Color and Image Statistics

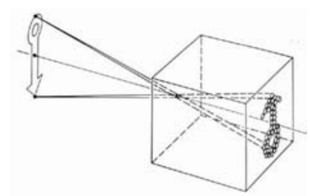
Today's Plan

- First part: Recap Last Lecture
  - how *positions* in the image relate to 3-d positions in the world.
- Second part:
  - how image *intensities* relate to surface and lighting properties in the world.
- Third part: ***Color and Image Statistics***
  - Object Detection using color models

Last Time

- **Pinhole camera models the geometry of perspective projection**
- **Lenses make it work in practice**
  - Refraction: Snell's law
  - Thin Lens Law
  - Image creation and chages in**:**
    - Focal length
    - Aperture
    - Focus distance …
- **Projections**
  - Perspective Projection
    - Non-linear projection
    - Pinhole, Camera
  - Weak Perspective Projection – linear
  - Orthographic – models telephoto lens

Pinhole camera model

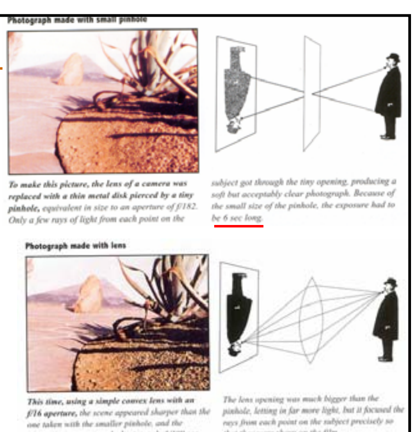


- Pinhole model:
  - Captures **pencil of rays** – all rays through a single point
  - The point is called **Center of Projection (COP)**
  - The image is formed on the **Image Plane**
  - **Effective focal length *f*** is distance from COP to Image Plane

Slide by Steve Seitz

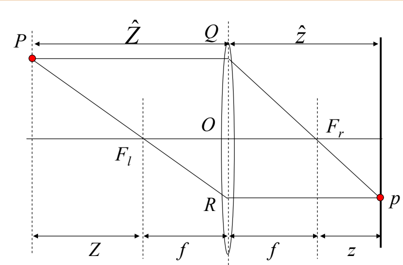
Lenses

- gather more light!
- But need to be focused



From Photography, London et al.

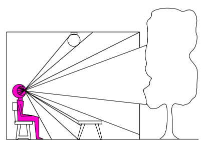
The Thin Lens Law

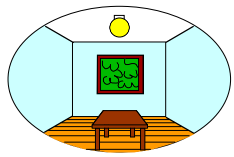


$$\frac{1}{\hat{z}} + \frac{1}{\hat{Z}} = \frac{1}{f}$$

Dimensionality Reduction Machine (3D to 2D)

3D world *2D image*

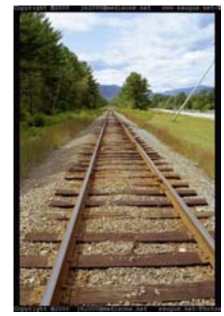




- What have we lost?
  - Angles
  - Distances (lengths)

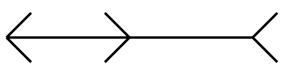
Figures © Stephen E. Palmer, 2002

Funny things happen…

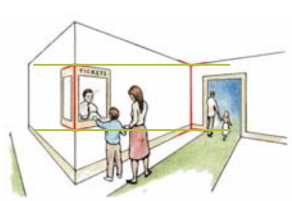


Parallel lines aren't…

…but humans adopt!



Müller-Lyer Illusion

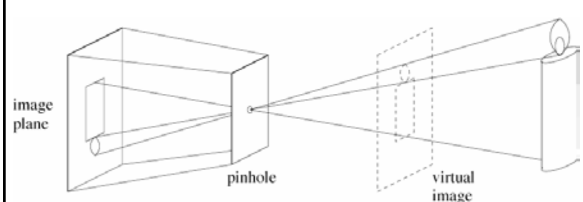


We don't make measurements in the image plane

http://www.michaelbach.de/ot/sze_muelue/index.html

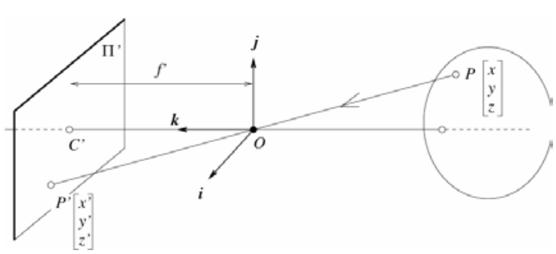
Perspective projection

- **Abstract camera model - box with a small hole in it**
- **In an ideal pinhole camera everything is in focus**



Forsyth&Ponce

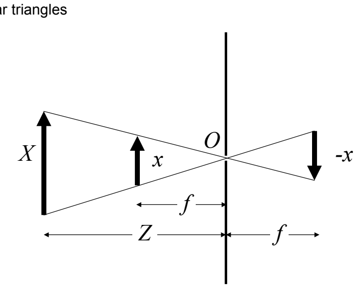
The equation of projection



2D image *3D world*

Perspective Projection

Similar triangles



$$x = -X\frac{f}{Z}$$

The equation of projection

- Cartesian coordinates:
  - We have, by similar triangles, that

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \rightarrow \begin{bmatrix} f\dfrac{X}{Z} \\ f\dfrac{Y}{Z} \\ -f \end{bmatrix}$$

  - Ignore the third coordinate, and get

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \rightarrow \begin{bmatrix} f\dfrac{X}{Z} \\ f\dfrac{Y}{Z} \end{bmatrix}$$

Homogeneous coordinates

- Is this a linear transformation?
  - no—division by z is nonlinear
- Trick: add one more coordinate:

$$(x, y) \Rightarrow \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \qquad (x, y, z) \Rightarrow \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

homogeneous image coordinates — homogeneous scene coordinates

- Converting *from* homogeneous coordinates

$$\begin{bmatrix} x \\ y \\ w \end{bmatrix} \Rightarrow (x/w, y/w) \qquad \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} \Rightarrow (x/w, y/w, z/w)$$

Slide by Steve Seitz

The camera matrix

- Turn previous expression into homogeneous coordinates
  - HC's for 3D point are *(X,Y,Z,t)*
  - HC's for point in image are *(u,v,w)*

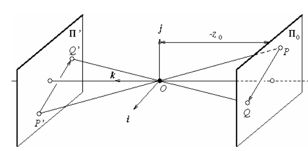
$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/f & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ t \end{bmatrix}$$

- Position of the point in the image from HC

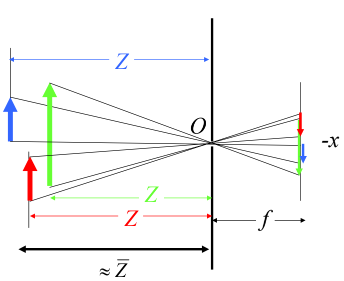
$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} X \\ Y \\ Z/f \end{pmatrix} \xrightarrow{\text{normalize by } w} \frac{1}{w}\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \frac{f}{Z}\begin{pmatrix} X \\ Y \\ Z/f \end{pmatrix} = \begin{pmatrix} fX/Z \\ fY/Z \\ 1 \end{pmatrix}$$

Weak perspective

- Issue
  - perspective effects, but not over the scale of individual objects
  - collect points into a group at about the same depth, then divide each point by the depth of its group
  - Adv: easy
  - Disadv: wrong



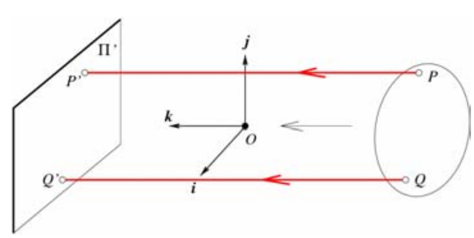
Weak Perspective Projection



$$x \approx X\frac{f}{\bar{Z}} = const \cdot X$$

Orthographic Projection

Telescope projection can be modeled by orthographic projection



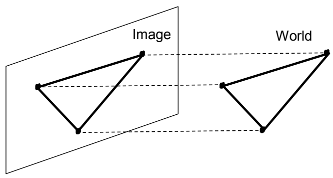
$$\boxed{\begin{cases} X = x \\ Y = y \end{cases}}$$

When the camera is at a (roughly constant) distance from the scene, take m=1.

Marc Pollefeys

Orthographic Projection

- Special case of perspective projection
  - Distance from the COP to the PP is infinite



  - Also called “parallel projection”
  - What’s the projection matrix?

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \Rightarrow (x, y)$$

Slide by Steve Seitz

Projection Summary:

1. Perspective

$$x = X\frac{f}{Z} \qquad y = Y\frac{f}{Z}$$

2. Weak perspective

$$x = const\ X \qquad y = const\ Y$$

3. Orthographic

$$x = X \qquad y = Y$$

x, y = image coordinates
X, Y, Z = world coordinates
Z = depth
f = focal length of the camera

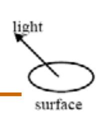
Today’s Plan

- First part:
  - how *positions* in the image relate to 3-d positions in the world.
- Second part:
  - how image *intensities* relate to surface and lighting properties in the world.
- Third part:
  - Color Spaces
  - Object Detection using color models

Radiometry

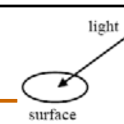
- Relationship between the world and its image
- Scene Radiance
  - Amount of light radiating from a surface point
- Image Irradiance
  - Amount of light incident at an image point

Radiance, L



- Amount of light radiated from a surface into a given solid angle per unit area (watts per square meter per steradian).
- Note: the area is the foreshortened area, as seen from the direction that the light is being emitted.
- Informally, radiance tells you the “brightness”.

Irradiance, E



- Light power per unit area (watts per square meter) incident on a surface.
- The units tell you what to integrate over to find the energy impinging on a given area.
- E times pixel area, times exposure time gives the pixel intensity out (for linear sensor response)

Image irradiance/scene radiance relationship

- The definition of scene radiance is constructed so that image irradiance is proportional to scene radiance.

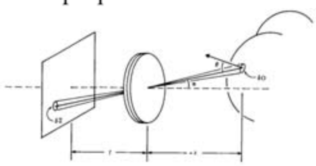


Figure 10-4. To see how image irradiance is related to the radiance of the surface, we must determine the size of the region in the image that corresponds to the patch on the surface.

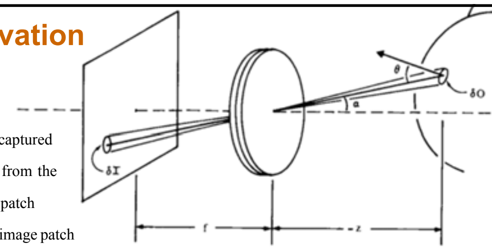
Horn, sect. 10.3

Scene radiance

$$E = L\frac{\pi}{4}\left(\frac{d}{f}\right)^2 \cos^4(\alpha)$$

Image irradiance

Derivation



δP = power captured by lens from the surface patch

δi = area of image patch

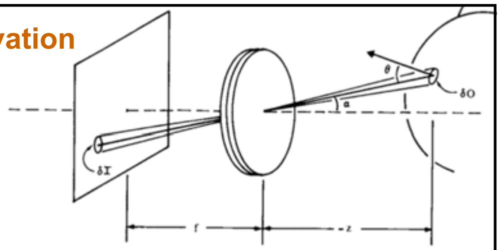
$$E = \frac{\delta P}{\delta I}$$

$$\delta P = L(\delta O \cos\theta)\frac{\pi\left(d/2\right)^2 \cos\alpha}{(z/\cos\alpha)^2}$$

$$\frac{(\delta I \cos\alpha)}{(f/\cos\alpha)^2}$$

$$\delta P = L\frac{\delta I \cos^2\alpha}{(f/\cos\alpha)^2}\pi\left(d/2\right)^2$$

Derivation



$$\frac{\delta P}{\delta I} = L\frac{\cos^2\alpha}{(f/\cos\alpha)^2}\pi\left(d/2\right)^2$$

$$= L\frac{\pi}{4}\cos^4\alpha\left(\frac{d}{f}\right)^2$$

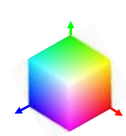
Color Spaces

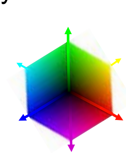
Different choices for color spaces

- RGB
- Normalized RGB
- HIS, HSV, HSL
  - Fleck HSV
- TSL
- YcrCb
- Perceptually uniform colors
  - CIELAB, CIELUV
- Others
  - YES, YUV, YIQ, CIE-xyz

RGB – Red, Green, Blue

- Most common color space used to represent images.
- Was developed with CRT as an additive color space
- [1] – Rehg and Jones used this color space to study the separability of the color space





Normalized RGB – rg space

$$r = \frac{R}{R+G+B} \quad g = \frac{G}{R+G+B} \quad b = \frac{B}{R+G+B} = 1 - r - g$$

- 2D color space as 'b' component is redundant
  - $r + g + b = 1$
- For matte surfaces (ignoring ambient light):
  - Normalized RGB is invariant to changes of surface orientation relatively to the light source
    [Skarbek and Koschan 1994]

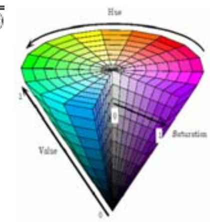
HSV, HSI, HSL

(hue, saturation, value/intensity/luminance)

$$H = \arccos \frac{\frac{1}{2}((R-G)+(R-B))}{\sqrt{((R-G)^2+(R-B)(G-B))}}$$

$$S = 1 - 3\frac{\min(R,G,B)}{R+G+B}$$

$$V = \frac{1}{3}(R+G+B)$$



- High cost of conversion
- Based on intuitive values
- Invariant to highlight at white light sources
- Pixel with large and small intensities are discarded as HS becomes unstable.
- Can be 2D by removing the illumination component

Y Cr Cb

$$Y = 0.299R + 0.587G + 0.114B$$
$$C_r = R - Y$$
$$C_b = B - Y$$

- *YCrCb* is an encoded nonlinear RGB signal, commonly used by European television studios and for image compression work.
- Y – Luminance component,
- C – Chrominance

Color Models for Object Detection

- Example: Skin Color Tracker

Skin color detection

- Are Skin and Non-skin colors separable?
  - Illumination changes over time.
  - Skin tones vary dramatically within and identical image. across individuals.
  - Different cameras have different output for the
  - Movement of objects cause blurring of colors.
  - Ambient light, shadows change the apparent color of the image.
- What color space should we use?
- How should the color distribution be modelled?

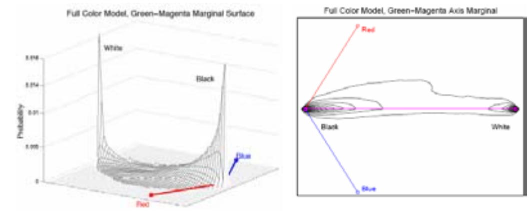
Color Models

- Desired Properties:
  - Increased separability between skin and non skin classes
  - Decreased separability among skin tones
  - Stability of color space (at extreme values)
    - Cost of conversion for real time applications
- Multiple choices for color spaces:
  - Stability of color space (at extreme values)
  - Keeping the Illumination component – 2D color space vs. 3D color space
- Multiple choices of color distribution model

Perceptually uniform colors

- “skin color” is not a physical property of an object, rather a perceptual phenomenon and therefore a subjective human concept.
- Color representation similar to the color sensitivity of human vision system should
- Complex transformation functions from and to RGB space, demanding far more computation than most other color spaces

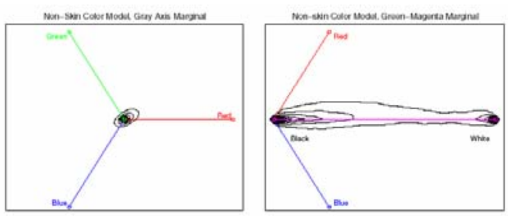
General Color Model



Observations:

1. Most colors fall on or near the gray line;
2. Black and white are by far the most frequent colors, with white occuring slightly more frequently;
3. There is a marked skew in the distribution toward the red corner of the color cube.

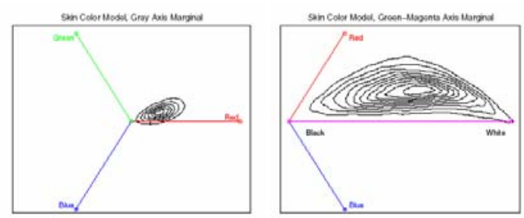
Non-Skin Pixel Distribution



(d) Contour plot for non-skin model, marginalized along the gray axis.

(c) Contour plot for non-skin model, marginalized along the green-magenta axis.

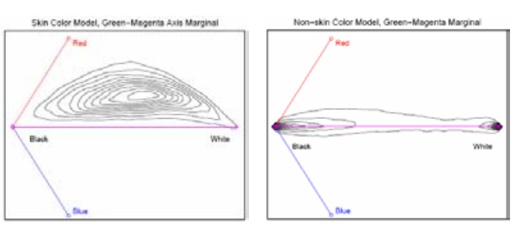
Skin Pixel Distribution



(b) Contour plot for skin model, marginalized along the gray axis.

(a) Contour plot for skin model, marginalized along the green-magenta axis.

Skin and Non-skin Color Distribution



Observations:

1. Non-skin model is the general model without skin pixels (10% of pixels);
2. There is a significant degree of separation between the skin and non-skin models;

Results from Rehg & Jones

- Used 18,696 images to build a general color model.
- Density is concentrated around the gray line and is more sharply peaked at white than black.
- Most colors fall on or near the gray line.
- Black and white are by far the most frequent colors, with white occurring slightly more frequently.
- There is a marked skew in the distribution toward the red corner of the color cube.
- 77% of the possible 24 bit RGB colors are never encountered (i.e. the histogram is mostly empty).
- 52% of web images have people in them.

Modeling the color distribution

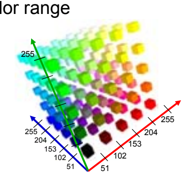
- **Non parametric** – Estimate skin color distribution from skin training data without deriving an explicit model of the skin.
  - Look up table or Histogram Model
- **Parametric** – Deriving a parametric model from skin training set
  - Gaussian Model

Histogram Color Models

- Images are organized in two sets:
  - Generic Training Set;
    - Used to compute a general histogram density;
  - Classifier Training Set;
    - Used to build the skin and non-skin models;
    - Manually separated into subsets containing skin and not containing skin;
    - Skin pixels are manually labeled;

Histogram/Look-Up Table

- Color space is quantized into a number of bins, where each bin corresponds to a color range



- Bins, forming a 3D histogram are referred to as the lookup table (LUT).
- Each bin stores the number of times a particular RGB color, **x,** occurred in the training skin samples

Skin Nomalized Histogram/Look-Up Table

- Likelihood that the RGB color, **x**, will correspond to skin
  - normalizing the histogram counts

$$P(\mathbf{x}|\text{skin}) = \frac{\text{value of the histogram bin, corresponding to color vector x}}{\text{sum of all histogram bin values}}$$

$p(\boldsymbol{x}|skin)$ - a probability of observing color $\boldsymbol{x}$, knowing that we see a skin pixel.

- Probability of encountering skin pixels

$$P(\text{skin}) = \frac{\text{total number of skin pixels}}{\text{total number of pixels}}$$

Non-Skin Nomalized Histogram

- Likelihood that the RGB color, **x**, will correspond to skin
  - normalizing the histogram counts

$$P(\mathbf{x}|\neg\text{skin}) = \frac{\text{value of the histogram bin, corresponding to color vector x}}{\text{sum of all histogram bin values}}$$

$p(\boldsymbol{x}|\neg skin)$ - a probability of observing color $\boldsymbol{x}$, knowing that we see a non-skin pixel.

- Probability of encountering non-skin pixels

$$P(\neg\text{skin}) = \frac{\text{total number of nonskin pixels}}{\text{total number of pixels}}$$

Skin Detection Using Color Models

- Given skin and non-skin histogram models, we can construct a skin pixel classifier
- Classifiers:
  - Maximum Likelihood Classifier
  - Bayes Classifier
- Skin classifier is useful in:
  - Detection and recognition of faces and figures;
  - Image indexing and retrieval

Bayesian Rule Classification

- Given: p(**x**|skin) and p(**x**|non-skin)
- Interested in finding the probability that a particular pixel belongs to skin class given its RGB value, **x**
- Probability of skin given a pixel's RGB value, **x**:

$$p(skin|\mathbf{x}) = \frac{p(\mathbf{x}|skin)\,p(skin)}{p(\mathbf{x}|skin)\,p(skin) + p(\mathbf{x}|\neg skin)\,p(\neg skin)}$$

$$> .5$$

Deriving ML from MAP (max. likelihood from max. a posteriori)

$$p(skin|\mathbf{x}) = \frac{p(\mathbf{x}|skin)\,p(skin)}{p(\mathbf{x}|skin)\,p(skin) + p(\mathbf{x}|\neg skin)\,p(\neg skin)}$$

$$p(\neg skin|\mathbf{x}) = \frac{p(\mathbf{x}|\neg skin)\,p(\neg skin)}{p(\mathbf{x}|skin)\,p(skin) + p(\mathbf{x}|\neg skin)\,p(\neg skin)}$$

$$\frac{p(skin|\mathbf{x})}{p(\neg skin|\mathbf{x})} = \frac{p(\mathbf{x}|skin)\,p(skin)}{p(\mathbf{x}|\neg skin)\,p(\neg skin)} > k$$

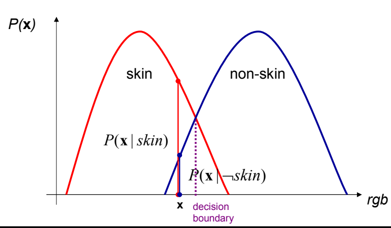
$$\frac{p(\mathbf{x}|skin)}{p(\mathbf{x}|\neg skin)} > \Theta \qquad \Theta = K\,\frac{(1 - p(skin))}{p(skin)}$$

$$p(skin) + p(\neg skin) = 1$$

Maximum Likelihood Classification

- a skin pixel classifier is derived through the standard likelihood ratio approach:

$$\frac{P(\mathbf{x}\,|\,skin)}{P(\mathbf{x}\,|\,\neg skin)} \geq \Theta \qquad \mathbf{x} = \begin{bmatrix} r \\ g \\ b \end{bmatrix}$$



Histogram-based Skin Classifier

- Qualitative observations:
  - θ = 0.4;
  - The classifier does a good job of detecting skin in most examples;
  - In particular, the skin labels form dense sets whose shape often resembles that of the true skin pixels;
  - The detector tends to fail on highly saturated or shadowed skin;
  - The performance of the skin classifier is surprisingly good considering the unconstrained nature of Web images;

Histogram-based Skin Classifier

Histogram-based Skin Classifier

- More qualitative observations:
  - The example photos also show the performance of the detector on non-skin pixels.
  - In photos such as the house (lower right) or flowers (upper right) the false detections are sparse and scattered.
  - More problematic are images with wood or copper-colored metal such as the kitchen scene (upper left) or railroad tracks (lower left).
  - These photos contain colors which often occur in the skin model and are difficult to discriminate reliably.
  - This results in fairly dense sets of false postives.

Histograms for object recognition

- Remarkable success of recognition methods using histograms of local image measurements:
  - [Swain & Ballard 1991] - *Color histograms*
  - [Schiele & Crowley 1996] - *Receptive field histograms*
  - [Lowe 1999] - *localized orientation histograms (SIFT)*
  - [Schneiderman & Kanade 2000] - *localized histograms of wavelet coef.*
  - [Leung & Malik 2001] - *Texton histograms*
  - [Belongie et.al. 2002] - *Shape context*
  - [Dalal & Triggs 2005] - *Dense orientation histograms*
- Likely explanation: *Histograms are robust to image variations such as limited geometric transformations and object class variability.*

Histogram-based Skin Classifier

- More quantitative observations:
  - The performance of the skin classifier is surprisingly good considering the unconstrained nature of Web images;
  - The best classifier (size 32) can detect roughly 80% of skin pixels with a false positive rate of 8.5%, or 90% correct detections with 14.2% false positives;
  - Its equal error rate is 88%.

Non-Parametric Models

- Advantages of non-parametric methods:
  - they are fast in training and usage:
    - use of the histogram model results in a fast classifier since only two table lookups are required to compute the probability of skin.
  - they are theoretically independent to the shape the color skin distribution
- Disadvantages:
  - large storage space required and
  - inability to interpolate or generalize the training data
  - ***performance directly depends on the representativeness of the training images set.***

Parametric Models

- Compact skin model representation
- Can generalize and interpolate the training data
- Models:
  - Single Gaussian Model for Skin
  - Mixture of Gaussians

Gaussian Model

- Two separate gaussian models (or mixtures of gaussians) can be trained for the skin and non-skin classes;
- Gaussian Model:

$$\mathbf{x} = \begin{bmatrix} r \\ g \\ b \end{bmatrix}$$

mean $\boldsymbol{\mu} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}_n$

covariance $\Sigma = \frac{1}{N-1}\mathbf{D}\mathbf{D}^T$

$$\mathbf{D} = [\mathbf{x}_1 - \boldsymbol{\mu} \quad \cdots \quad \mathbf{x}_N - \boldsymbol{\mu}]$$

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})}$$

Skin & Non-skin Color Gaussian Model

- Skin Model:
  - Conditional Density:
    - prob of RGB value, **x**, given a skin sample
      (d is the dimensionality of **x**)

$$p(\mathbf{x}|\text{skin}) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_{\text{skin}}|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_{\text{skin}})^T \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu}_{\text{skin}})}$$

- Non-skin Model
  - Conditional Density:
    - prob of RGB value, **x**, given a non-skin sample

$$p(\mathbf{x}|\text{non-skin}) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_{\text{non-skin}}|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_{\text{non-skin}})^T \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu}_{\text{non-skin}})}$$

Classification

- Bayes Rule Classification – Maximum A Posteriori

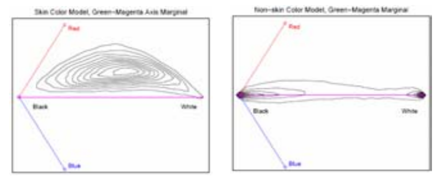
$$p(skin|\mathbf{x}) = \frac{p(\mathbf{x}|skin)\, p(skin)}{p(\mathbf{x}|skin)\, p(skin) + p(\mathbf{x}|\neg skin)\, p(\neg skin)}$$

- Maximum Likelihood

$$\frac{p(skin|\mathbf{x})}{p(\neg skin|\mathbf{x})} = \frac{p(\mathbf{x} \mid skin)}{p(\mathbf{x} \mid \neg skin)} \geq \Theta$$

Mixture of Gaussian Model

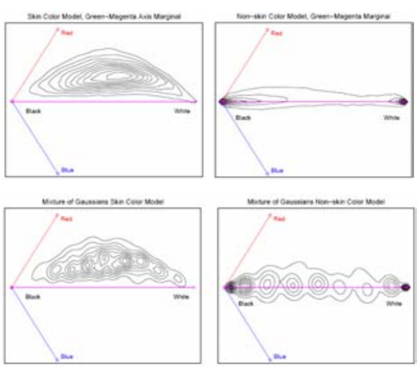
- Skin/Non-skin pixel color of have complicated distributions that are not easily described by a single gaussian each



Mixture of Gaussian Model

Skin and Non-Skin Color Distribution:

Mixture of Gaussians:



Mixture Model Classification

Skin Mixture Model:

$$p(\mathbf{x} \mid skin) = \sum_{g_s=1}^{G_s} w_{g_s} p_{g_s}(x \mid skin)$$

Non-Skin Mixture Model:

$$p(\mathbf{x} \mid \neg skin) = \sum_{g_n=1}^{G_n} w_{g_n} p_{g_n}(\mathbf{x} \mid \neg skin)$$

- Classification:
  - Maximum Likelihood
  - Bayes Rule Classification

Gaussian Models

Advantages:

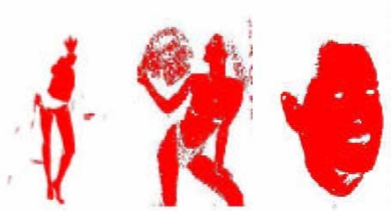
- One advantage of gaussian model (or mixture models) is that they can be made to generalize well on small amounts of training data;
- .From the standpoint of storage space, the gaussian (mixture of gaussian) model is a much more compact representation of the data.

Gaussian Models

Disadvantages:

- The mixture of Gaussian model is significantly more expensive to train than the histogram models;
- It took 24 hours to train both skin and non-skin mixture of gaussian models using 10 Alpha workstations in parallel. In contrast, the histogram models could be constructed in a matter of minutes on a single workstation;
- The mixture model is also slower to use during classification since all of the Gaussians must be evaluated in computing the probability of a single color value;

Adult Image Detector



(a) Examples of images correctly classified by our detector. Both images were classified as adult images.

(b) Example of an image misclassified as adult by our detector.

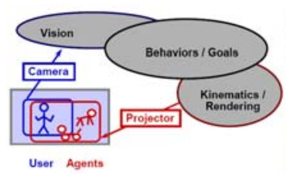
"Funny Mirrors"

ALIVE

http://vismod.www.media.mit.edu/cgi-bin/tr_pagemaker (TR 257)

ALIVE

- Real sensing for virtual world
- Tightly coupled sensing-behavior-action
- Vision routines: body/head/hand tracking



[Blumberg, Darrell, Maes, Pentland, Wren, ... 1995]

Conclusions

- Color distributions for skin and non-skin pixel classes learned from web images can be used as an accurate pixel-wise skin detector;
- The key is the use of a very large labeled dataset to capture the effects of the unconstrained imaging environment represented by web photos;
- Visualization studies show a surprising degree of separability in the skin and non-skin color distributions;
- They also reveal that the general distribution of color in web images is strongly biased by the presence of skin pixels.

Conclusions

- One possible advantage of using a large dataset is that simple learning rules may give good performance;
- A pixel-wise skin detector can be used to detect images containing naked people, which tend to produce large connected regions of skin;
- It is shown that a detection rate of 88% can be achieved with a false alarm rate of 11.3%, using a seven element feature vector and a neural network classifier;
- This performance is comparable to systems which use more elaborate and slower spatial image analysis;
- The results suggest that skin color is a very powerful cue for detecting people in unconstrained imagery.