Lecture-17

Mixture of Gaussians

Grimson

Algorithm

- Learn background model by watching 30 second video
- Detect moving object by measuring deviations from background model, and applying connected component to foreground pixels.
- Predict position of a region in the next frame using Kalman filter
- Update background and blob statistics

Summary

- Each pixel is an independent statistical process, which may be combination of several processes.
  - Swaying branches of tree result in a bimodal behavior of pixel intensity.
- The intensity is fit with a mixture of K Gaussians.

$$\Pr(X_t) = \sum_{j=1}^{K} \frac{w_j}{(2p)^{\frac{m}{2}} |\Sigma_j|^{\frac{1}{2}}} e^{-\frac{1}{2}(X_t - m_j)^T \Sigma_j^{-1} (X_t - m_j)}$$

Mixture of Gaussians

• The K distributions are stored in descending order of the term $\frac{w_j}{s_j}$

• Out of "k" distributions, the first B are selected

$$B = \arg\min_b \left[\frac{\sum_{j=1}^{b} w_j}{\sum_{j=1}^{K} w_j} > T \right]$$

Learning Background Model

• Every new pixel is checked against all existing distributions. The match is the first distribution such that the pixel value lies within 2 standard deviations of mean.

•If no match, introduce new distribution.

Updating

• The mean and s.d. of unmatched distributions remain unchanged. For the matched distributions they are updated as:

$$m_{j,t} = (1-r)m_{j,t-1} + rX_t$$

$$s_{j,t} = (1-r)s^2_{j,t-1} + r(X_t - m_{j,t})^T(X_t - m_{j,t})$$

• The weights are adjusted:

$$w_{j,t} = (1-a)w_{j,t-1} + a(M_{j,t}) \qquad M_{j,t} = \begin{cases} 1 & \text{if distributi on matches} \\ 0 & \text{otherwise} \end{cases}$$

Segmenting Background

- Any pixel that is more than 2 sd from all the distributions is marked as a part of foreground-moving object.
- Such pixels are then clustered into connected components.

Kanade

Summary

- Very similar to k-Gaussian with following differences:
  - uses only single Gaussian
  - uses gray level images, the mean and variance are scalar values

Algorithm

- Learn background model by watching 30 second video
- Detect moving object by measuring deviations from background model, and applying connected component to foreground pixels.
- Update background and region statistics

Detection

- During detection if intensity value is more than two sigma away from the background it is considered foreground:
  - keep original mean and variance
  - track the object with new mean and variance
  - if new mean and variance persists for sometime, then substitute the new mean and variance as the background model
  - If object is no longer visible, it is incorporated as part of background

W4 (Who, When, Where, What)

Davis

W4

• Compute “minimum”(M(x)), “maximum” (N(x)), and “largest absolute difference” (L(x)).

$$D_i(x,y) = \left\{ \begin{array}{c} 1 \quad if \quad |M(x,y) - f_i(x,y)| > L(x,y) or \\ |N(x,y) - f_i(x,y)| > L(x,y) \\ 0 \quad \ldots \quad otherwise \end{array} \right\}$$

- Theoretically, the performance of this tracker should be worse than others.
- Even if one value is far away from the mean, then that value will result in an abnormally high value of L.
- Having short training time is better for this tracker.

Limitations

- Multiple people
- Occlusion
- Shadows
- Slow moving people
- Multiple processes (swaying of trees..)

Webpage

- Http://www.cs.cmu.edu/~vsam

Skin Detection

Kjeldsen and Kender

Training

- Crop skin regions in the training images.
- Build histogram of training images.
- Ideally this histogram should be bi-modal, one peak corresponding to the skin pixels, other to the non-skin pixels.
- Practically there may be several peaks corresponding to skin, and non-skin pixels.

Training

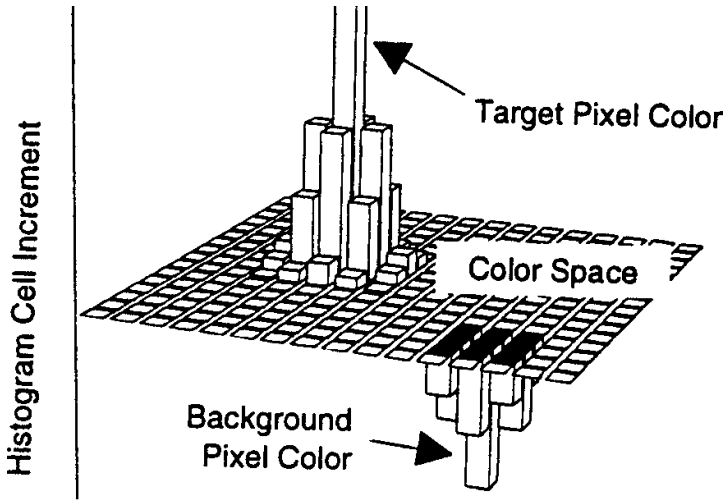
- Apply threshold to skin peaks to remove small peaks.
- Label all gray levels (colors) under skin peaks as “skin”, and the remaining gray levels as “non-skin”.
- Generate a look-up table for all possible colors in the image, and assign “skin” or “non-skin” label.

Detection

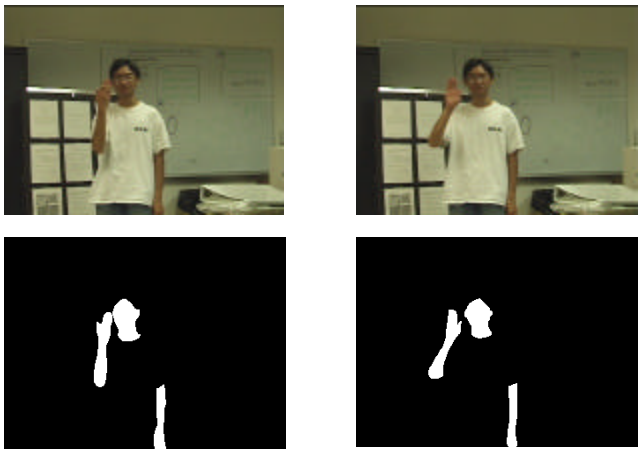
- For each pixel in the image, determine its label from the “look-up table” generated during training.

Building Histogram

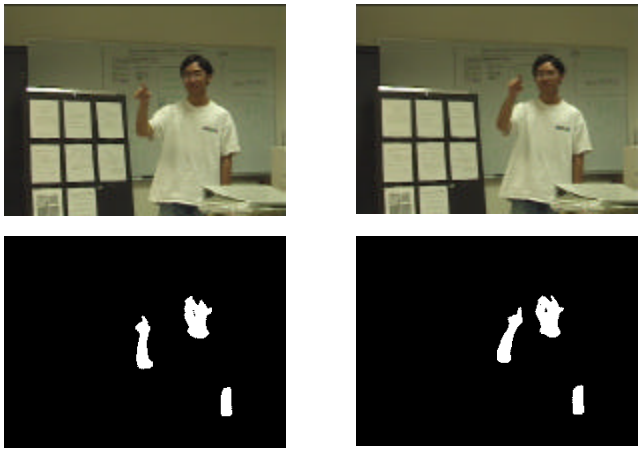
- Instead of incrementing the pixel counts in a particular histogram bin:
  - for skin pixel increment the bins centered around the given value by a Gaussian function.
  - For non-skin pixels decrement the bins centered around the given value by a smaller Gaussian function.



Example training images

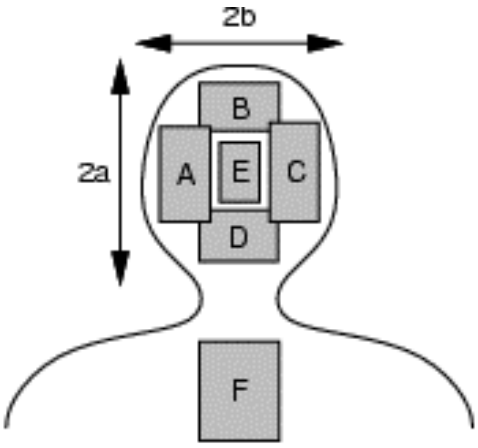


Results of skin detection



Tracking People Using Color

Fieguth and Terzopoulos



Fieguth and Terzopoulos

• Compute mean color vector for each sub region.

$$(r_i, g_i, b_i) = \frac{1}{|R_i|} \sum_{(x,y) \in R_i} (r(x,y), g(x,y), b(x,y))$$

Fieguth and Terzopoulos

- Compute goodness of fit.

$$\Psi_i = \frac{\max\left\{\frac{r_i}{\bar{r}_i}, \frac{g_i}{\bar{g}_i}, \frac{b_i}{\bar{b}_i}\right\}}{\min\left\{\frac{r_i}{\bar{r}_i}, \frac{g_i}{\bar{g}_i}, \frac{b_i}{\bar{b}_i}\right\}}$$

$(\bar{r}_i, \bar{g}_i, \bar{b}_i)$ (r_i, g_i, b_i)

Target Measurement

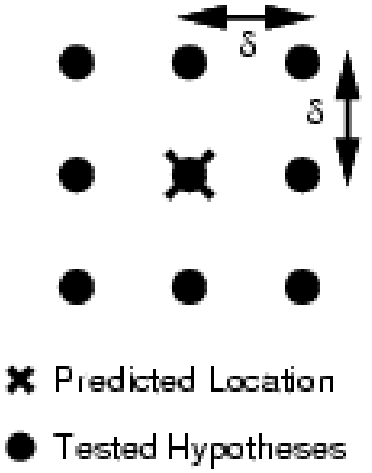
Fieguth and Terzopoulos

- Tracking

$$\Psi(x_H, y_H) = \sum_{i=1}^{N} \frac{\Psi_i(x_H + x_i, y_H + y_i)}{N}$$

$$(\hat{x}, \hat{y}) = \arg_{(x_H, y_H)} \min\{\Psi(x_H, y_H)\}$$

Fieguth and Terzopoulos



Fieguth and Terzopoulos

- Non-linear velocity estimator

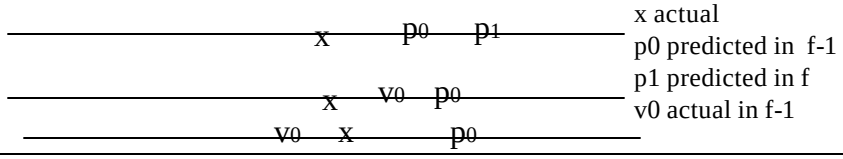
$$v(f) = v(f-1)$$

$$if \quad (r(f).r(f-1) > 0) \quad v(f) \; += \; d\frac{\text{sgn}(r(f))}{\Delta t}$$

$$if \quad (r(f).v(f-1) < 0) \quad v(f) \; += \; d\frac{\text{sgn}(r(f))}{\Delta t}$$

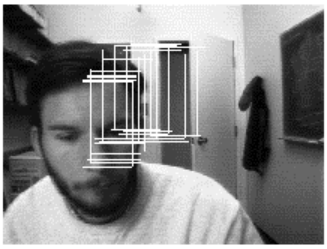
$$if \quad (r(f) = 0) \quad v(f) \; -= \; d\frac{\text{sgn}(v(f))}{2\Delta t}$$

r is an error in prediction.



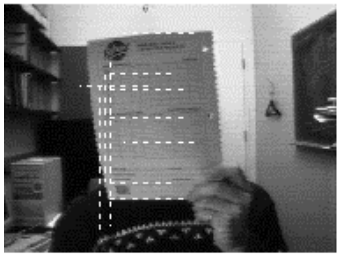
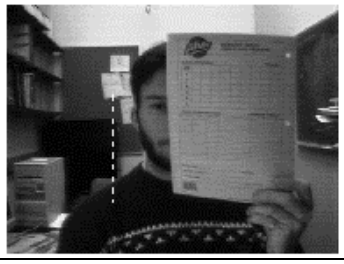
x actual
p0 predicted in f-1
p1 predicted in f
v0 actual in f-1

Fieguth and Terzopoulos


Fieguth and Terzopoulos



Bibliography

- .J. K. Aggarwal and Q. Cai, "Human Motion Analysis: A Review", *Computer Vision and Image Understanding*, Vol. 73, No. 3, March, pp. 428-440, 1999
- .Azarbayejani, C. Wren and A. Pentland, "Real-Time 3D Tracking of the Human Body", MIT Media Laboratory, Perceptual Computing Section, TR No. 374, May 1996
- .W.E.L. Grimson *et. al.*, "Using Adaptive Tracking to Classify and Monitor Activities in a Site", *Proceedings of Computer Vision and Pattern Recognition,* Santa Barbara, June 23-25, 1998, pp. 22-29

Bibliography

- .Takeo Kanade *et. al.* "Advances in Cooperative Multi-Sensor Video Surveillance", *Proceedings of Image Understanding workshop*, Monterey California, Nov 20-23, 1998, pp. 3-24
- .Haritaoglu I., Harwood D, Davis L, "*W*^{4} - Who, Where, When, What: A Real Time System for Detecting and Tracking People", *International Face and Gesture Recognition Conference,* 1998
- .Paul Fieguth, Demetri Terzopoulos, "Color-Based Tracking of Heads and Other Mobile Objects at Video Frame Rates", *CVPR 1997,* pp. 21-27