# Light and Shading: Core Ideas

D.A. Forsyth, UIUC

1

## Key issues

- Physical
  - what makes a pixel take its brightness values?
- Inference
  - what can we recover from the world using those brightness values?
- Human
  - What can people do?
    - which suggests problems we might be able to solve

Slides to accompany Forsyth and Ponce "Computer Vision - A Modern Approach" 2e by D.A. Forsyth

2

# Model

- Light arrives at a surface
  - from a light source
  - from another surface
- It is reflected into the camera
  - many possible effects
- It arrives at a sensor at the back of the camera
  - and a record is made
  - this could be a linear or a non-linear function of the amount of light



3

# Effects in camera

- Film
  - Record is made by chemical processes in the film
  - These are non-linear
    - Typically,
      - dark patches are lighter than they should be
      - light patches are darker than they should be
      - so that more detail is visible
- CCD
  - Linear devices
    - with non-linearities produced by electronics to mimic film
- Calibration
  - Can be hard to find curves of camera response vs light input
  - Instead, use calibration algorithms (later)



4

## Reflection at a surface

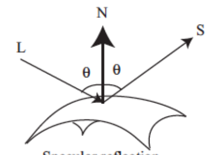
- Many effects when light strikes a surface -- could be:
  - absorbed; transmitted; reflected; scattered
    - eg some people can see arteries, veins under their skin because
      - light is transmitted through skin, reflected at blood vessel, transmitted out
  - Simplify
    - Assume that
      - surfaces don't fluoresce
      - surfaces don't emit light (i.e. are cool)
      - all the light leaving a point is due to that arriving at that point



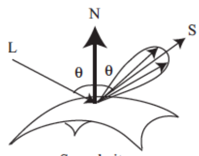
5

## The important reflection modes

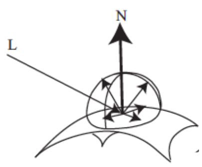
- Specular reflection (mirror like)
  - Pure mirror:
    - incoming, outgoing directions and normal are coplanar
    - incoming, outgoing angles to normal are equal
  - Most specular surfaces:
    - some light leaves the surface along directions near to the specular direction as well
- Diffuse reflection
  - Light leaves in equal amounts in each direction
    - so surface looks equally bright from each viewing direction



Specular reflection



Specularity



Diffuse reflection



6

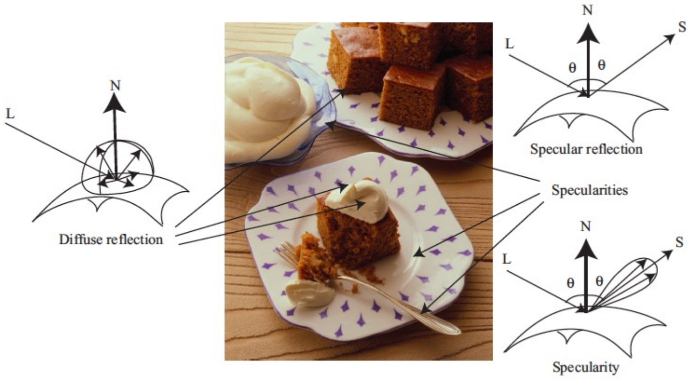


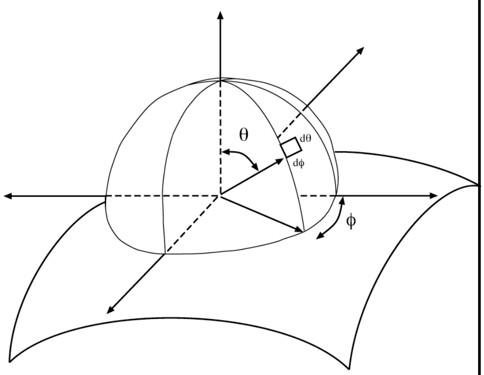

FIGURE 2.1: The two most important reflection modes for computer vision are diffuse reflection (**left**), where incident light is spread evenly over the whole hemisphere of outgoing directions, and specular reflection (**right**), where reflected light is concentrated in a single direction. The specular direction $S$ is coplanar with the normal and the source direction ($L$), and has the same angle to the normal that the source direction does. Most surfaces display both diffuse and specular reflection components. In most cases, the specular component is not precisely mirror like, but is concentrated around a range of directions close to the specular direction (**lower right**). This causes specularities, where one sees a mirror like reflection of the light source. Specularities, when they occur, tend to be small and bright. In the photograph, they appear on the metal spoon and on the plate. Large specularities can appear on flat metal surfaces (arrows). Most curved surfaces (such as the plate) show smaller specularities. Most of the reflection here is diffuse; some cases are indicated by arrows. *Martin Brigdale © Dorling Kindersley, used with permission.*

Slides to accompany Forsyth and Ponce "Computer Vision - A Modern Approach" 2e by D.A. Forsyth

7

# Radiometry

- Questions:
  - how "bright" will surfaces be?
  - what is "brightness"?
    - measuring light
    - interactions between light and surfaces
- Core idea - think about light arriving at a surface
- around any point is a hemisphere of directions
- Simplest problems can be dealt with by reasoning about this hemisphere




Slides to accompany Forsyth and Ponce "Computer Vision - A Modern Approach" 2e by D.A. Forsyth

8

# Foreshortening

- **Principle:** two sources that look the same to a receiver must have the same effect on the receiver.
- **Principle:** two receivers that look the same to a source must receive the same amount of energy.
- "look the same" means produce the same input hemisphere (or output hemisphere)
- **Reason:** what else can a receiver know about a source but what appears on its input hemisphere? (ditto, swapping receiver and source)
- **Crucial consequence:** a big source (resp. receiver), viewed at a glancing angle, must produce (resp. experience) the same effect as a small source (resp. receiver) viewed frontally.

Slides to accompany Forsyth and Ponce "Computer Vision - A Modern Approach" 2e by D.A. Forsyth

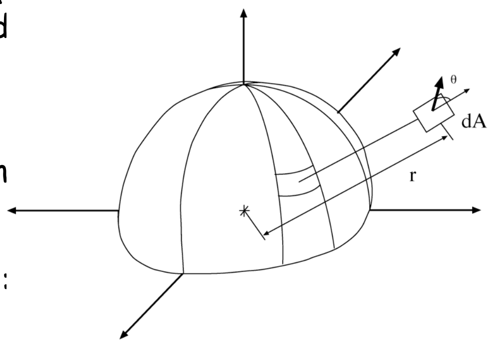
9

# Solid Angle

- By analogy with angle (in radians), the solid angle subtended by a region at a point is the area projected on a unit sphere centered at that point
- The solid angle subtended by a patch area dA is given by

$$d\omega = \frac{dA\cos\vartheta}{r^2}$$

- Another useful expression:




$$d\omega = \sin\vartheta (d\vartheta)(d\phi)$$

Slides to accompany Forsyth and Ponce "Computer Vision - A Modern Approach" 2e by D.A. Forsyth

10

# Measuring Light in Free Space

- **Desirable property:** in a vacuum, the relevant unit does not go down along a straight line.
- How do we get a unit with this property? Think about the power transferred from an infinitesimal source to an infinitesimal receiver.
- We have
**total power leaving s to r = total power arriving at r from s**
- Also:
**Power arriving at r is proportional to:**
  - solid angle subtended by s at r (because if s looked bigger from r, there' d be more)
  - foreshortened area of r (because a bigger r will collect more power



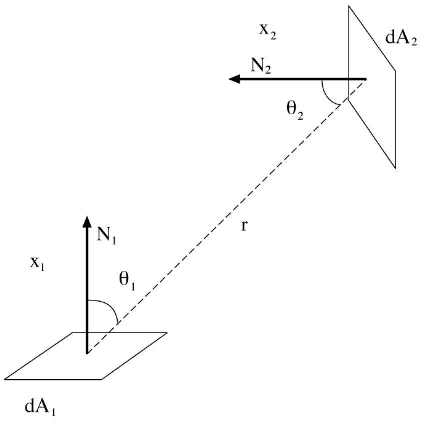
# Radiance

- All this suggests that the light transferred from source to receiver should be measured as:
*Radiant power per unit foreshortened area per unit solid angle*
- This is radiance
- Units: watts per square meter per steradian ($wm^{-2}sr^{-1}$)
- Usually written as:

$$L(\underline{x},\vartheta,\varphi)$$

- Crucial property:
In a vacuum, radiance leaving p in the direction of q is the same as radiance arriving at q from p
  - which was how we got to the unit

# Radiance is constant along straight lines




- Power 1->2, leaving 1:

$$L(\underline{x}_1,\vartheta,\varphi)(dA_1\cos\vartheta_1)\left(\frac{dA_2\cos\vartheta_2}{r^2}\right)$$

- Power 1->2, arriving at 2:

$$L(\underline{x}_2,\vartheta,\varphi)(dA_2\cos\vartheta_2)\left(\frac{dA_1\cos\vartheta_1}{r^2}\right)$$

- But these must be the same, so that the two radiances are equal

Slides to accompany Forsyth and Ponce "Computer Vision - A Modern Approach" 2e by D.A. Forsyth

13

# Irradiance

- How much light is arriving at a surface?
- Sensible unit is *Irradiance*
- Incident power per unit area *not foreshortened*
- This is a function of incoming angle.
- A surface experiencing radiance L(x,q,f) coming in from dw experiences irradiance

$$L(\underline{x},\vartheta,\varphi)\cos\vartheta d\omega$$

- Crucial property: Total power arriving at the surface is given by adding irradiance over all incoming angles --- this is why it's a natural unit
- Total power is

$$\int_{\Omega} L(\underline{x},\vartheta,\varphi)\cos\vartheta\sin\vartheta d\vartheta d\varphi$$

Slides to accompany Forsyth and Ponce "Computer Vision - A Modern Approach" 2e by D.A. Forsyth

14

# Light at surfaces

- Many effects when light strikes a surface -- could be:
  - absorbed
  - transmitted
    - skin
  - reflected
    - mirror
  - scattered
    - milk
  - travel along the surface and leave at some other point
    - sweaty skin
- Assume that
  - surfaces don't fluoresce
    - e.g. scorpions, washing powder
  - surfaces don't emit light (i.e. are cool)
  - all the light leaving a point is due to that arriving at that point



15

# The BRDF

- Assuming that
  - surfaces don't fluoresce
  - surfaces don't emit light (i.e. are cool)
  - all the light leaving a point is due to that arriving at that point
- Can model this situation with the Bidirectional Reflectance Distribution Function (BRDF)
- the ratio of the radiance in the outgoing direction to the incident irradiance

$$\rho_{bd}(\underline{x},\vartheta_o,\varphi_o,\vartheta_i,\varphi_i) = \frac{L_o(\underline{x},\vartheta_o,\varphi_o)}{L_i(\underline{x},\vartheta_i,\varphi_i)\cos\vartheta_i d\omega}$$



16

# Specularities

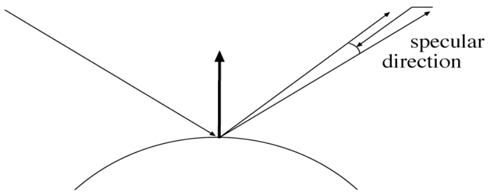
- Mirrors are bright
  - Reflect most incoming light
- Most specular surfaces aren't pure mirrors
  - eg plastics; rough or brushed metal surfaces; lacquers; varnishes
  - The only significant specular reflection is the light source
  - Result: small, bright patches on specular surfaces
    - Specularities
    - Move when the light source moves
    - Move when the viewing direction moves
    - Shape, motion depend on local geometry of the surface
- Specular albedo
  - percentage of incoming light that is specularly reflected



17

# Phong's model

- Exact shape of the specular lobe seldom matters.
- Typical cases:
  - very, very small --- mirror
  - small -- blurry mirror
  - bigger -- see only light sources as "specularities"
  - very big -- faint specularities
- Phong's model
  - reflected energy falls off with $\cos^n(\delta\vartheta)$






18

# Diffuse reflection

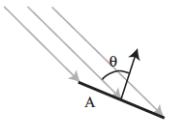
- Light leaves the surface evenly in all directions
  - eg, cotton cloth, carpets, matte paper, matte paints, most “rough” surfaces
- Described by one parameter: Albedo
  - percentage of light arriving that leaves
  - range 0-1
    - practical range is smaller
- Light leaving is (Albedo)x(Light arriving)
  - Ambiguity: A surface could be dark because
    - It reflects a small percentage of the light arriving
    - There isn’t very much light arriving

Slides to accompany Forsyth and Ponce “Computer Vision - A Modern Approach” 2e by D.A. Forsyth

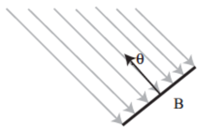
19

# How much light arrives?

- Assume source is far away
  - So light travels in parallel rays
  - (Light arriving) proportional to (number of rays striking surface)
  - Surface A below receives less light than surface B
- Drawing yields
  - (number of rays striking surface) proportional to cos theta
    - where theta is angle between normal and direction of travel
- Shadows
  - If point can’t see the light source, it is in shadow



Diffuse reflection,
dark



Diffuse reflection,
bright

Slides to accompany Forsyth and Ponce “Computer Vision - A Modern Approach” 2e by D.A. Forsyth

20

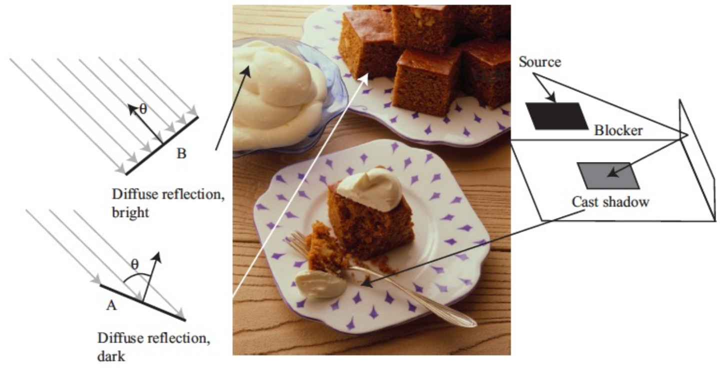



FIGURE 2.2: The orientation of a surface patch with respect to the light affects how much light the patch gathers. We model surface patches as illuminated by a distant point source, whose rays are shown as light arrowheads. Patch A is tilted away from the source ($\theta$ is close to $90^0$) and collects less energy, because it cuts fewer light rays per unit surface area. Patch B, facing the source ($\theta$ is close to $0^0$), collects more energy, and so is brighter. Shadows occur when a patch cannot see a source. The shadows are not dead black, because the surface can see interreflected light from other surfaces. These effects are shown in the photograph. The darker surfaces are turned away from the illumination direction. *Martin Brigdale © Dorling Kindersley, used with permission.*

Slides to accompany Forsyth and Ponce "Computer Vision - A Modern Approach" 2e by D.A. Forsyth

21

# Diffuse+Specular model

- Most surfaces can be modeled as diffuse+specular
  - surface parameters:
    - diffuse albedo,
    - specular albedo,
    - phong parameter

Seldom known, hard to measure, usually not important

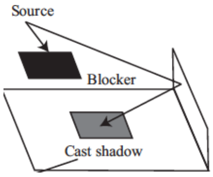
- This justifies the following strategy for many analyses
  - Find and remove specularities
    - which are small, and bright
  - Ignore the missing points, and treat the rest of the image as diffuse

Slides to accompany Forsyth and Ponce "Computer Vision - A Modern Approach" 2e by D.A. Forsyth

22

# Shadows

- Most shadows aren't dark
  - because shadow points get light from other surfaces, not just light source
- Area sources
  - Large, bright areas
  - eg diffuser boxes, the sky
  - Yield smooth, blurry shadows
    - Points that can see the whole source are brighter
    - Points that can see only part of the source are darker (penumbra)
    - Points that can see no part of the source are darkest (umbra)




- Other surfaces behave like area sources
  - Smooth, blurry shadows are common (and sometimes too faint to see)

Slides to accompany Forsyth and Ponce "Computer Vision - A Modern Approach" 2e by D.A. Forsyth

23

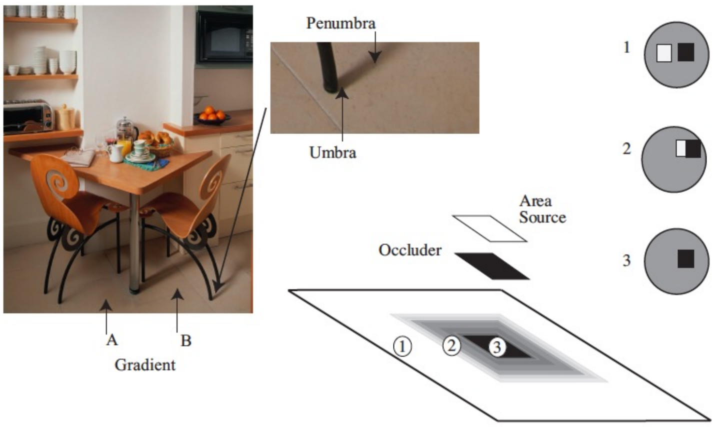



FIGURE 2.3: Area sources generate complex shadows with smooth boundaries, because from the point of view of a surface patch, the source disappears slowly behind the occluder. **Left:** a photograph, showing characteristic area source shadow effects. Notice that A is much darker than B; there must be some shadowing effect here, but there is no clear shadow boundary. Instead, there is a fairly smooth gradient. The chair leg casts a complex shadow, with two distinct regions. There is a core of darkness (the *umbra*—where the source cannot be seen at all) surrounded by a partial shadow (*penumbra*—where the source can be seen partially). A good model of the geometry, illustrated **right**, is to imagine lying with your back to the surface looking at the world above. At point 1, you can see all of the source; at point 2, you can see some of it; and at point 3, you can see none of it. *Peter Anderson © Dorling Kindersley, used with permission.*

Slides to accompany Forsyth and Ponce "Computer Vision - A Modern Approach" 2e by D.A. Forsyth

24

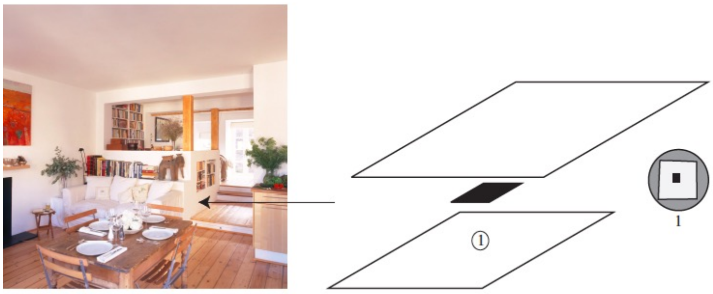



FIGURE 2.4: The photograph on the **left** shows a room interior. Notice the lighting has some directional component (the vertical face indicated by the arrow is dark, because it does not face the main direction of lighting), but there are few visible shadows (for example, the chairs do not cast a shadow on the floor). On the **right**, a drawing to show why; here there is a small occluder and a large area source. The occluder is some way away from the shaded surface. Generally, at points on the shaded surface the incoming hemisphere looks like that at point 1. The occluder blocks out some small percentage of the area source, but the amount of light lost is too small to notice (compare figure 2.3). *Jake Fitzjones © Dorling Kindersley, used with permission.*





# Light and shading - Crucial points

- Image brightness is affected by
  - amount of light arriving at surface
  - surface type (diffuse, specular) and amount reflected at surface
  - camera sensitivity
- There are significant ambiguities
  - eg low albedo surface in bright light
    - vs high albedo surface in low light
  - each might reflect about the same amount
- Most surfaces can be modeled as diffuse + specular
  - generally, find and remove specularities
  - treat the rest as diffuse

# Shadows are everywhere!

27

# Shadow detection is challenging

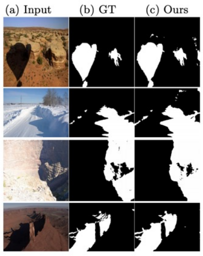
Appear anywhere!
Different shapes!
Varied appearance!

28

## Deep Learning Shadow Detection and Removal with Physical Illumination Constraints

Shadow detection

(a) Input (b) GT (c) Ours



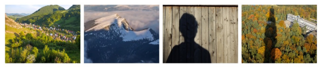
Shadow Removal

Input ST-CGAN CVPR18 Ours Ground-Truth

Shadow Editing

Synthetic image Real image Synthetic image

Shadow Video Dataset



**A+D Net: Training a Shadow Detector with Adversarial Shadow Attenuation**, Hieu Le, Tomas Vicente, Vu Nguyen, Minh Hoai, Dimitris Samaras, ECCV18
**Shadow Removal via Shadow Image Decomposition**, Hieu Le , Dimitris Samaras, ICCV 19
**From Shadow Segmentation to Shadow Removal**, Hieu Le , Dimitris Samaras, ECCV 20

29

## Deep Learning Shadow Detection and Removal

# From Shadow Segmentation to Shadow Removal

Hieu Le, Dimitris Samaras

Stony Brook University

Stony Brook University

ECCV'20 ONLINE 23-28 AUGUST 2020

ECCV18
**Shadow Removal via Shadow Image Decomposition**, Hieu Le , Dimitris Samaras, ICCV 19
**From Shadow Segmentation to Shadow Removal**, Hieu Le , Dimitris Samaras, ECCV 20

30

# Light and Shading: Inference from Shading

- D.A. Forsyth, UIUC

Slides to accompany Forsyth and Ponce "Computer Vision - A Modern Approach" 2e by D.A. Forsyth

41

# Inference from shading

- Recover some information about the world from shading
- Cases
  - photometric stereo
    - recover shape and albedo of surfaces from multiple shaded images
  - recovering surface albedo
    - from image data
  - determining radiometric calibration
    - how much light is required to produce a particular number in the image

Slides to accompany Forsyth and Ponce "Computer Vision - A Modern Approach" 2e by D.A. Forsyth

42

# Photometric stereo

- Assume:
  - a set of point sources that are infinitely distant
  - a set of pictures of an object, obtained in exactly the same camera/object configuration but using different sources
  - pictures in an orthographic camera
  - A Lambertian object (or the specular component has been identified and removed)



43

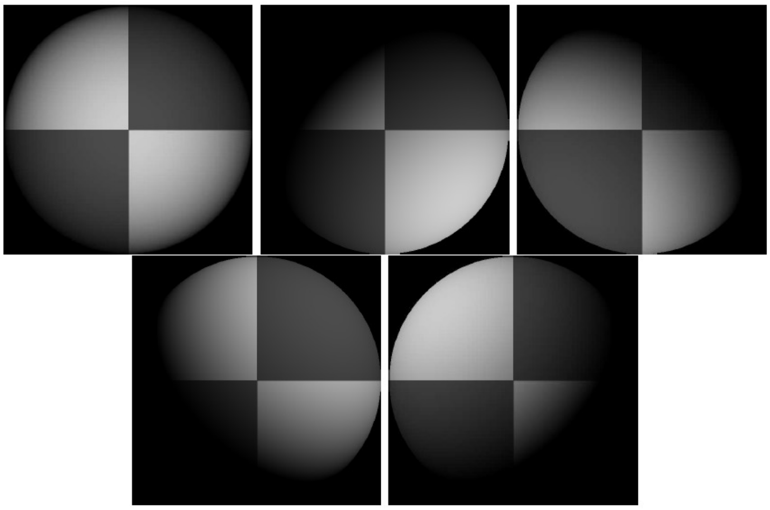


FIGURE 2.11: Five synthetic images of a sphere, all obtained in an orthographic view from the same viewing position. These images are shaded using a local shading model and a distant point source. This is a convex object, so the only view where there is no visible shadow occurs when the source direction is parallel to the viewing direction. The variations in brightness occuring under different sources code the shape of the surface.

Slides to accompany Forsyth and Ponce "Computer Vision - A Modern Approach" 2e by D.A. Forsyth

44

# Imaging model

- Write B for radiosity, S_1 for source vector
- Image value at (x, y) is

$$\begin{aligned} I(x,y) &= kB(\boldsymbol{x}) \\ &= kB(x,y) \\ &= k\rho(x,y)\boldsymbol{N}(x,y)\cdot \boldsymbol{S}_1 \\ &= \boldsymbol{g}(x,y)\cdot \boldsymbol{V}_1, \end{aligned}$$

where $\boldsymbol{g}(x,y) = \rho(x,y)\boldsymbol{N}(x,y)$ and $\boldsymbol{V}_1 = k\boldsymbol{S}_1$, where $k$ is the constant connecting the camera response to the input radiance.

- Notice g(x, y) tells us about the surface, V_1 about the source and camera

Slides to accompany Forsyth and Ponce "Computer Vision - A Modern Approach" 2e by D.A. Forsyth

45

# Measurement setup

- Take multiple photographs, using different sources
  - We can determine V's for these sources by
    - first principles
    - or imaging an object of known geometry
- Stack the V's into a matrix

$$\mathcal{V} = \begin{pmatrix} \boldsymbol{V}_1^T \\ \boldsymbol{V}_2^T \\ \dots \\ \boldsymbol{V}_n^T \end{pmatrix}$$

- Stack the image values at each point into a vector

$$\boldsymbol{i}(x,y) = \{I_1(x,y), I_2(x,y), \dots, I_n(x,y)\}^T$$

- and get the linear system

$$\boldsymbol{i}(x,y) = \mathcal{V}\boldsymbol{g}(x,y)$$

Slides to accompany Forsyth and Ponce "Computer Vision - A Modern Approach" 2e by D.A. Forsyth

46

## Measurement setup - II

- Solve the linear system for g(x, y)
- Reconstruct
  - Albedo from

    $$|\boldsymbol{g}(x,y)| = \rho(x,y)$$

    - this gives a check on the measurements, because albedo < 1
  - Surface from N(x, y), the unit normal, given by $\boldsymbol{N}(x,y) = \frac{\boldsymbol{g}(x,y)}{|\boldsymbol{g}(x,y)|}$
    - this gives a check on the measurements, because N must be integrable

Slides to accompany Forsyth and Ponce "Computer Vision - A Modern Approach" 2e by D.A. Forsyth

47

## Normals

- Not every vector field is a surface normal field
- For a surface $(x, y, f(x,y))$ the normal is

$$\boldsymbol{N}(x,y) = \frac{1}{\sqrt{1 + \frac{\partial f}{\partial x}^2 + \frac{\partial f}{\partial y}^2}} \left\{ \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, 1 \right\}^T$$

Slides to accompany Forsyth and Ponce "Computer Vision - A Modern Approach" 2e by D.A. Forsyth

48

# Integrability

Assume that the measured value of the unit normal at some point $(x, y)$ is $(a(x,y), b(x,y), c(x,y))$. Then

$$\frac{\partial f}{\partial x} = \frac{a(x,y)}{c(x,y)} \text{ and } \frac{\partial f}{\partial y} = \frac{b(x,y)}{c(x,y)}.$$

We have another check on our data set, because

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x},$$

so we expect that

$$\frac{\partial \left(\frac{a(x,y)}{c(x,y)}\right)}{\partial y} - \frac{\partial \left(\frac{b(x,y)}{c(x,y)}\right)}{\partial x}$$

should be small at each point.



49

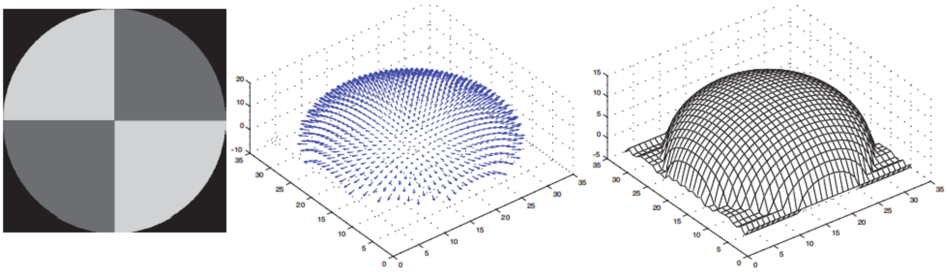


FIGURE 2.12: The image on the **left** shows the magnitude of the vector field $\boldsymbol{g}(x,y)$ recovered from the input data of Figure 2.11 represented as an image—this is the reflectance of the surface. The **center** figure shows the normal field, and the **right** figure shows the height field.



50

FIGURE 2.13: Photometric stereo could become the method of choice to capture complex deformable surfaces. On the **top**, three images of a garment, lit from different directions, which produce the reconstruction shown on the **top right**. A natural way to obtain three different images at the same time is to use a color camera; if one has a red light, a green light, and a blue light, then a single color image frame can be treated as three images under three separate lights. On the **bottom**, an image of the garment captured in this way, which results in the photometric stereo reconstruction on the **bottom right**. *This figure was originally published as Figure 6 of "Video Normals from Colored Lights," G. J. Brostow, C. Hernández, G. Vogiatzis, B. Stenger, and R. Cipolla, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011 © IEEE, 2011.*

Slides to accompany Forsyth and Ponce "Computer Vision - A Modern Approach" 2e by D.A. Forsyth

51

# Estimating surface albedo

- Albedo tends to appear in constant patches
  - separated by edges
- Illumination tends to produce smooth gradients
  - but not always - for example, shadow edges outdoors
- Strategy
  - Take image log, so that multiplication becomes addition
  - Differentiate
  - Now small derivatives are due to illumination, large are due to albedo
  - Integrate, estimate missing constant
  - Known as Retinex
  - This strategy remains reliable, near as good as best known

Slides to accompany Forsyth and Ponce "Computer Vision - A Modern Approach" 2e by D.A. Forsyth

52

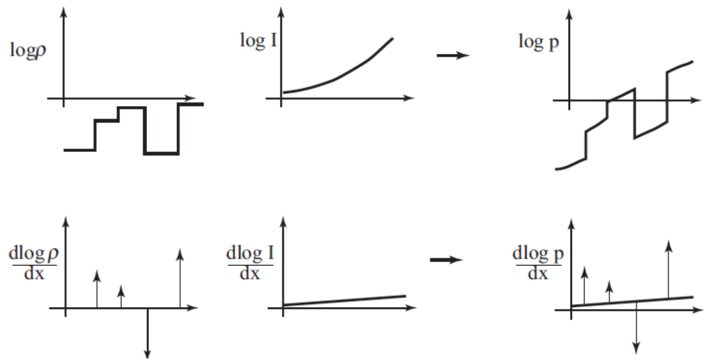



FIGURE 2.8: The lightness algorithm is easiest to illustrate for a 1D image. In the **top row**, the graph on the **left** shows $\log \rho(x)$, that in the **center** $\log I(x)$, and that on the **right** their sum, which is $\log C$. The log of image intensity has large derivatives at changes in surface reflectance and small derivatives when the only change is due to illumination gradients. Lightness is recovered by differentiating the log intensity, thresholding to dispose of small derivatives, and integrating at the cost of a missing constant of integration.

Slides to accompany Forsyth and Ponce "Computer Vision - A Modern Approach" 2e by D.A. Forsyth

53

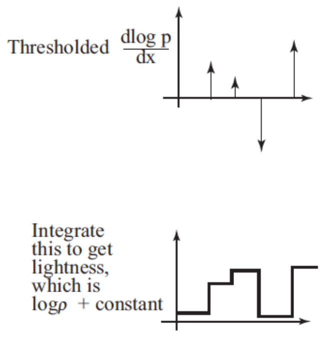



FIGURE 2.8: The lightness algorithm is easiest to illustrate for a 1D image. In the **top row**, the graph on the **left** shows $\log \rho(x)$, that in the **center** $\log I(x)$, and that on the **right** their sum, which is $\log C$. The log of image intensity has large derivatives at changes in surface reflectance and small derivatives when the only change is due to illumination gradients. Lightness is recovered by differentiating the log intensity, thresholding to dispose of small derivatives, and integrating at the cost of a missing constant of integration.

Slides to accompany Forsyth and Ponce "Computer Vision - A Modern Approach" 2e by D.A. Forsyth

54

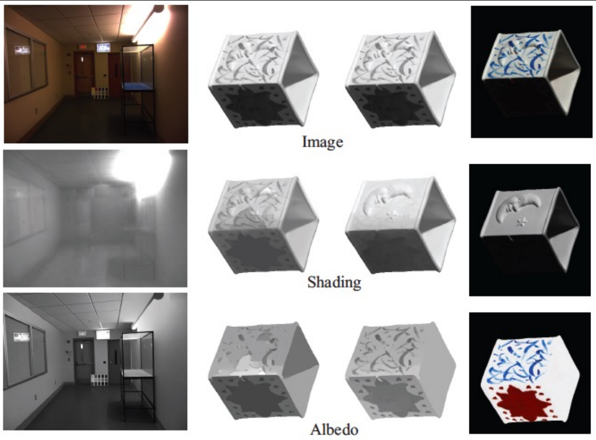



FIGURE 2.9: Retinex remains a strong algorithm for recovering albedo from images. Here we show results from the version of Retinex described in the text applied to an image of a room (**left**) and an image from a collection of test images due to Grosse *et al.* (2009). The **center-left** column shows results from Retinex for this image, and the **center-right** column shows results from a variant of the algorithm that uses color reasoning to improve the classification of edges into albedo versus shading. Finally, the **right** column shows the correct answer, known by clever experimental methods used when taking the pictures. This problem is very hard; you can see that the albedo images still contain some illumination signal. *Part of this figure courtesy Kevin Karsch, U. Illinois. Part of this figure was originally published as Figure 3 of "Ground truth dataset and baseline evaluations for intrinsic image algorithms," by R. Grosse, M. Johnson, E. Adelson, and W. Freeman, Proc. IEEE ICCV 2009, © IEEE, 2009.*





# Radiometric calibration - I

- What image output is produced by a given input radiance?
- Why?
  - To interpret image intensities more accurately
  - To produce high-dynamic range images
    - Take several exposures with different shutter speeds
      - using a radiometrically calibrated camera
      - multiple exposures because some pixels saturated/too dark in some
    - Now use calibration to infer radiance at each pixel from exposures

# Radiometric calibration - II

- What image output is produced by a given input radiance?
- Strategies
  - ask camera manufacturer (sometimes works)
  - look it up on the web (can work quite often)
  - calibrate
- Calibration relies on reciprocity
  - camera response is governed by energy
  - image patch with intensity E for time t - energy is E t
  - image patch with intensity k E for time t/k - energy is E t

Each gives the same camera response



57

# Radiometric calibration - III

- Strategy
  - Image the same patches for different, known, periods of time
    - by manipulating shutter speeds
  - Record input/output pairs for i, j'th pixel
    - input is t E_ij
    - output is f(t E_ij)
  - data looks like

$$\{(t_1 E_{ij}, f(t_1 E_{ij})), (t_2 E_{ij}, f(t_2 E_{ij})), \ldots, (t_n E_{ij}, f(t_n E_{ij}))\}$$

    - where E_ij and f are unknown, but t's and values f(t E_ij) are known
  - fit a model of f



58

# Fitting a model

- Numerous strategies
  - assume f is polynomial, infer parameters and missing E values
  - assume f is invertible and very smooth, fit to g=Inverse(f)
- Fitting g
  - take logs to get

$$\log g(I_{ij}^{(k)}) = \log E_{ij} + \log \Delta t_k$$

  - Fit g by minimizing

$$\sum_{i,j,k} (\log g(I_{ij}^{(k)}) - (\log E_{ij} + \log \Delta t_k))^2 + \text{smoothness penalty on } g$$

Slides to accompany Forsyth and Ponce "Computer Vision - A Modern Approach" 2e by D.A. Forsyth

59

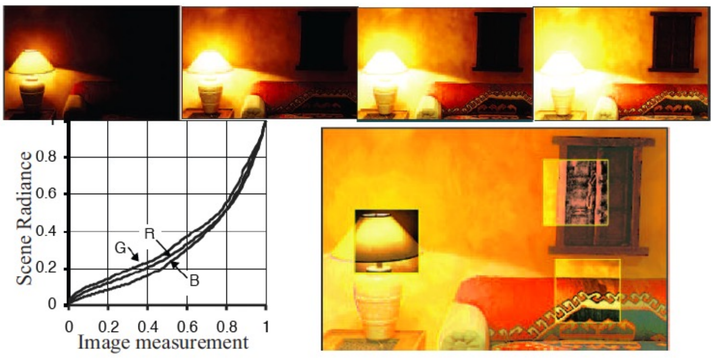



FIGURE 2.5: It is possible to calibrate the radiometric response of a camera from multiple images obtained at different exposures. The **top** row shows four different exposures of the same scene, ranging from darker (shorter shutter time) to lighter (longer shutter time). Note how, in the dark frames, the lighter part of the image shows detail, and in the light frames, the darker part of the image shows detail; this is the result of non-linearities in the camera response. On the **bottom left**, we show the inferred calibration curves for each of the R, G, and B camera channels. On the **bottom right**, a composite image illustrates the results. The dynamic range of this image is far too large to print; instead, the main image is normalized to the print range. Overlaid on this image are boxes where the radiances in the box have also been normalized to the print range; these show how much information is packed into the high dynamic range image. *This figure was originally published as Figure 7 of "Radiometric Self Calibration," by T. Mitsunaga and S. Nayar, Proc. IEEE CVPR 1999, © IEEE, 1999.*

Slides to accompany Forsyth and Ponce "Computer Vision - A Modern Approach" 2e by D.A. Forsyth

60

# Case Study: Neural Face Editing

Neural Face Editing with Intrinsic Image Disentangling, Zhixin Shu, Ersin Yumer, Sunil Hadap, Kalyan Sunkavalli, Eli Shechtman, and Dimitris Samaras, CVPR 2017.

61

# Motivation

- Physically based face editing


Expression editing | Illumination editing | Pose editing

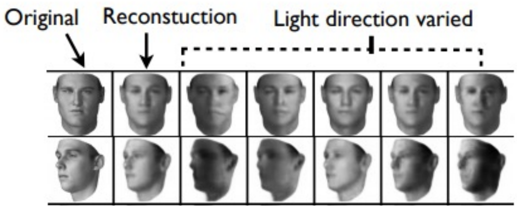
[Blanz and Vetter 1999]
[Wen et al. 2003]
[Weyrich et al. 2005]
[Wang et al. 2009]

[Yang et al. 2011]
[Kemelmacher-Shlizerman et al. 2014]
[Cao et al. 2014]
[Thies et al. 2015]
...

62

# Motivation

- Neural network based face editing




[Gardner et al. 2015]
[Kulkarni et al. 2015]
[Brock et al. 2016]
[Chen et al. 2016]

[Yan et al. 2016]
[Perarnau et al. 2016]
[Upchurch et al. 2017]
[Tran et al. 2017]
...

63

# Motivation

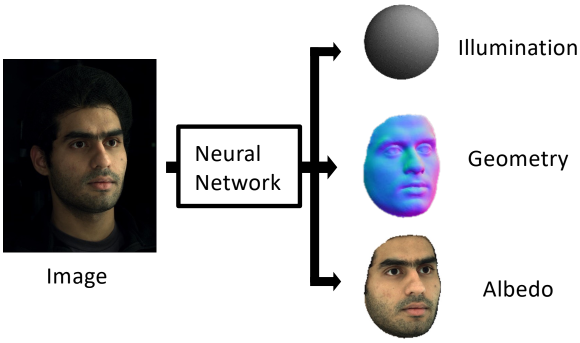
Physically based face editing

Neural network based face editing

Our goal:
Physically-grounded-neural-network-based face editing

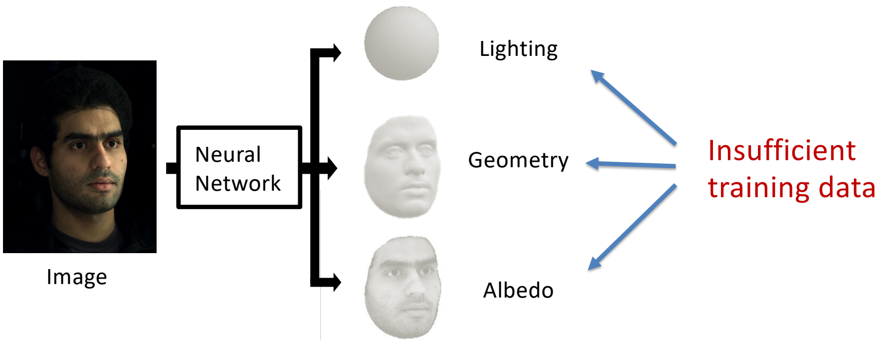
64

# Motivation



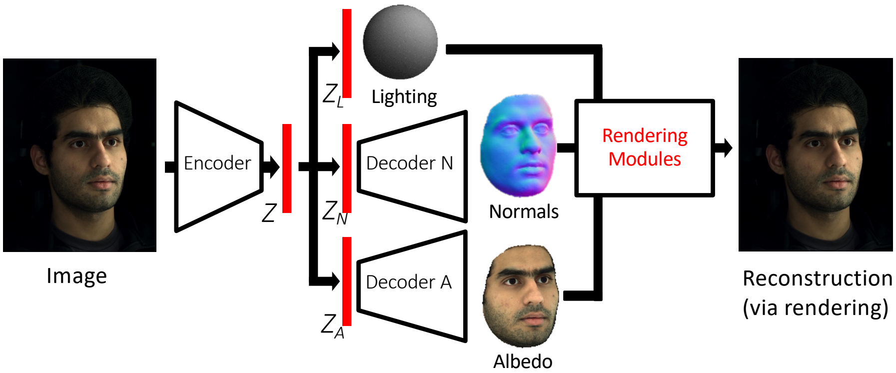

65

# Challenge



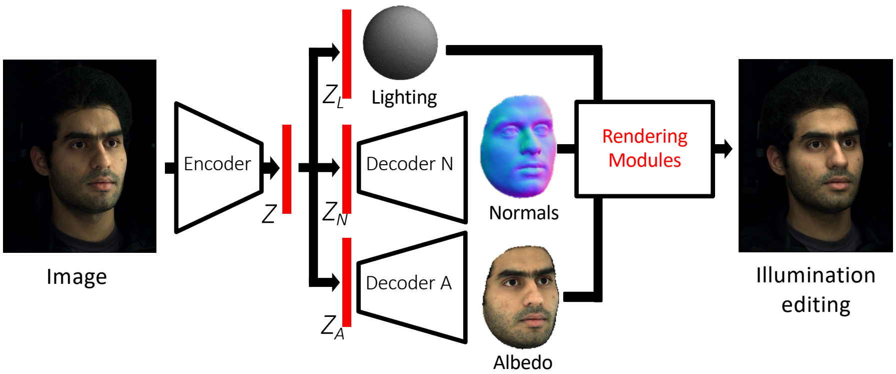

66

# Introduction



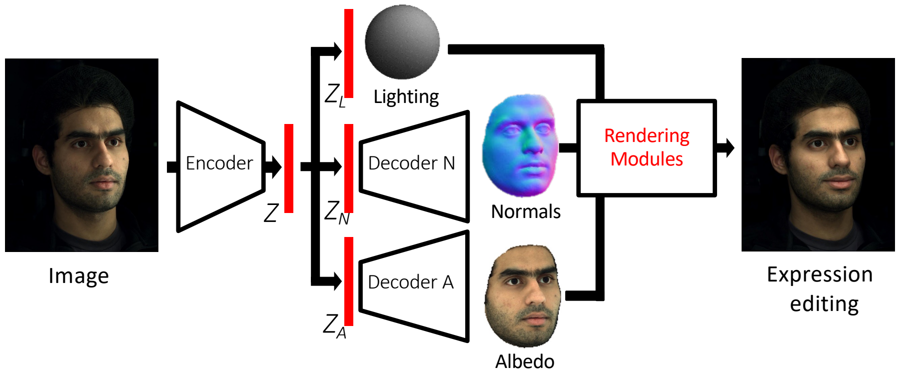

67

# Introduction



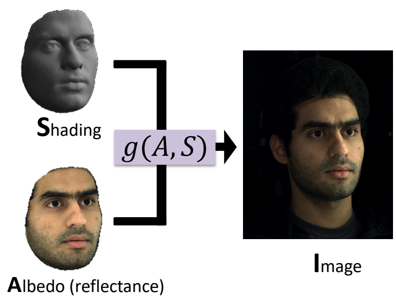

68

# Introduction



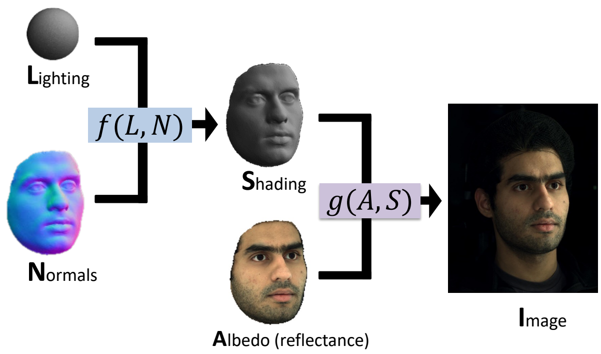

69

# Background: face image formation




70

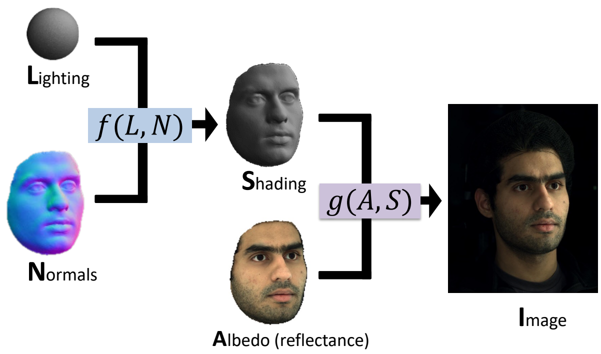
## Background: face image formation




71

## Background: face image formation

$g = A \circ S$

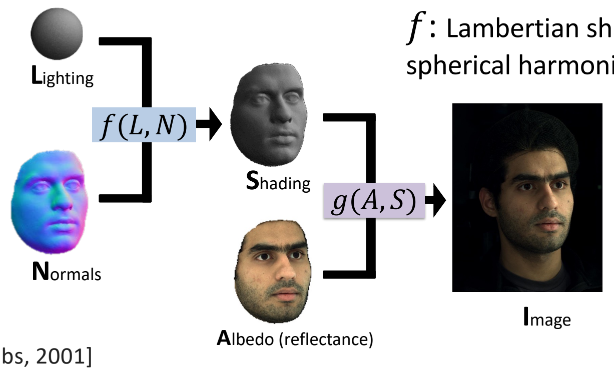



72

## Background: face image formation

$g = A \circ S$

$f$: Lambertian shading with spherical harmonics lighting



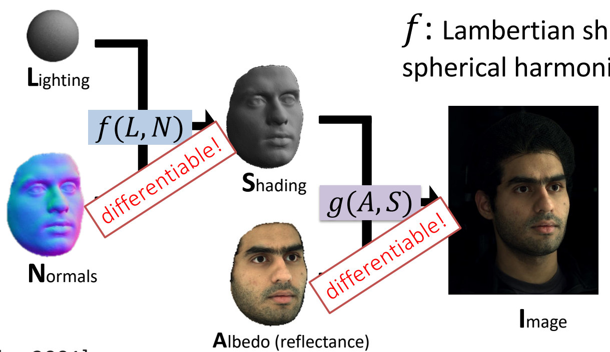

[Basri and Jacobs, 2001]
[Ramamoorthi and Hanrahan, 2001]

73

## Background: face image formation

$g = A \circ S$

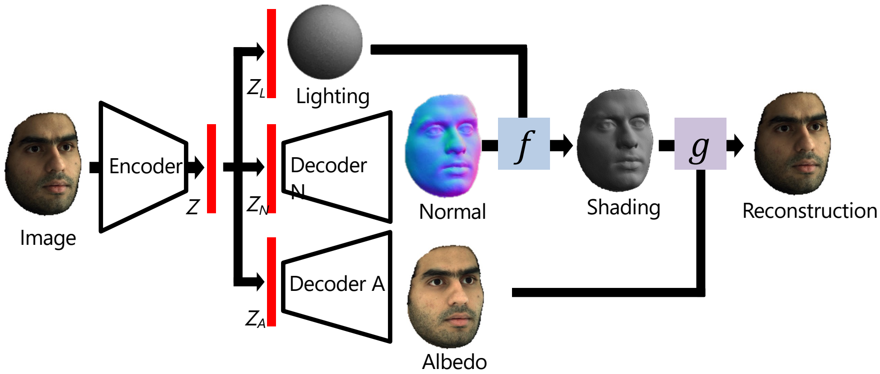
$f$: Lambertian shading with spherical harmonics lighting




[Basri and Jacobs, 2001]
[Ramamoorthi and Hanrahan, 2001]

74

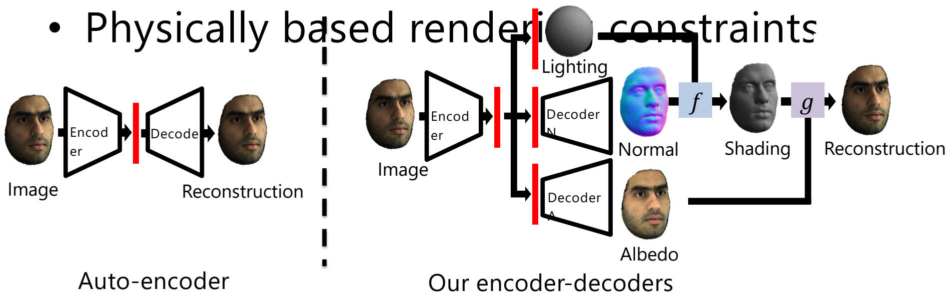
## Network with rendering based reconstruction




75

## Comparison to a standard auto-encoder

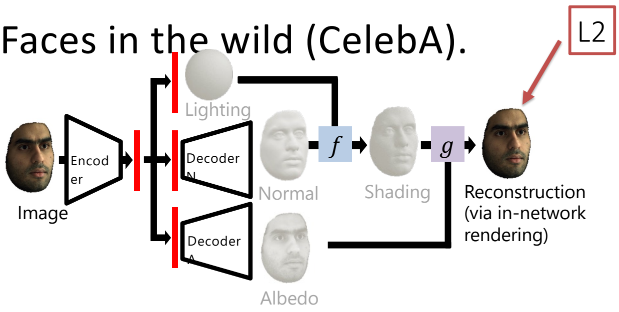
- Explicit access to lighting, normals and albedo
- Physically based rendering constraints




76

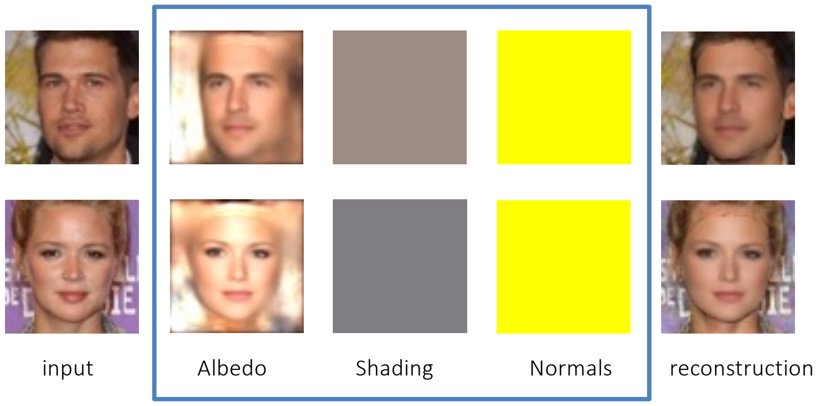
# Unsupervised learning

- Unsupervised training with reconstruction loss (L2).
- Data: Faces in the wild (CelebA).



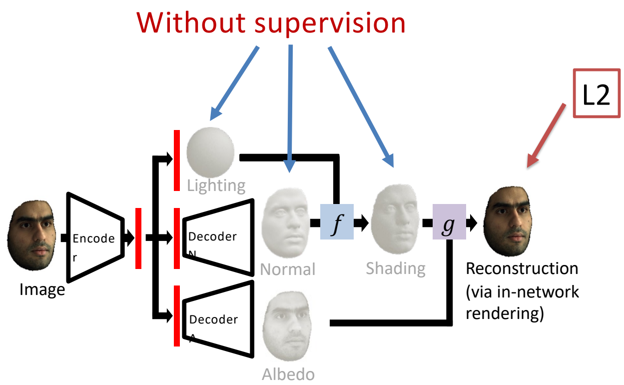

77

# Unsupervised learning



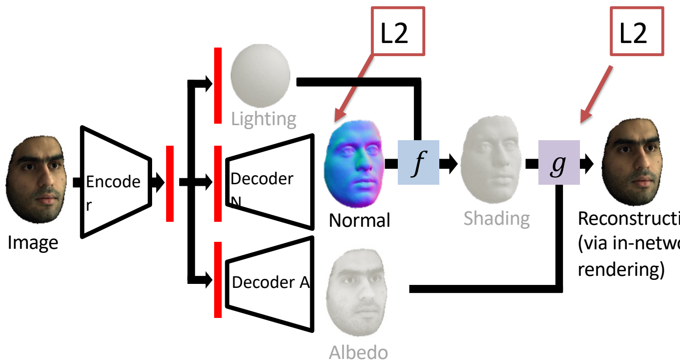

78

# We need some supervision!




79

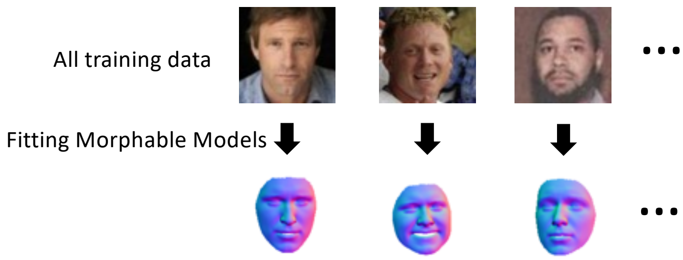
# Weak supervision: guiding loss




80

# Weak supervision: guiding loss

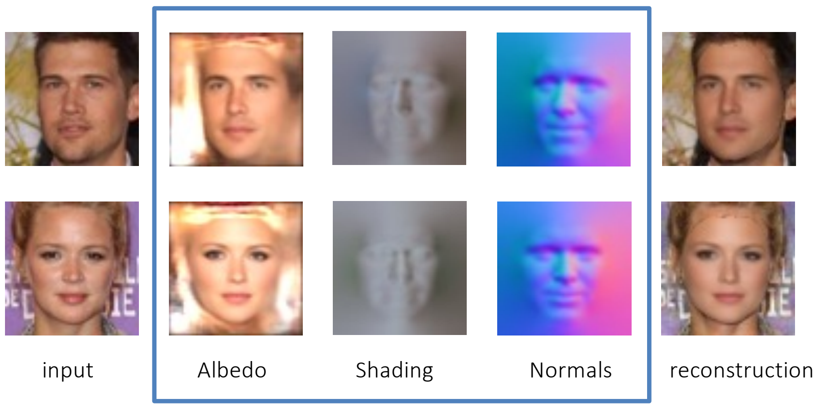
- Guiding loss for Normals: a proxy from Morphable Models




[Yang et al., 2012]

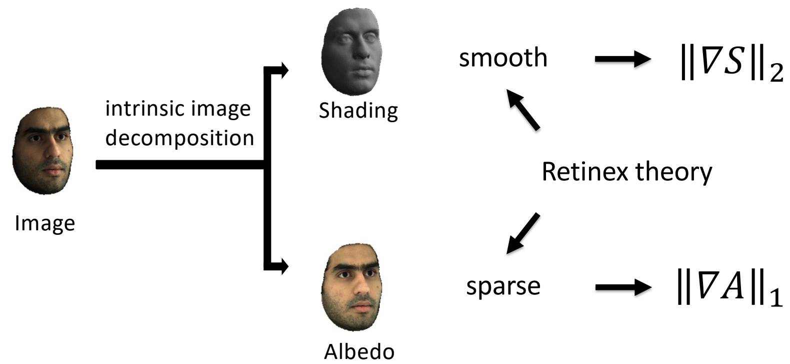
81

# Weak supervision: Normals




82

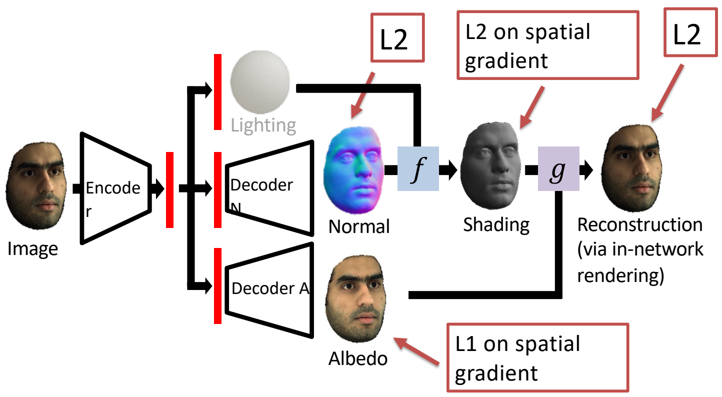
# Weak supervision: intrinsic images




[Barrow and Tenenbaum 1978]
[Land et al. 1971]

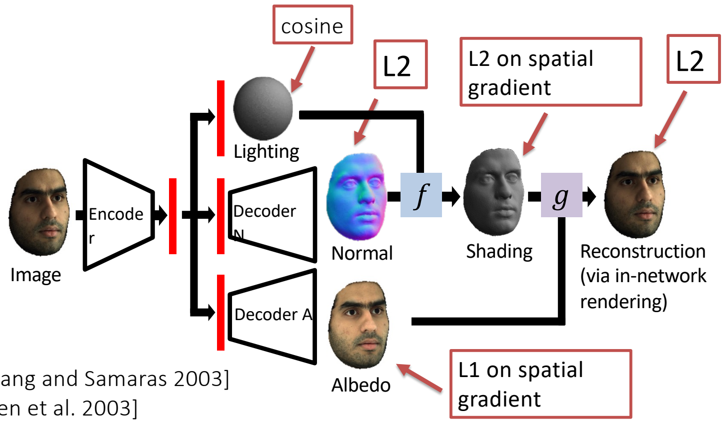
83

# Weak supervision: intrinsic images




84

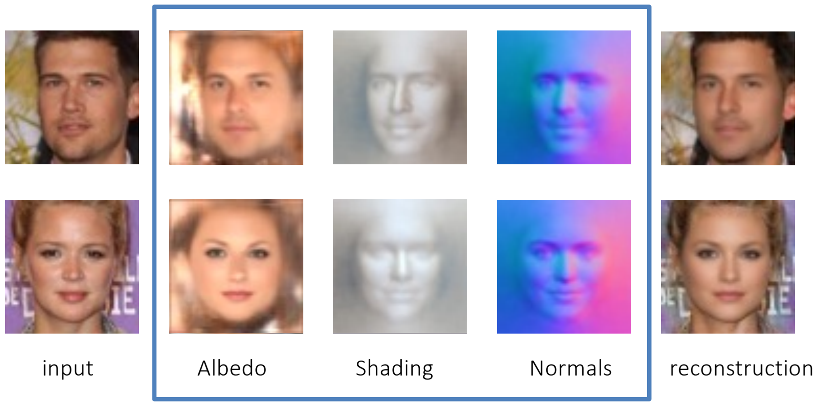
# Weak supervision




[Zhang and Samaras 2003]
[Wen et al. 2003]

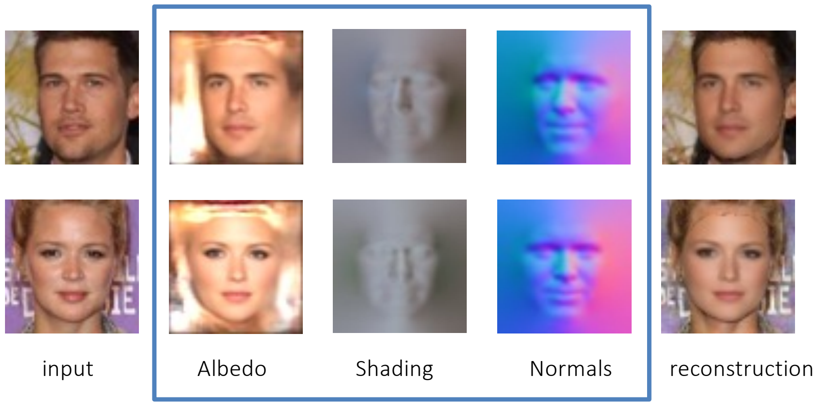
85

# Weak supervision: Normals & intrinsic images



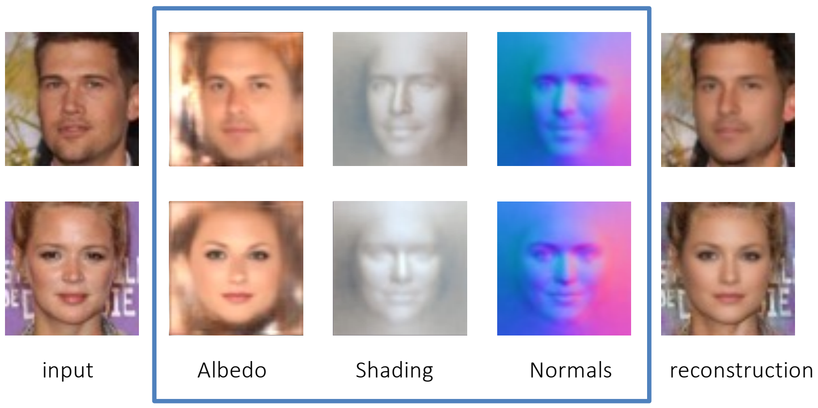

86

## Weak supervision: Normals




87

## Weak supervision: Normals & intrinsic images



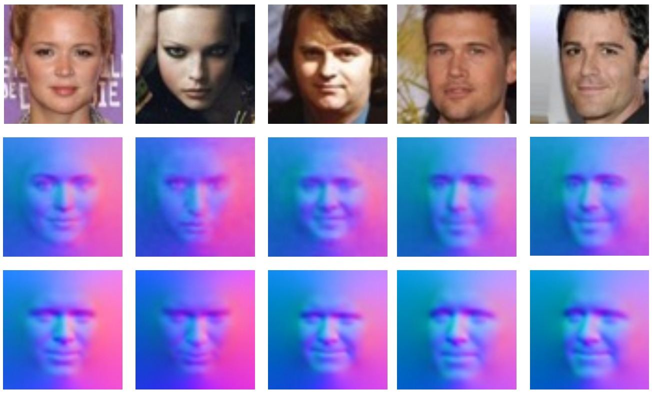

88

## Weak supervision: Normals & intrinsic images

input

decoded Normals

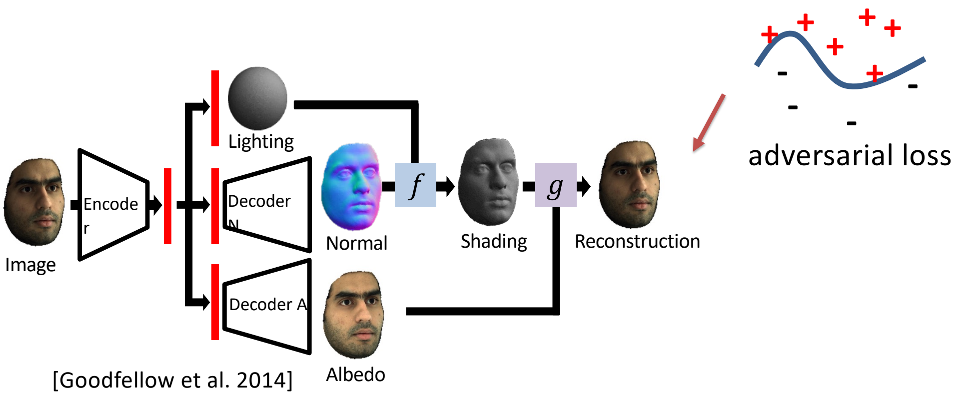
Normals from morphable model*



*Morphable model constructed from dataset with only 16 identities
*Used as guiding loss

89

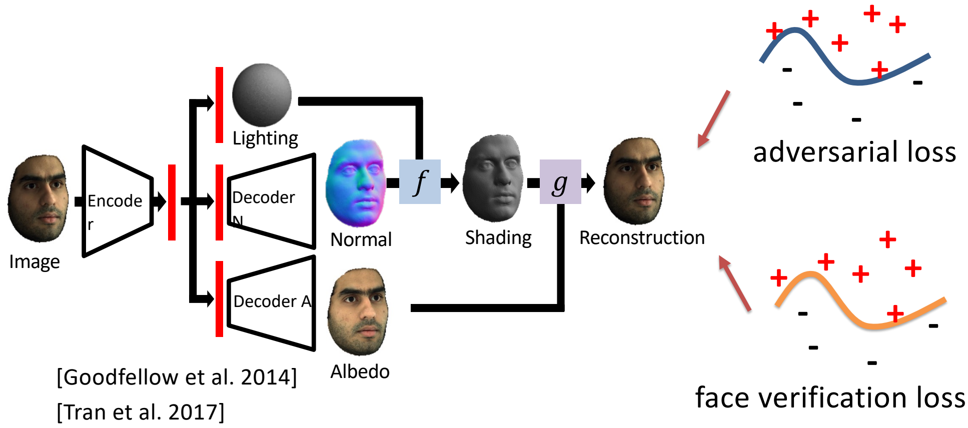
## Adversarial and face verification loss




[Goodfellow et al. 2014]

90

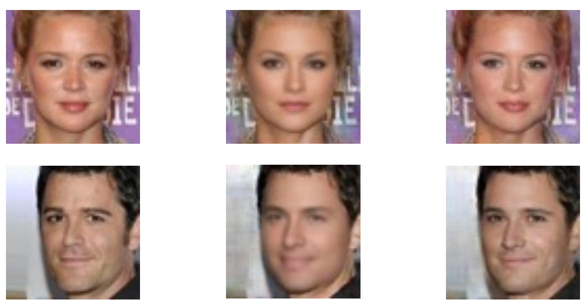
## Adversarial and face verification loss




[Goodfellow et al. 2014]
[Tran et al. 2017]

91

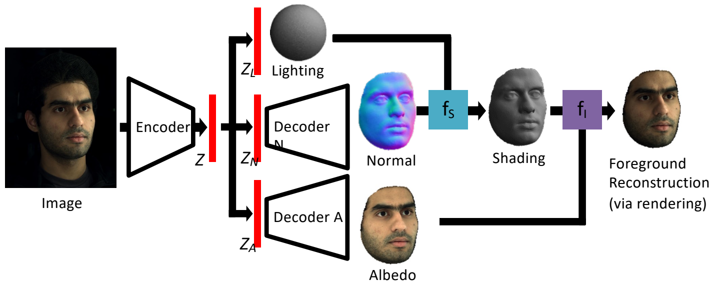
## Adversarial and face verification loss



| input | reconstruction without adversarial and verification loss | reconstruction with adversarial and verification loss |
|---|---|---|

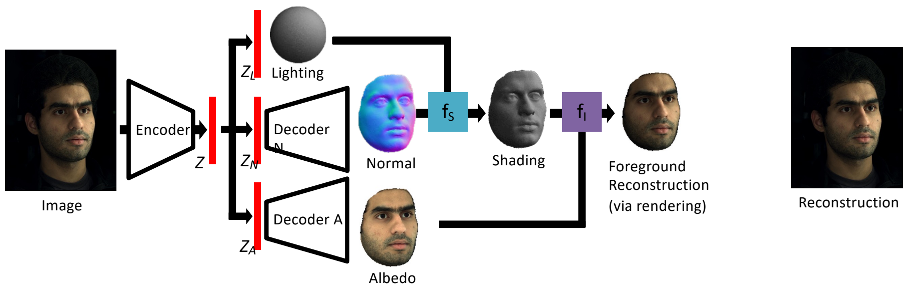
92

## Complete network architecture



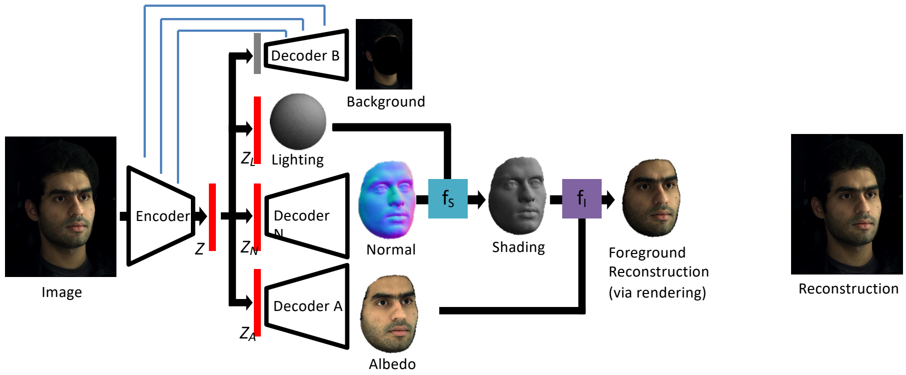

93

## Complete network architecture



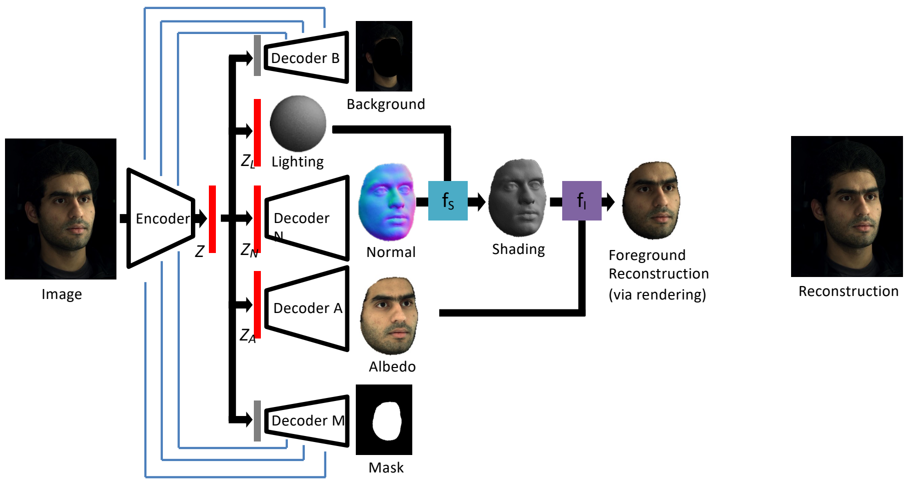

94

## Complete network architecture



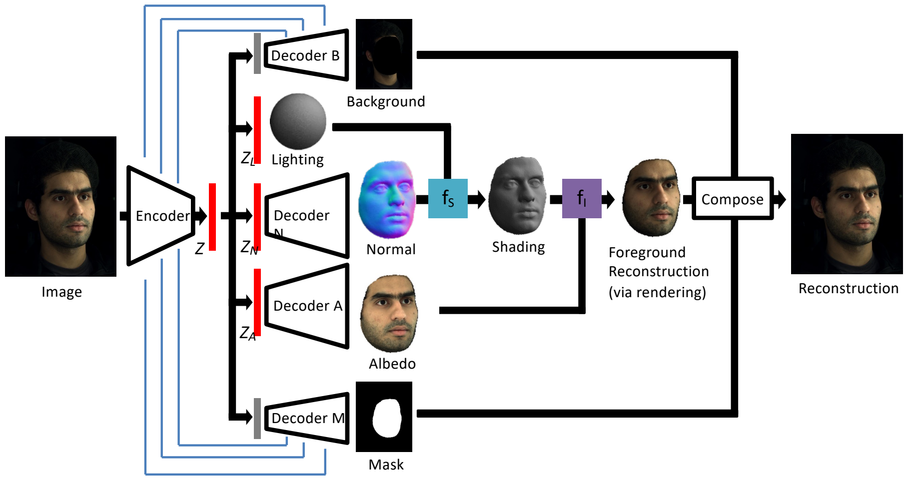

95

## Complete network architecture



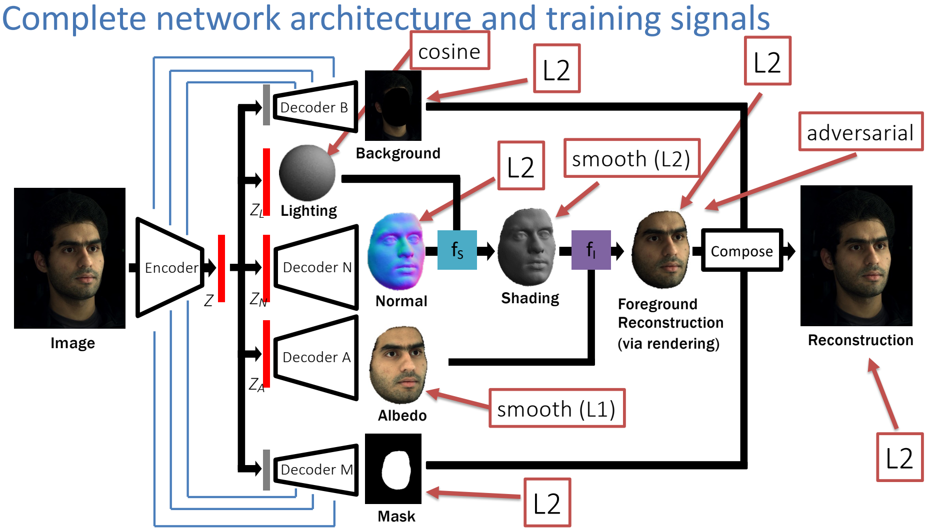

96

## Complete network architecture




97

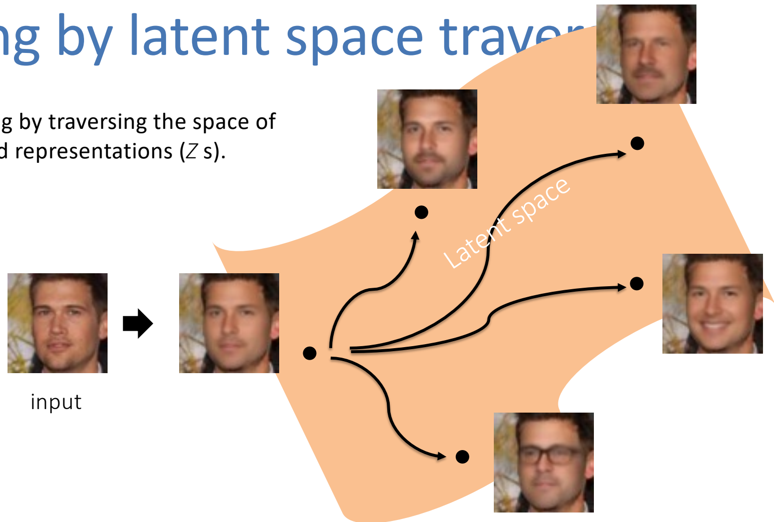
## Complete network architecture and training signals




98

## Editing by latent space traversal

Image editing by traversing the space of deep learned representations ($Z$ s).

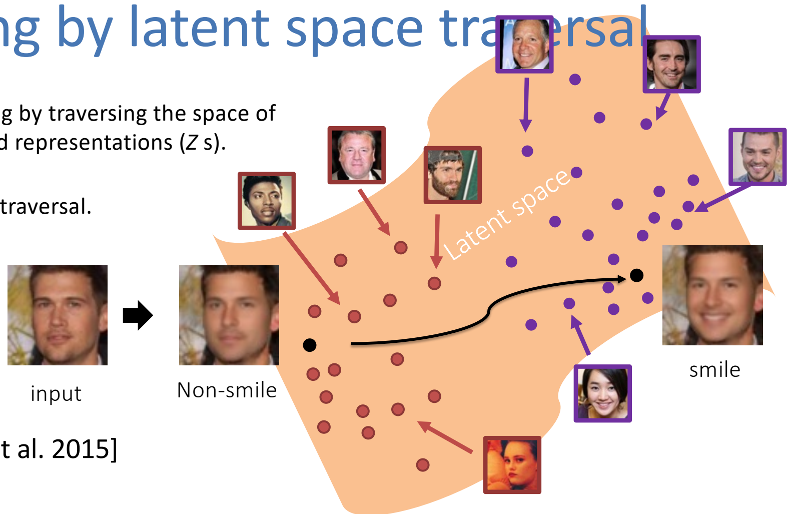



99

## Editing by latent space traversal

Image editing by traversing the space of deep learned representations ($Z$ s).

Data-driven traversal.

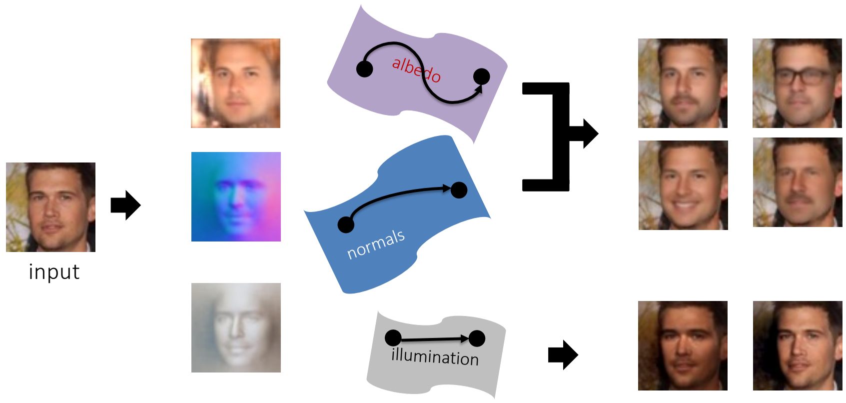



[Gardner et al. 2015]

100

# Editing by latent space traversal

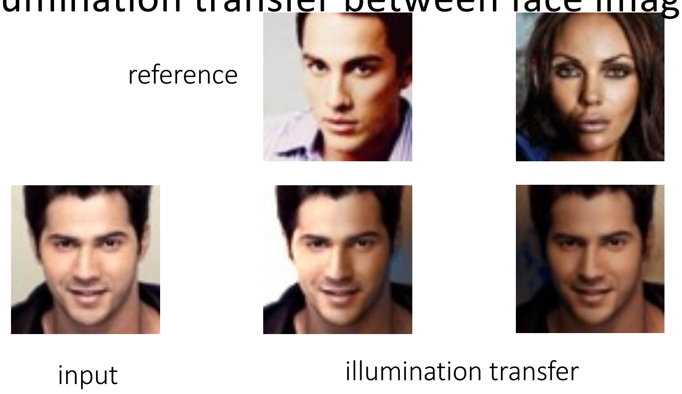
Editing in disentangled representations.






# Relighting

- Illumination transfer between face images



# Face editing results: smile

Image resolution: 64 x 64

103

# Face editing results: smile

smile 1
shorter traversal

Image resolution: 64 x 64

104

# Face editing results: smile

smile 1
longer traversal

Image resolution: 64 x 64

105

# Face editing results: smile

smile 1
shorter traversal

Image resolution: 64 x 64

106

# Face editing results: smile

Image resolution: 64 x 64

107

# Face editing results: smile

smile 1
shorter traversal

Image resolution: 64 x 64

108

## Face editing results: smile

smile 1
longer traversal

Image resolution: 64 x 64

109

## Face editing results: smile

input

smile 1
shorter traversal
λ= 0.1

smile 2
longer traversal
λ= 0.05

Image resolution: 64 x 64

110

# Face editing results: aging

Image resolution: 64 x 64

111

# Face editing results: aging

aging 1
shorter traversal

Image resolution: 64 x 64

112

## Face editing results: aging

aging 2
shorter traversal

Image resolution: 64 x 64

113

## Face editing results: aging

aging 1
shorter traversal

Image resolution: 64 x 64

114

# Face editing results: aging

Image resolution: 64 x 64

115

# Face editing results: aging

aging 1
shorter traversal

Image resolution: 64 x 64

116

## Face editing results: aging

aging 2
shorter traversal

Image resolution: 64 x 64

117

## Face editing results: aging

input

aging 1
shorter traversal
λ= 0.1

aging 2
longer traversal
λ= 0.05

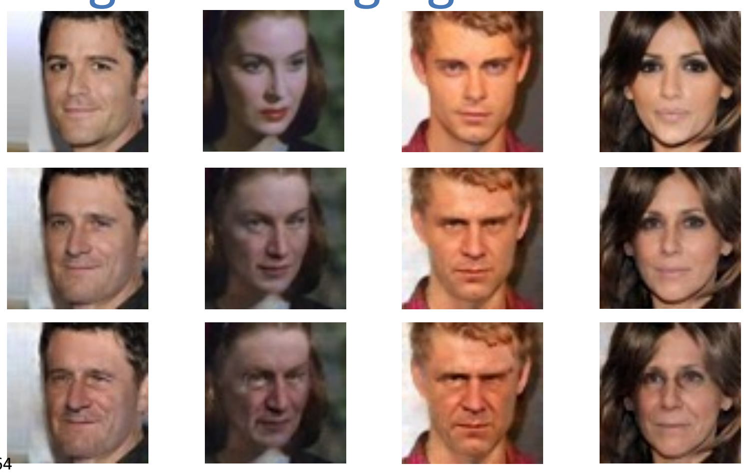


Image resolution: 64 x 64

118

## Face editing results: growing a beard

input

beard 1
shorter traversal
$\lambda = 0.1$

beard 2
longer traversal
$\lambda = 0.05$

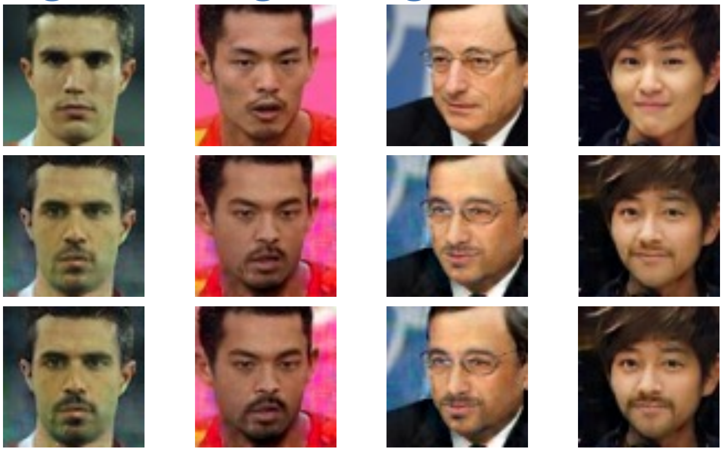
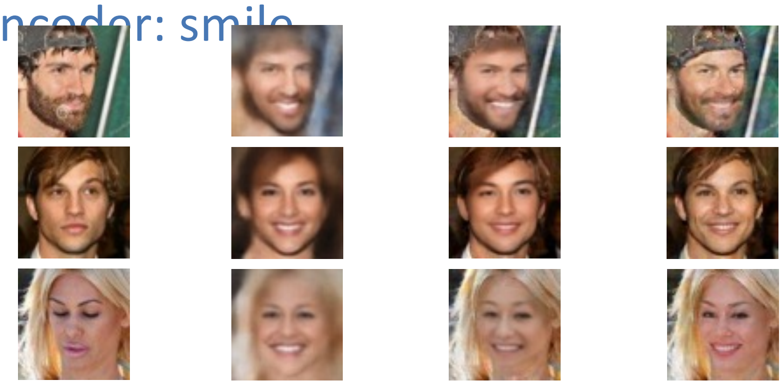


Image resolution: 64 x 64

119

## Comparison with a standard auto-encoder: smile



Input

Editing with standard auto-encoder

Editing with our encoder-decoders without adversarial and verification loss

Editing with our encoder-decoders with adversarial and verification loss

Image resolution: 64 x 64

120

## Conclusion

- In-network physically based rendering
- Constraints on physical components
  - Intrinsic image decomposition
- Weakly supervised training
  - No need for shape and reflectance ground-truth
- Editing as manifold traversal
  - Illumination
  - Expression
  - Semantic attributes

- Future work
  - explicit shape recovery, pose changes, high resolution editing

121