Robbery Detection Camera

Vincenzo Caglioti
Simone Gasparini
Giacomo Boracchi
Pierluigi Taddei
Alessandro Giusti

ARTIFICIAL INTELLIGENCE and ROBOTICS LABORATORY

Camera and DSP

Camera used

- VGA camera (640x480)
- [Y, Cb, Cr] color coding, chroma interlaced
  - Y is 640x480, Cb and Cr are 320x480
- Auto white balance
- Auto luminance balance
- Ethernet connection
- The program is stored into the camera DSP:
  - Image processing is done directly onboard

Algorithm outline

- Skin Pixel Detection
- Background Subtraction
- Skin Segmentation
  - Morphological Image Processing
  - Blob flooding
- Blob Metric Analysis
- Hand Detector
- Face Detector
- Alarm Trigger

Colour Images

YCbCr Model

» $Y = 0.299 * R + 0.587 * G + 0.114 * B$

» $Cb = 128 - 0.168736 * R - 0.331264 * G + 0.5 * B$

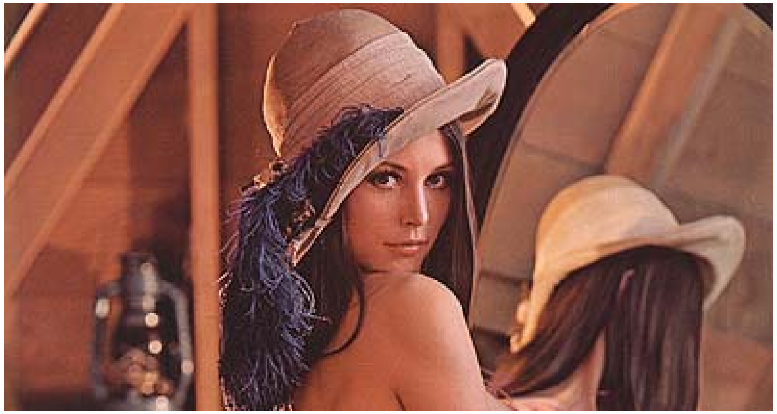
» $Cr = 128 + 0.5 * R - 0.418688 * G - 0.081312 * B$

Advantages

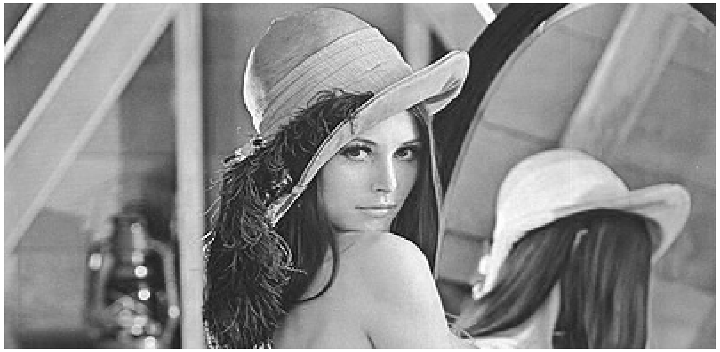
- **Chrominance characterized by Cb,Cr.**
- Suited for image Compression (JPEG, Selea Camera Coding).
- Linear combination of “familiar” RGB.

Color Models

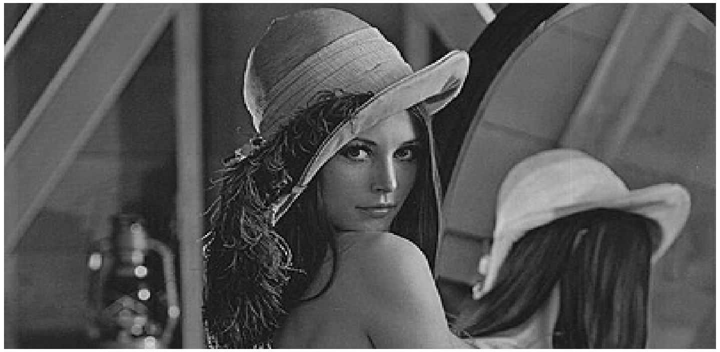
RGB Image



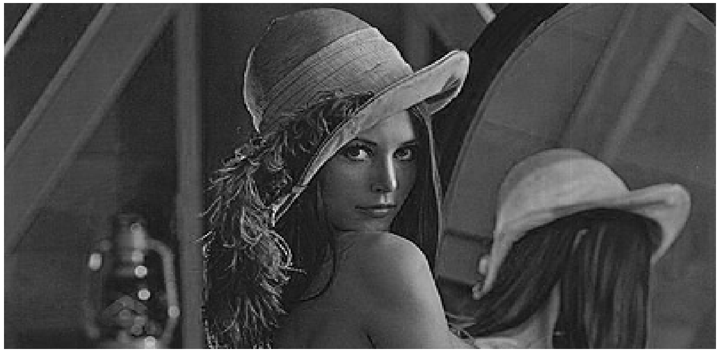
R



G

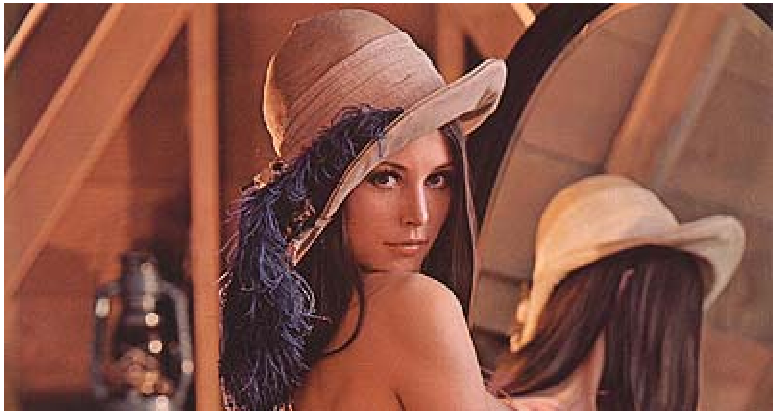


B

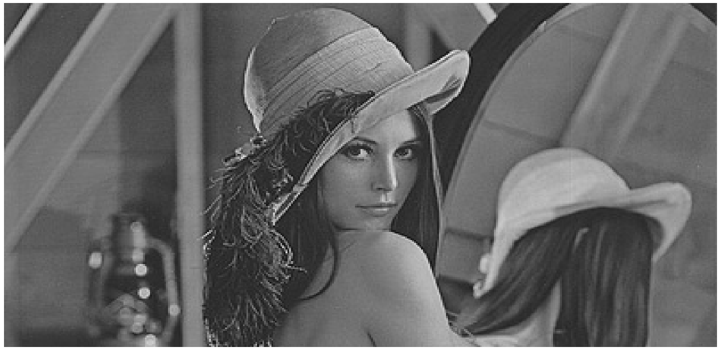


Color Models

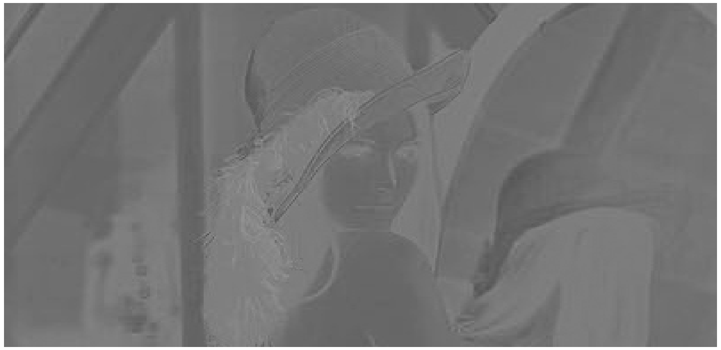
RGB Image



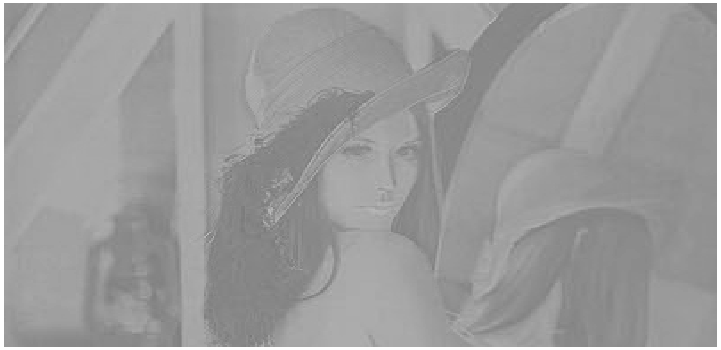
Y



Cb



Cr



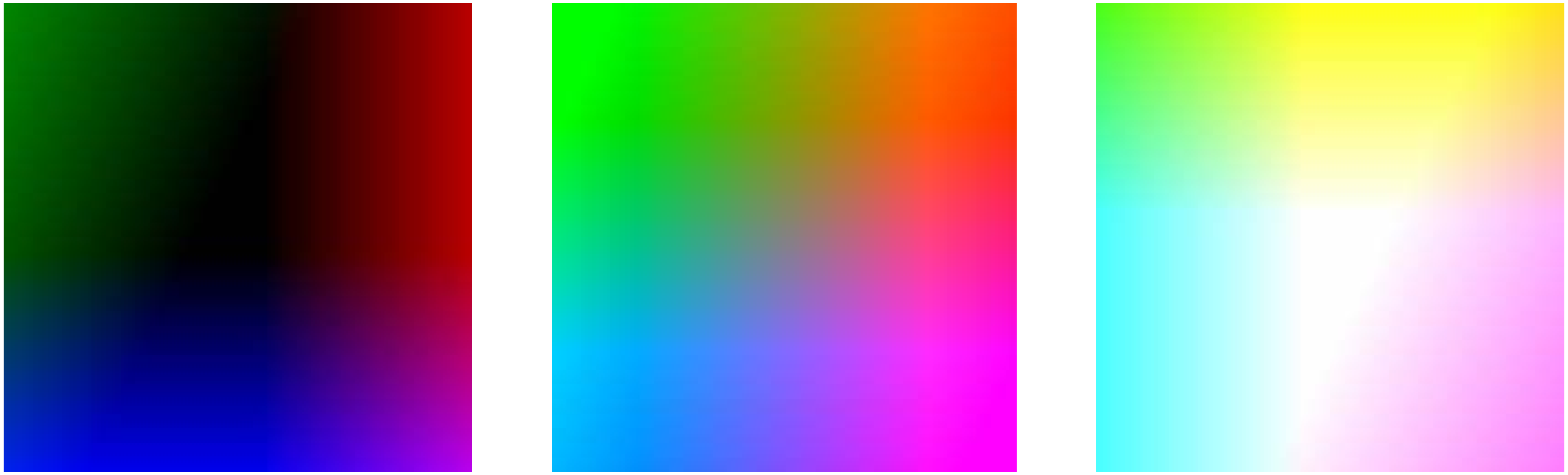
Human eyes are less sensitive to chromatic change than luminance change.

JPEG uses rough approximation in Cb Cr channels

Skin Detector

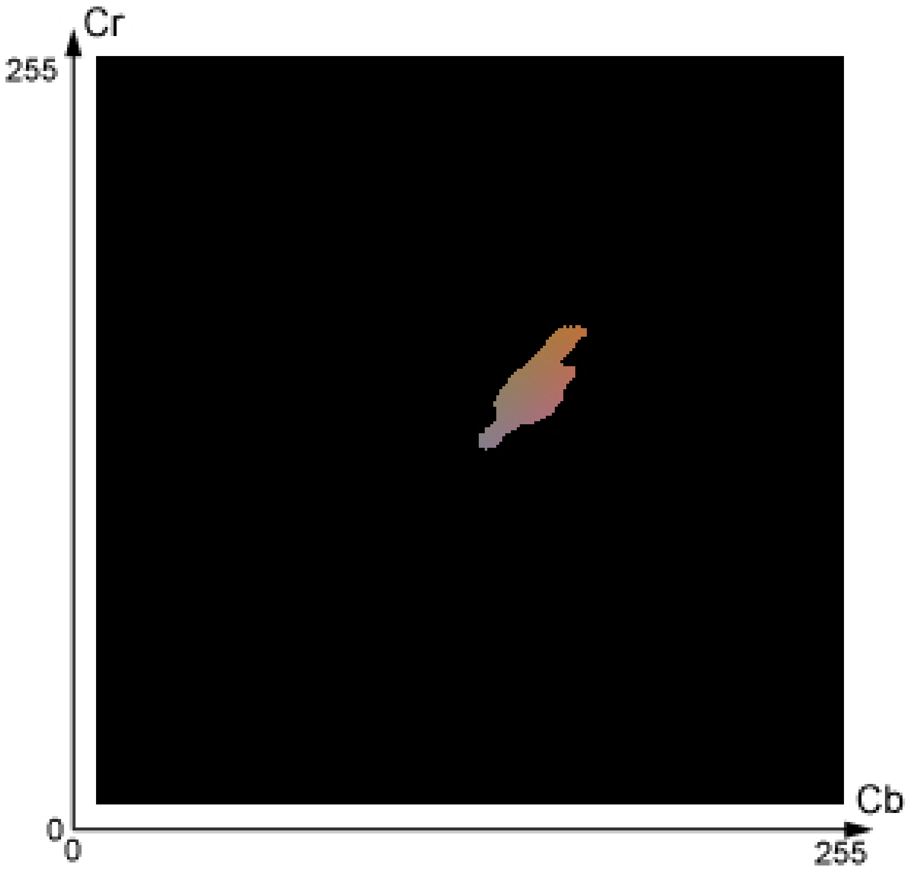
- Skin pixels can be identified by Cb-Cr components.
- Skin region can be draw in Cb-Cr plane.
- Previous models [] [] didn’t work properly on SELEA camera, we collected some skin images to build our own lookup table.

- Cb Cr plane for Y=0, Y=0.5, Y=1

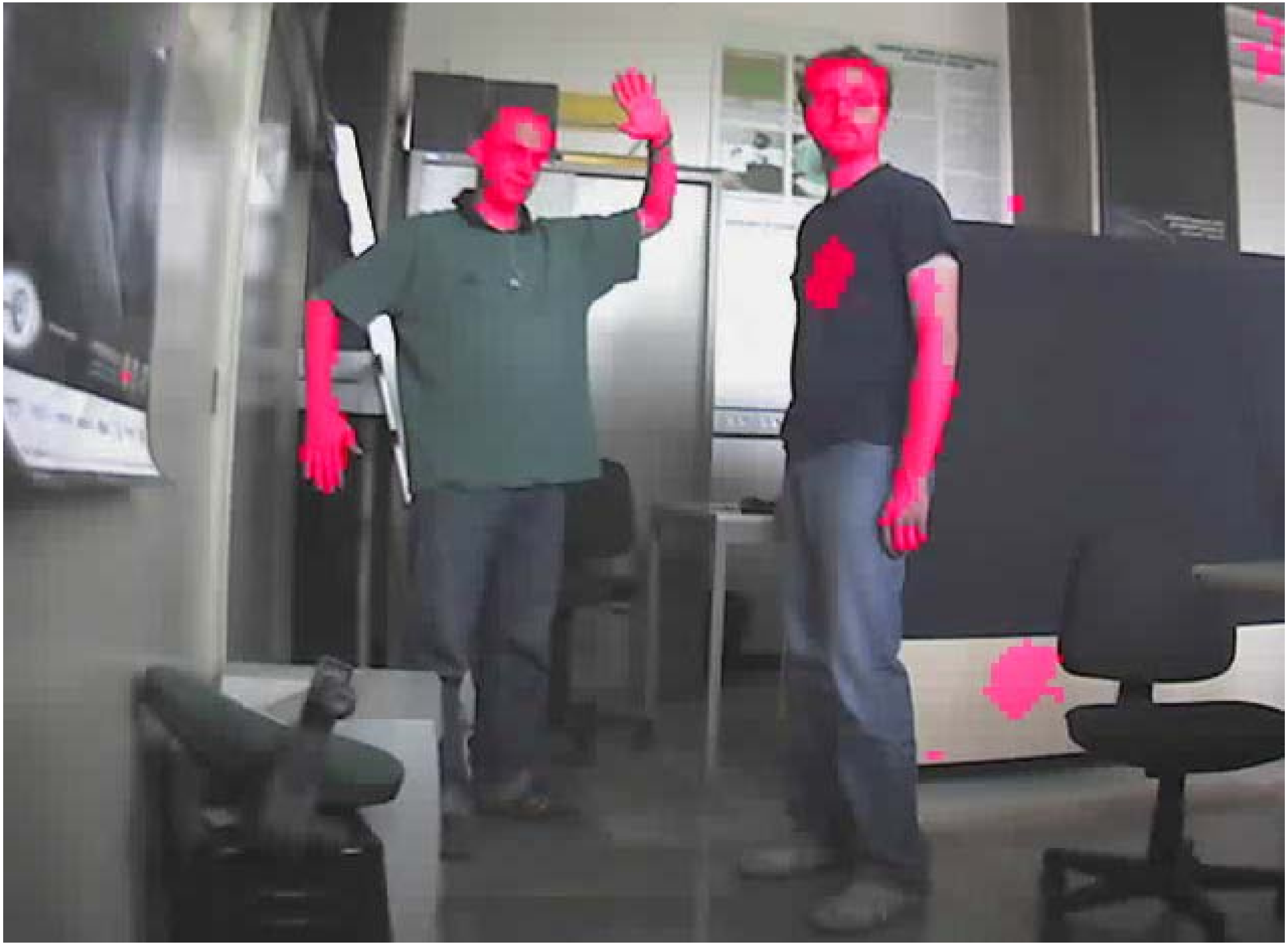


Skin Detection

- Skin Region in Cb Cr plane



Skin Detection Results



Skin Detection Problems

Noisy spots in Cb and Cr images are considered skin both in foreground and background

- Low pass
- Downsampling (160x120 image)
- Tresholding
- Weighted Ordered Statistic Filtering (erode, median, orderd filtering)

Some still objects in the background considered as skin

- Background subtraction

Background Acquisition

- Background Acquisition for Y, Cb, Cr channels
  - Average on N (~30) consecutive frames, separately

$$B(x) = \frac{1}{N}\sum_{i=0}^{N} f_i(x)$$

  - Noise is assumed Additive Gaussian White
  - Nose std is computed from differences between consecutive frames

$$var_i = 1/2(\sum_x (f_i(x) - f_{i-1}(x))^2) \qquad std = \frac{1}{29}(\sum_i \sqrt{var_i})$$

  - Noise std is used to build background confidence intervals

Background Suppression

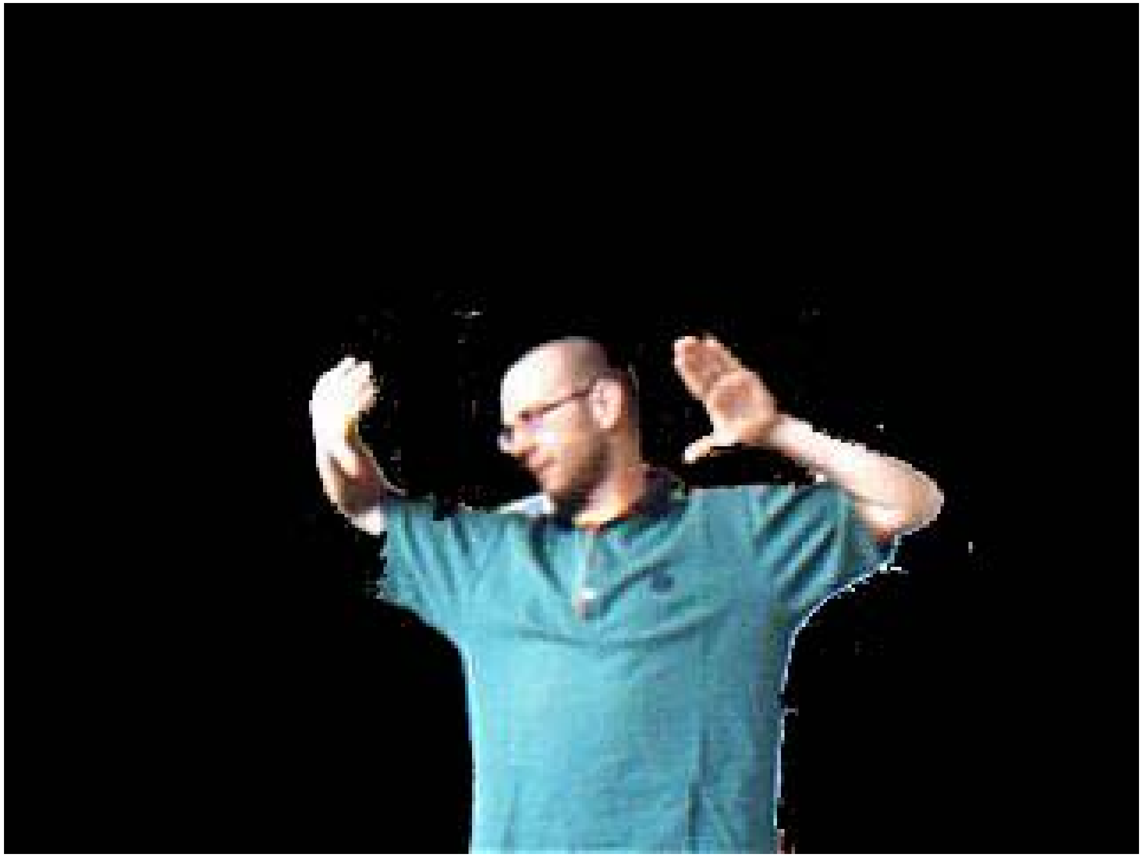
Given a new frame, separately on each channel , every pixel is compared with background.

$$|f_i(x) - B(x)| > T^{\cdot} std$$

T depends on channel considered

- Higher on Y to increase robustness to illumination changes

Background Suppression Result



Background Subtraction Problems

Again, Noise produces isolated foreground pixels

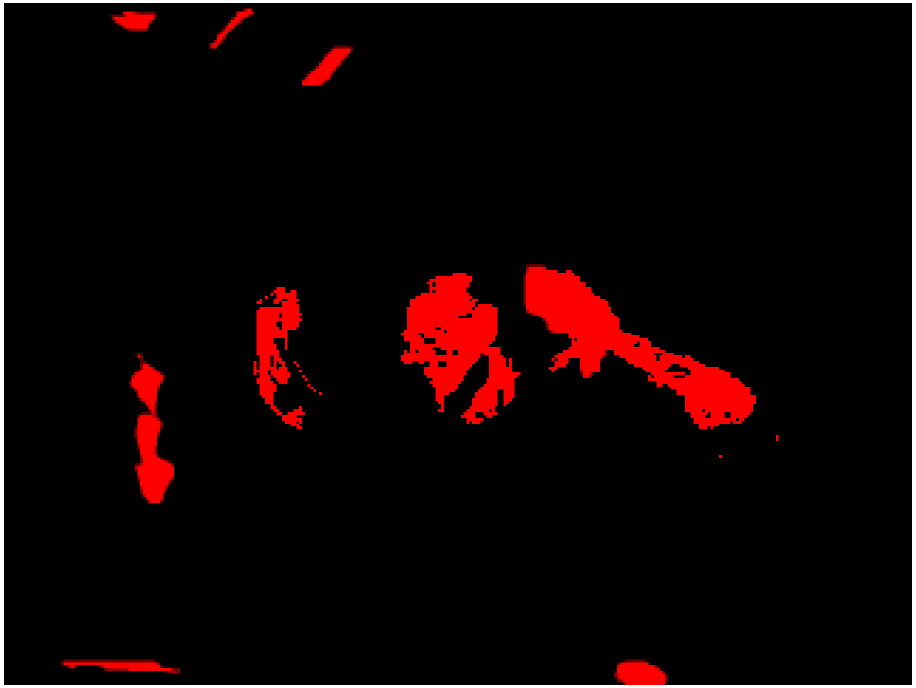
Sometimes background pixels appears in foreground area

Solution:

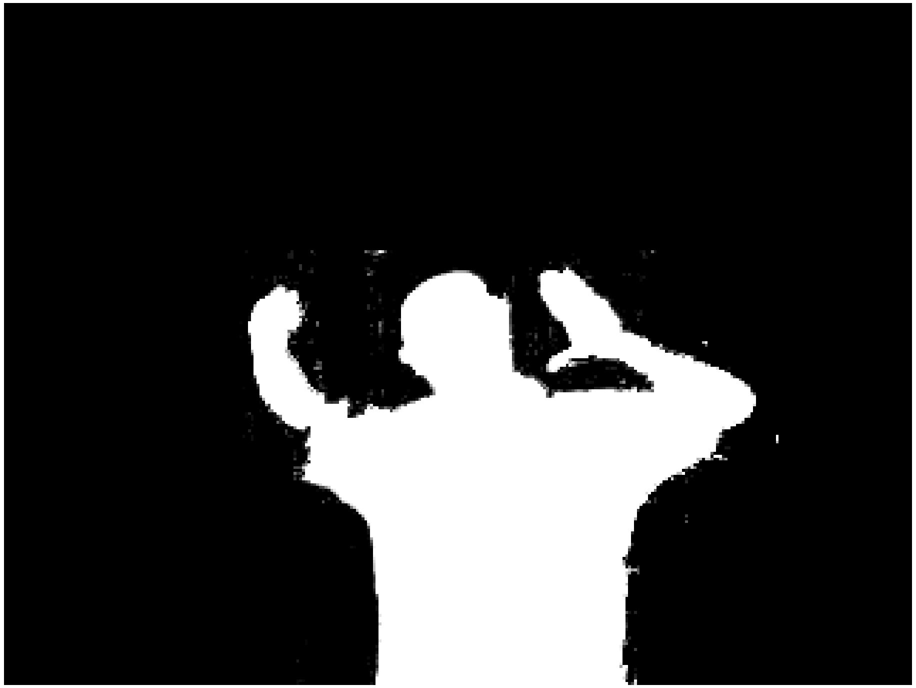
- Low pass
- Downsampling (160x120 image)
- Threshold
- Morphological image Processing

Intersection of maps

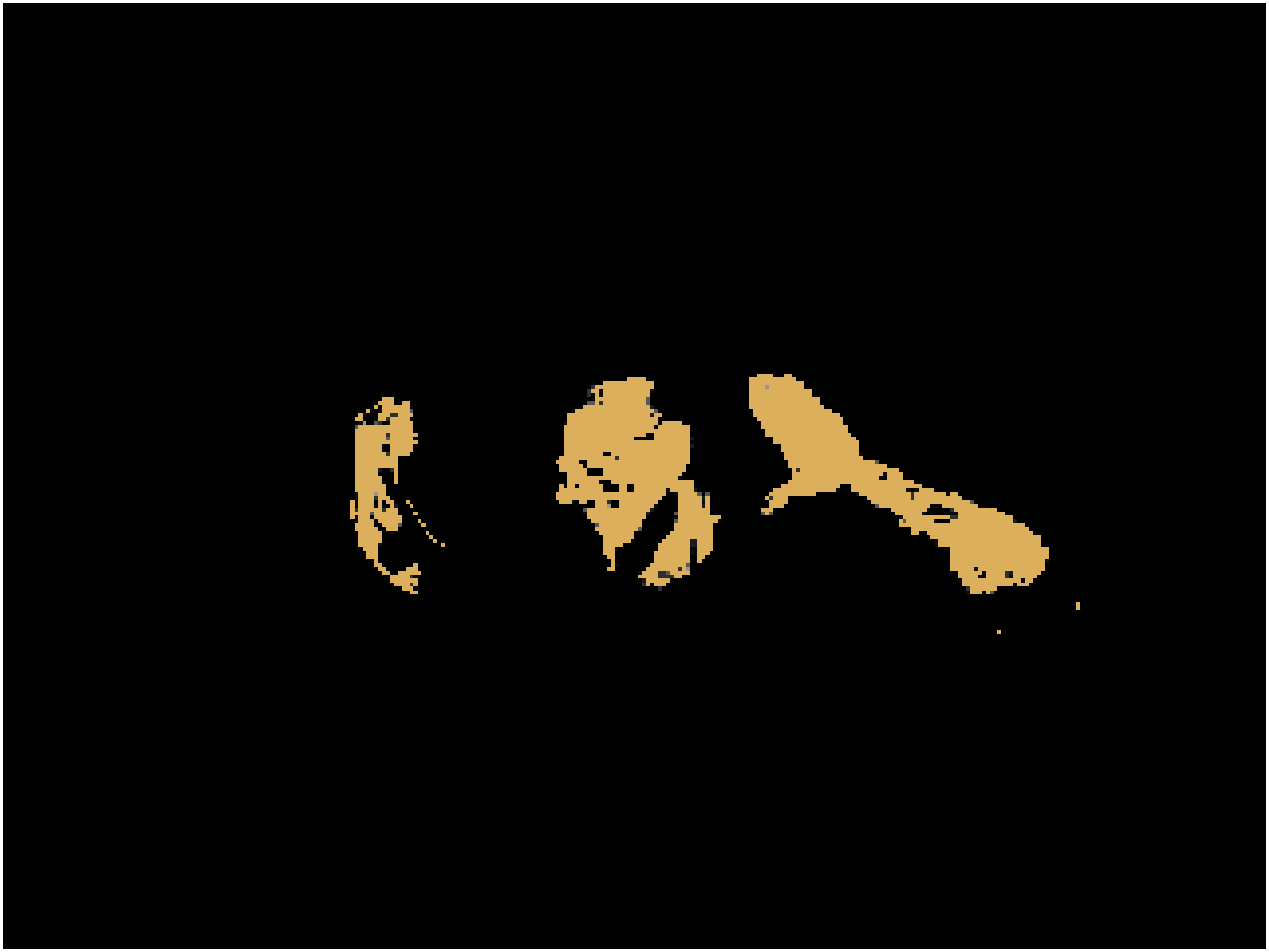
Skin



Foreground

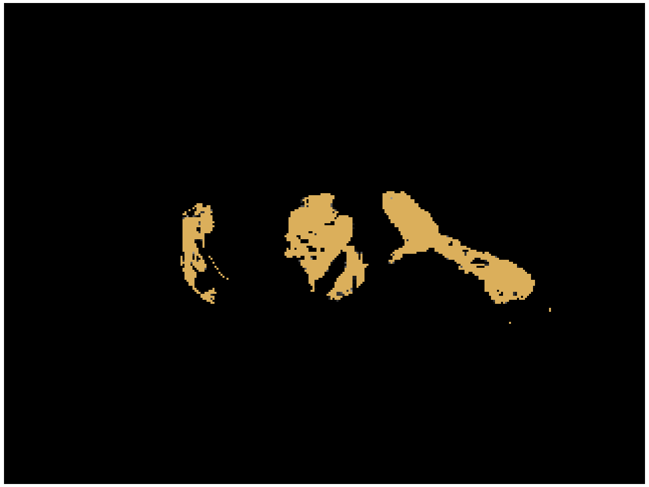


Intersection



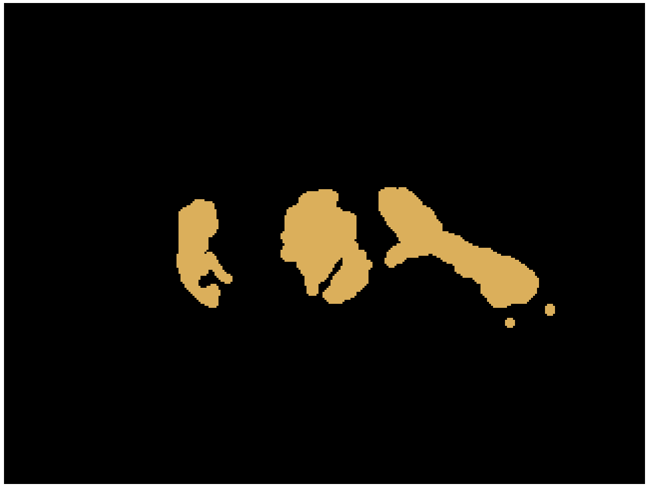
Morphological Processing

Morphological Filtering is run on black and white maps



Skin Pixel map (160x120)

Foreground Pixel map (160x120)



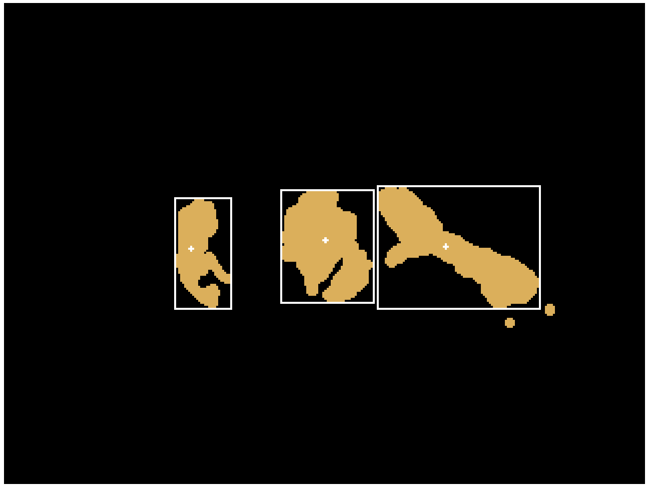
Weighted Ordered Statistic filtering

On 3x3 pixel mask, sort pixel values.

- "Erosion" get minimum
- Median
- "Dilate" get maximum value

Flooding Algorithm

Algorithm pseudo-code



Blobs are 4 way connected

For each blob we compute

- Area
- Centroid
- Bounding Box

Blob Metric Analysis

We are looking for triplets of blobs representing
Left Hand - Head - Right Hand in alarm position

- Discard blob having area smaller than Threshold (30 pixels)

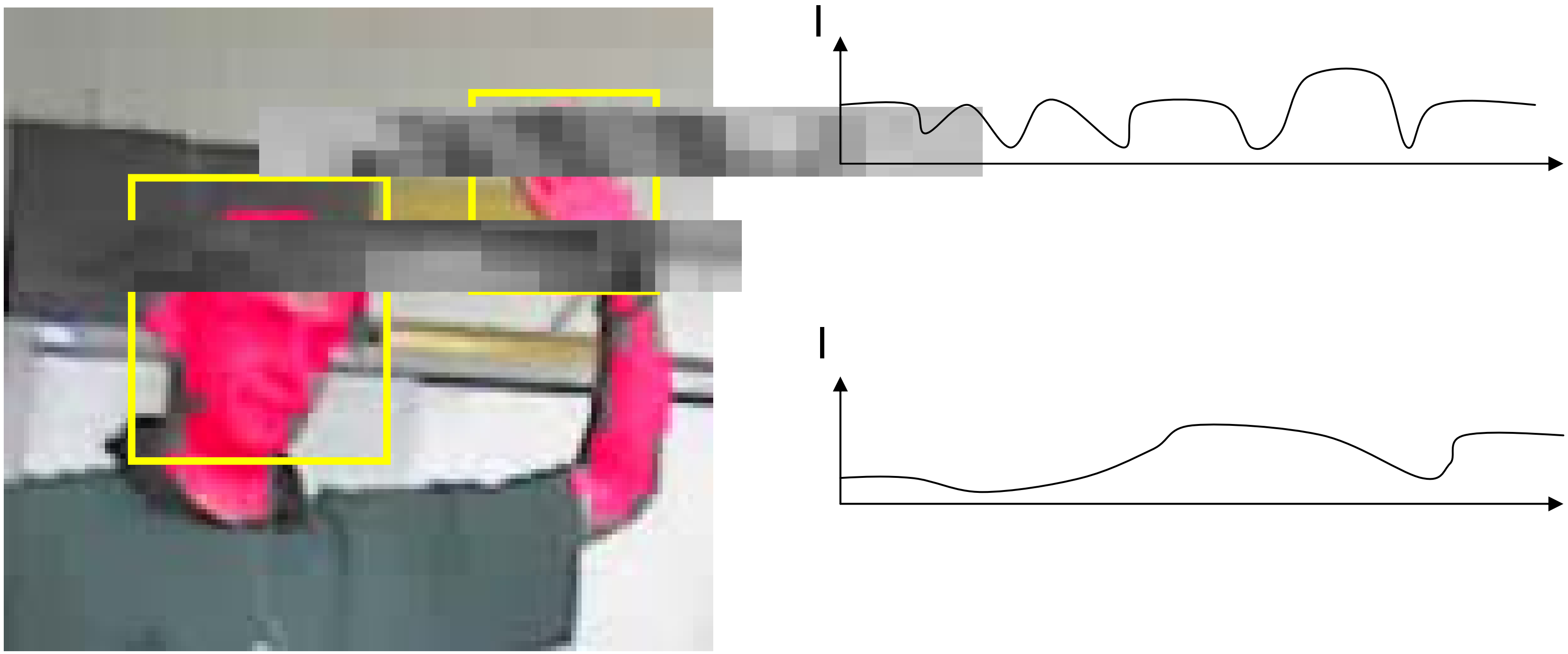
Blob Metric Analysis

For each blob, check if it can be an head

- Check bonding box ratio ~ 1,6 (golden ratio)
- Blob Area ~ ¾ bounding box area
- Look for two hands candidate
  - Relative Position Analysis
  - Area Ratio Analysis
- Only for blob triplets satisfying LH-H-RH detection are run.

Hand Detector

- Hands are not correlation friendly because of fingers
  - Very sensible to rotation and scale
- Approach: analyze luminance (double resolution than CbCr) along given horizontal profile (1/5 blob height from top) → forehead or fingers
  - Compute 1D gradient (first derivative) and count peaks



Face Detector, Normalized Cross Correlation

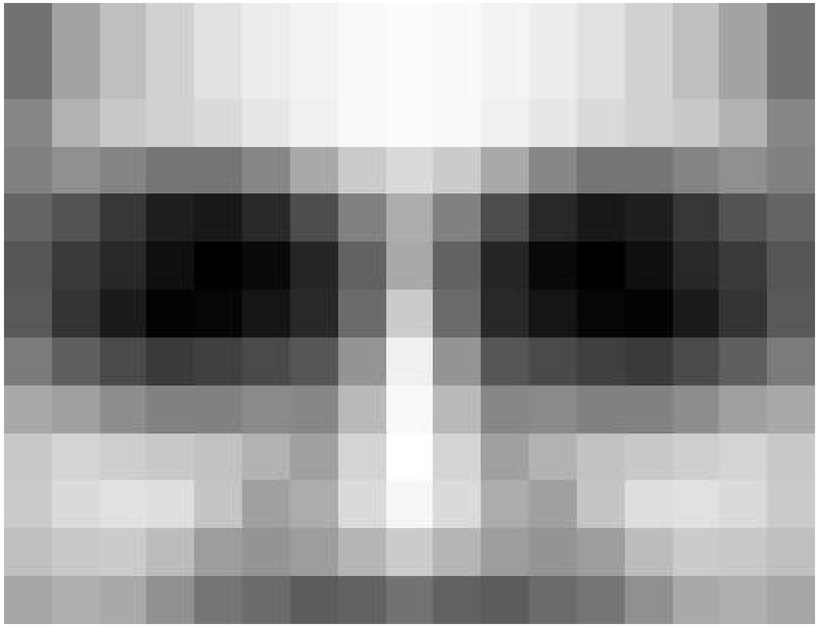
Measure distance between image *f* in pixel *(u,v)* and template *t*.

$$\Gamma(u,v) = \frac{\sum_{x,y}[f(x,y)-\bar{f}_{u,v}][t(x-u,y-v)-\bar{t}]}{\{\sum_{x,y}[f(x,y)-\bar{f}_{u,v}]^2 \sum_{x,y}[t(x-u,y-v)-\bar{t}]^2\}^{0.5}}$$

- Gives a measure between [-1,1]
- Independent on Image Intensity
- Fast Implementation using running sum to compute local averages

Face Detector, Template

- Normalized template used to match faces
- 19x11 mask, average of 100 faces from a training dataset



Alarm Trigger

- There is a global alarm level which over time tends to zero.
- Each alarm frame increases such level.
- When Alarm level rises over a given threshold, Police is coming!

Implementation details

- In one shot only
  - Low pass
  - Downsampling (160x240 image)
  - Threshold
- WOS are done without sorting elements (binaries entry).
- NCC is implemented using running sums

Improvements and Problems

- Background updating procedure
- Skin region/look up table
- Adaptive threshold
- Improve Hand Detector Robustness
- Change Face Template
- Change Correlation order in Face Matching

- Camera Automatic White Balance causes problem in skin detection