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

# Computational Illumination

**Prof. Ramesh Raskar October 2, 2009** **October 2, 2009**
***Scribe: Anonymous MIT student*** **Lecture 4**

## Poll: When will Google Earth go live?

With the increasing camera capacities, as a figment of imagination one can visualize Google Earth going live. The question is, how long will it take the present day computational photography field, to actually implement this as a free community service for a city? What kind of camera would be best suitable for this? How much of computational power will it require?

Some of the arguments that were made during the discussion:

(1) "People would not like to compromise their privacy for this" vs "One can always blur out faces and omit sensitive areas from coverage".
(2) "Google does not have enough camera infrastructures to do this" vs "People would be happy to place a webcam out of their window if Google pays them for it".
(3) "Do we have good enough cameras for satellite images, Google trucks cannot serve live feeds everywhere " vs "Satellite imagery can be used".

Image removed due to copyright restrictions.
See video linked at reference [1].

<fig 1: Recent research from Georgia Tech which shows a sample of Google Earth + Augmented Reality to make Google Earth Live [1]>

As a consensus it was agreed that although this raises more questions than it answers, the day when one can fire up the browser and actually check whether a kid in California is going to school or not is not far from today. With the latest announcement from researchers from Georgia Institute of Technology [1], that talks about a technology to capture real time videos from multiple locations and perspectives to stitch them together, interpolate and animate (to hide identities) the real time movements on a smaller part of the city; having Google Earth live does not seem to be a distant dream. An interesting project from Tokyo city predicted rain conditions based on data accumulated from wiper movements of hundreds of cars in the region. Hereafter we no longer would need this, we can actually record the rain, snow and hurricane histories over time in Google earth world history archives! ***"The next decade, is going to be the decade of visual computing"***

## Computational Illumination

How can we create programmable lighting that minimizes critical human judgment at the time of capture? And provided incredible control over post-capture manipulation for hyper realistic imagery? Computational illumination [2] is by and large illuminating a scene in a coded, controllable fashion-programmed to highlight favorable scene properties or aid in information extraction.

Following parameters of auxiliary photographic lighting are programmable:

(1) Presence or Absence
- – Flash/No-flash

(2) Light position
- – Multi-flash for depth edges
- – Programmable dome (image re-lighting and matting)

(3) Light color /wavelength

(4) Spatial Modulation
- – Synthetic Aperture Illumination

(5) Temporal Modulation
- – TV remote, Motion Tracking, Sony ID-cam, RFIG

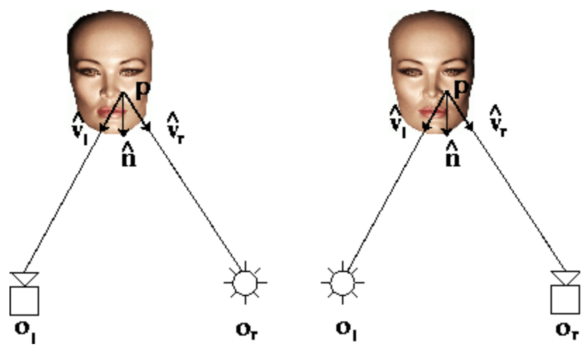
(6) Exploiting (uncontrolled) natural lighting condition
- – Day/Night Fusion

One can also exploit the change in natural lighting.

## Dual Photography

Helmholtz reciprocity: Helmholtz reciprocity says that if a ray from the light source at intensity 'I' reflected from the object reaches the sensor at some intensity 'kI', then a ray following the exact reverse path will experience the same attenuation.

The idea that the flow of light can be effectively reversed without altering its transport properties, can be used cleverly to overcome some scene view-point limitations, or to obtain information about scenes not in view.






<Fig. 2: Capturing a reciprocal pair of images>
Source: Todd Zickler, Harvard http://www.eecs.harvard.edu/~zickler/helmholtz.html

If an imaging sensor and a projector are used in pair , we can project light on the scene pixel by pixel using projector and capture the lighted scene for each case by camera. This kind of a measurement of illumination of pixel xi is analogous to measuring the illumination of pixel xi as created when sensor was at the light source, according to the reciprocity principle. Interesting point to note here is, by accumulating millions of such single pixel illumination images, it is possible to create the view of a scene as would be seen from the projector's point of view. This produces many interesting possibilities like reading out opponents' cards which are not in line of sight while playing poker!

**Note:** The dual photograph however has shadows formed according to the ray directions in which the light placed at the camera position will produce. Hence a region occluded/in shadow in one photograph might be completely illuminated in its dual.

John Baird in early twentieth century perfected the Flying spot Scanner technology which is still used in CRT displays. Similar principles are used in scanning electron microscopes and confocal microscopy for taking 3D pictures of biological specimens.

One catch in the dual photography technique however is that the projector in first place has to point at the scene point of interest which is not in line of sight of the camera. Hence at least the projector has to be in line of sight of the points of interest. This puts a limitation on use of this technique for espionage.
On a side note, it is possible to retrieve the CRT display image on a monitor in a room by simply capturing the glow of light coming out of window if we can capture the images at the rate at which CRT displays scan the screen.

## Relighting using Dual photography

The idea of dual photographs can be used with a camera and a projector instead of a photodetector and a projector. The 4D light field can be obtained for a scene in a (u,v) plane using how it looks from every point in (s,t) plane gives a 4D dataset (u,v,s,t).This is equivalent to turning one pixel from a projector on at a time and capturing how the u,v plane looks from the s-t plane perspective.
Applying the concept of dual photographs here, we can now obtain the image as would be seen from where the projector is. It is notable that this can be done even in cases when the projection source is at a place where camera can possibly be never kept.

*Mathematical illustration:*

Let us assume that our projector has p *q pixels and the camera of m*n pixels.
Let P = pq*1 vector made up of all the pixels of the projector, and
similarly C = mn*1 vector for the camera pixels (mn = 1 for a photodetector).

In primal domain, the image obtained at the camera is given by: a particular vector C' given the illumination P' at the projector by multiplying

$$C' = TP'$$

Here, T: light transform matrix of mn*pq elements.

Let us compute the elements in T. We do this with the flying spot principle, by turning on one pixel at a time in P′,
(i.e., $P' = e_i$, all zeros except one 1 in one row i).

At the camera end, we achieve one column of T (the $i_{th}$ column).

What is the structure of T here? For a completely specular flat surface, each ray will be reflected in one particular direction and thus each column of T will only have one non-zero element.
If we arrange the pixels in P and C, T will be a diagonal matrix.

**Note:** The light transform will be sparse if the scene has little inter-reflections, and dense otherwise (reflections, scattering, translucency, etc.).

One of the most important limitations of this method of dual photography is the amount of time spent in collecting T. If the projector has p*q = 1Mp resolution (this resolution of the projector will decide resolution of the reconstructed dual image), then we will be required to take a million pictures. How can we speed this up? We could use a high speed camera in combination with a CRT to take them really quickly. A better technique would be to use the fact that if no camera pixel sees contribution from two pixels projected at once, we can extract the two corresponding columns of T at once, refer to [3] for details. For a sparse T, compressed sensing can be used to obtain elements from T to overcome the limitation of time spent in collecting T elements.

Once we have matrix T ready with us, we can now proceed to see how to obtain the dual photograph from it. We use simple linear algebra techniques to demonstrate how this happens mathematically in theory. In a dual space, the light source is expected to be at camera and the capture has to be done at projector.

This essentially means we have to find P'' from C''.

$$P'' =?\ C''$$

We need not use the traditional linear algebra methods to solve this by finding inverse if T here (note that it is most likely to be a mega pixel by megapixel dimension matrix). The trick here is to exploit reciprocity and use Transpose of T for obtaining P'' from C''

$T_{ij}$ represents the attenuation of intensity of light coming from projected pixel j when measured at camera pixel i. According to the reciprocity principle, $T_{ji}$ will give the capture of pixel i projected from the camera from pixel j at the projector. Hence,

$$P'' = T^T C''$$

where $T^T$ is the transpose of T.

P'' is the dual photograph in this context.
As can be seen from the horse and emblem objects in the sample scenes from the slides (images from paper), the shadows have now been seen as they would have been if the light source had been kept at initial position of camera.

How does that fact help us?: We can achieve relighting since the light source location change can create nice effects on how the scene looks. For example we can make a specular surface glow from other side and mix the two photographs together to create novel view. An important point to note here is: We can do relighting using single illumination setting (we previously used multiple or changed illumination settings to create relighting effects).

**Relighting Effects**

We upgrade slowly from one photodetector to a full camera with multiple pixels. Each additional photodetector increases our relighting capacity by one more source of light. In case of dual photography, increasing our relighting capacity does not require taking more images. For assignment 1, we were required to take n pictures for n different light sources. To relight the scene with 4D light fields (e.g., projectors surrounding the scene on a plane), we can turn the problem into dual, replace projectors with cameras (and vice versa) and achieve significant savings in the effort required to collect the transform matrix- the trick here is: We can turn on all cameras in parallel but cannot do so for projectors!

## Separation of Direct and Global Illumination

Consider the setup with a light source, and a surface that reflects light. The captured image would generally have radiance from ’first bounce’ but there could also be indirect illumination due to inter reflections, we shall name it ‘second bounce’. Other reasons for indirect light: subsurface scattering (e.g., in the human skin), volumetric scattering (smoke, water), translucent objects, etc. So we have the direct bounce and all indirect bounces. This shows that the images that we capture are made up of a good mixture of effects due to direct bounce and indirect bounces. If we could separate direct bounce effect from indirect bounce effect we would be able to apply this for multitude of interesting problems like how we can see a people standing behind the bush, or differentiate a fake apple form an original one (more about it later).

We can separate the illumination contents as follows: use very high frequency illumination pattern (checkerboard- which can be easily used with its negative: a shifted checkerboard) to illuminate the scene. Suppose in this case the particular patch of interest is illuminated, and we do get the direct bounce back. What’s important to note is that the rest of the scene still contributes to the lighting of the patch. Now, about half of them continue to contribute, while the rest is removed. We can subsequently project the inverse pattern to get the other half of indirect contribution and none of the direct bounce.

In mathematical terms: (for a checkerboard and inverse checkerboard), collected global illumination is half of the total global illumination.

$$I_1 = I_{direct} + I_{global}/2$$
$$I_2 = I_{global}/2$$

So,

$$I_1 - I_2 = I_{direct}$$

The method discussed above still ahs quite a few practical limitations. We assume that we get roughly half of the global component. To achieve this we need to use a very high frequency pattern (changing faster than the radiance characteristics). Still high frequency reflectors (e.g., mirrors, mirror balls) will cause nasty artefacts (in the shape of the pattern we used to illuminate, e.g. checkerboard).

## Using Direct and global illumination for testing fake objects

Interestingly, this technique for separating global and direct illumination can be used for testing the inherent scattering and inter-reflecting properties of objects. An interesting result is that ability to determine the amount of subsurface scattering allows us to identify real fruit from fake fruit. On the other hand, human skin pigment is subsurface, thus the direct light image does not reveal the race of the photographed person.

## Day long photo capture to estimate geographical location

An interesting application of deriving inferences from illumination contents is to see if one can estimate the geographical location of the place from where the photographs have been taken. These photographs when taken over a period of one year can also show interesting facts about the normal duration of the day during different seasons. One can also make use of sky as a mask to estimate the locations from the direct illumination patterns. Since it is easy to know how a particular location on earth is likely to have been lighted up by sun, one can use it in a reverse way to know what was the position from the kind of direct illumination the image shows. Researchers from CMU Computer Vision group have recently come up with interesting results in this area [4].

## Guidelines for Assignment 2

Consider the practical limitations: how many rays needed? How many images will be needed? Exploit reciprocity to reduce the effort! How dark are the dark pixels? Explore the framework with more cameras/projectors/sensors. To emulate a single sensor, use a camera and add all the pixels together.

The problem seems easy: project two checkerboards. In practice not that easy, due to alignment, contrast and focus. The suggested method is to take a bunch of photos with shifted pattern (about 16 in a 4 by 4 window) then take the minimum of photos as global illumination and maximum as global + direct.

## References:


[1] Revolution magazine article on Google earth going live

http://www.revolutionmagazine.com/news/938024/Google-Earth-+-augmented-reality---wow/

[2] Raskar et al, Computational Illumination, ICCGIT 2006 course

http://portal.acm.org/citation.cfm?id=1185739

[3] Sen et al, Dual Photography, SIGGRAPH 2005

http://graphics.stanford.edu/papers/dual_photography/

[4] Jean-François Lalonde, Alexei A. Efros, and Srinivasa G. Narasimhan. Estimating Natural Illumination from a Single Outdoor Image,ICCV 2009

http://graphics.cs.cmu.edu/projects/outdoorIllumination/