiplexing

Example: several bags with different weights. To determine the weight of the different bags: either weigh every one and therefore get each weight. Also possible: weigh multiple bags at once (useful if scale works well in the mid range, not too low or too high weight). Put a group of them 3 of them: simple linear system w1+w2, w2+w3, w3+w1 take about half of the bags, take measurement, other half, take measurement. (Total of 9 measurements with 4 or 5 bags put together).

Disadvantage: lot of computation

Hadamard and cosine is one and the same thing. Mask can be optimized in many ways between all 0 (blocking) and all 1 (transparent).

Talk by Matt Hirsch about BiDi screen

BiDi screen was inspired by lcd screens with embedded optical sensing and depth sensing cameras. Optical touch sensors provides a one-to-one correspondence between object touching the screen and pixel. However, this correspondence is lost as the object moves away from the scene, resulting in a blur. To bring that correspondence back, the approach is to seperate the sensor by a little margin from the display and then using a mask as described in the previous section. This way, the LCD acts as both a display device and a mask, which allows us to decode the scene in software later on.

An interesting feature of the screen is that it produces an orthographic image.

Spatial heterodyning: like in radio, voice is multiplied with a high-frequency signal, the ray is multiplied with a spatial high-frequency signal through the mask. Due to the offset between mask and sensor, the lightfield created is actually skewed by the time it reaches the sensor. Therefore, we get a lightfield which is modulated in the frequency domain.

Implementation: Since no sensor the size of the screen is currently available, the setup uses a diffusor (like a movie screen) and a camera filming the screen.

To avoid interference between the content displayed on the screen and capturing the light field, the screen constantly switches between these two modes.

The mask which used in the setup is binary and is optimized for light efficiency (50% compared to 1-2% of pinhole camera).

The screen captures a 20x20 light field angularely with 100x80 pixels resolution. From the captured light field, the image is computationally refocused to different planes. The algorithm then finds regions with hight contrast in order to create a depthmap from multiple refocused planes.

The main practical problem in terms of realtime implementation was remapping the different information coming from the sensor to create 4d light field due to memory access, another problem is synchronizing the display hardware with the computer.

Matt first used a printed mask before experimenting with LCD (printed by PageWorks in Cambridge).

1 Camera for HCI: Part 1

Light field camera for video stabilisation [2]: Switch between different cameras in the array acquired from the light field to stabilize the output image of a moving camera.

HoloWall by Matsushita and Rekimoto is well known early research for tracking hands through a diffuse surface [1]. Don't follow this path as many similar projects have been done in the last 12 years.

VideoMouse by Hinckley et al. [9] is a mouse with an embedded video camera instead of the optical sensor. By measuring the distortion of a dot pattern and markers printed on the mousepad, the mouse is able to determine not only x and y position, but also height, tilt and rotation.

FTIR (Frustrated Total Internal Reflection) in itself is not new and has been used in many fields before. However, it was beautifully applied to track multi-touch interaction by Jeff Han [6]. If the refractive index of glass is sufficiently different from air and light hits the glass at sufficiently steep angle, the light will be reflected between the bounds of the glass. Thats exactly how fiber optics and light pipes work as well. For fiber optics, multiple layers with different index of refraction are used to optimize this reflection effect. The FTIR effect occurs when someone touches the glass with a finger. In that case, the total internal reflection is frustrated and the point of contact emits a diffuse glow. This glow is tracked as a blob with a camera viewing the acrylic surface. The company CamFPD builds the Wedge Display based on the TIR principle [5]. This display is based on a projector, which emits an image at certain angles into a plexiglass wedge. The shape of the wedge determines where each light ray is reflected internally and where it exits and is subsequently projected onto a diffusor. The principle of TIR is also used for see-through head-mounted displays.

Mouse 2.0 [4] explores different mouse form factors and novel forms of interacting with them, some based on optical sensing like FTIR and projected light planes. illuminating a plane with a sheet of light and tracking fingers intersecting this sheet is also part of smartboards

The **Wii Remote** [7] has an embedded XGA camera running at 100 Hz, which can track 4 blobs on-board and transmit them wirelessly to the Wii. Johnny Lee exploited this hardware for a number of HCI projects. [8] A similar tracking principle is used by the Vicon motion capturing system, where IR light is emitted onto retroreflective dots, which are then tracked as blobs by IR cameras

By using **multi-flash photography** with colored light sources at the same time, it's possible to track shapes robustly in real-time.

Pens developed by **Anoto** have a built-in IR camera, which they use to analyze an encoded pattern of displaced dots on paper. This way, they are able to determine their unique coordinates on the pattern.

Gaze tracking is still not a completely solved issue. Most solutions use IR light an analyze the reflection of the eyes to determine gaze.

Thermal IR motion detectors only use two thermal pixels. Both sensors have a narrow field-of-view. The motion detection hardware measures the difference between them. This way, an object with a temperature different from it's surrounding is detected when moving from one region to the next. Ambient changes in temperature do not trigger a false detection, as they effect both sensors simultaneously.

References

[1] Nobuyuki Matsushita and Jun Rekimoto, *HoloWall: Designing a Finger, Hand, Body, and Object Sensitive Wall*, Proceedings of UIST '97, 1997

[2] Brandon M. Smith, Li Zhang, Hailin Jin, Aseem Agarwala, *Light Field Video Stabilization*, ICCV 2009

[3] Chia-Kai Liang and Tai-Hsu Lin and Bing-Yi Wong and Chi Liu and Homer Chen, *Programmable Aperture Photography: Multiplexed Light Field Acquisition*, ACM Transactions on Graphics, 2008

[4] Villar, Nicolas and Izadi, Shahram and Rosenfeld, Dan and Benko, Hrvoje and Helmes, John and Westhues, Jonathan and Hodges, Steve and Ofek, Eyal and Butler, Alex and Cao, Xiang and Chen, Billy, *Mouse 2.0: multi-touch meets the mouse*, UIST '09: Proceedings of the 22nd annual ACM symposium on User interface software and technology, 2009

[5] CamFPD Wedge Display, *http://www.eng.cam.ac.uk/news/stories/flatscreen_tv/*

[6] Han, J. Y. *Low-cost multi-touch sensing through frustrated total internal reflection.* In Proceedings of the 18th Annual ACM Symposium on User interface Software and Technology (Seattle, WA, USA, October 23 - 26, 2005). UIST '05

[7] Wii Remote, *http://en.wikipedia.org/wiki/Wii_Remote*

[8] Johnny Lee, *http://johnnylee.net/projects/wii/*

[9] Hinckley, K., Sinclair, M., Hanson, E., Szeliski, R., Conway, M., *The VideoMouse: A Camera-Based Multi-Degree-of-Freedom Input Device*, ACM UIST'99 Symposium on User Interface Software and Technology, 1999

[10] Ng, R., Levoy, M., Brdif, M., Duval, G., Horowitz, M., and Hanrahan, P. *Light field photography with a hand-held plenoptic camera.*, Tech. rep., Stanford University, 2005