Pose and Lighting Invariant Face Detection

Vijayan Asari
Old Dominion University
Department of Electrical and Computer Engineering
Norfolk, Virginia 23529.
Face Detection

- Images at various lighting conditions are enhanced to a uniform lighting environment.
- In order to reduce the search space for faces in an image frame, the human skin regions are extracted using the color information.
- Skin regions are classified into faces and non-faces using feature matrices developed during a training phase.
Nonlinear Enhancement

- Dynamic range compression by nonlinear transformation
- Need to consider the information of the surrounding pixels in each pixel processing
  - Filtering with multi-level Gaussian function
- Transformation to display domain
  - Gain (contrast) and offset (brightness) compensation
- Prevention of luminance dropping
- Color restoration and color saturation adjustment

Input images

Enhanced images

Logarithmic transformation

Multi-level Gaussian function
Adaptive Technique for Nonlinear Enhancement of Color Images

**Luminance Enhancement**

$I(x, y)$ is the gray scale image

$$I(x, y) = R \times W_r + G \times W_g + B \times W_b$$

$$W_r = \frac{R}{R+G+B} \quad W_g = \frac{G}{R+G+B} \quad W_b = \frac{B}{R+G+B}$$

$I_n(x, y)$ is the normalized gray level

$$I_n(x, y) = \frac{I(x, y)}{255}$$

Non-linear transfer function

$$I'_{n} = \left( \frac{I_n^{0.75}(x + 0.25)}{2} + 0.6 \left( 1 - I_n \right) \left( 1 - x \right) + I_n^{2-x} \right)$$

(Curve 6, when $x = 0$)
The non-linear transfer function is image dependent with a parameter $x$, which is related to the image histogram and is defined as:

$$x = \begin{cases} 
0 & \text{for } L \leq 50 \\
\frac{L-50}{100} & \text{for } 50 < L < 150 \\
1 & \text{for } L \geq 150 
\end{cases}$$

As $x$ approaches 0, we get the top curve. As $x$ approaches 1, the transfer function curve gets closer to the identity transformation.
Image Enhancement by AINDANE

Original image

Enhanced image
Image Enhancement by AINDANE

Original image

Enhanced image
Image Enhancement by AINDANE

Original image

Enhanced image
Image Enhancement by AINDANE

Original image

Enhanced image
Image Enhancement by AINDANE

Original image

Enhanced image
Enhancement of Video Stream

- Surveillance camera – Sony EVI-D30, frame size is 360×240 pixels.
- Simulation environment: Xeon (3.06GHz) Desktop, RAM size of 1GB in Windows-XP platform, language C++.
- Processing rate is 18 frames/sec.
Skin Region Extraction

- Skin patches were collected by manually cropping from images.
- A 2D skin distribution map (SDM) of dimension $256 \times 256$ representing the sparse network is trained using the Winnow update rule with skin pixels to learn their distribution in Cb-Cr color space.
- False positives in the initial skin detection results are used to update the SDM to “unlearn” the non-skins.
- After training the SDM, human skin regions are extracted automatically in any arbitrary input image.
Skin Segmentation Results

Original image

Enhanced image

Detected faces

Skin segmented image

Vision and VLSI Systems Laboratory, Department of Electrical and Computer Engineering, Old Dominion University
Face Detection using SNoW

- **Sparse Network of Winnows:**
  A learning architecture of linear units over a common feature space (pixel coordinate vs. intensity value).
  The architecture learns the most likely intensity of each pixel for faces and non-faces during the training process.

- Encode the image into a set of active features whose index is,

  \[ [(y \times w + x), I(x, y)] \]

  image width  pixel intensity

- Train using faces and non-faces and obtain two weight matrices.
Update Rule for Weights

- **Update rule for weight matrix generation.**

<table>
<thead>
<tr>
<th>Prediction</th>
<th>Correct response</th>
<th>Update action</th>
<th>Update step</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>$w_i = \beta w_i \text{ for } 0 &lt; \beta &lt; 1$</td>
<td>demotion step</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>$w_i = \alpha w_i \text{ for } \alpha &gt; 1$</td>
<td>promotion step</td>
</tr>
</tbody>
</table>

- Same update rule is followed for both face and non-face weight matrices.
- The activation function at the output of each node is just the linear sum of the active weights in the example.
- Winner takes all.
- The face weight matrix learns the most likely intensities, but most of the weights are close to zero.
Downsampling and Classification

Downsampling:
- Each 19 X 19 block in the input image and its down-sampled versions are histogram-equalized and provided for classification.

Classification:
- Classifiers are trained to classify the incoming probe image irrespective of a face or a non-face into appropriate pose and orientation.
- Final step is to determine if the block is a face or a non-face using the trained network for a particular orientation.
An architecture which retains the accuracy of individual classification but reduces the total number of classifiers.

Classifiers are trained to distinguish faces with maximum disparity in orientation to reduce the risk of misinterpretation.

10 classifiers would be enough to classify all the face orientations and poses.

Face profile information will be obtained at the final stage of classification.

1. Up and upside-down classification
2. Left rotated against right rotated
3. Frontal or pose
4. Further classification to reduce the angle disparity
Test Results
Face Detection Results on FRGC Database

Experiment 1: Images With Controlled Indoor Lighting.

<table>
<thead>
<tr>
<th>Face Detection</th>
<th>Detection Rate</th>
<th>FRR</th>
<th># Non-faces</th>
<th>Automatic</th>
</tr>
</thead>
<tbody>
<tr>
<td>FD</td>
<td>100%</td>
<td>0%</td>
<td>61</td>
<td>Yes</td>
</tr>
<tr>
<td>SD + FD</td>
<td>100%</td>
<td>0%</td>
<td>53</td>
<td>Yes</td>
</tr>
<tr>
<td>En + FD</td>
<td>100%</td>
<td>0%</td>
<td>59</td>
<td>Yes</td>
</tr>
</tbody>
</table>

FD = Face Detection, SD = Skin Detection, En = Enhancement, FRR = False Rejection Rate
Face Detection Results on FRGC Database

Experiment 4 : Images With Uncontrolled Indoor Lighting.

<table>
<thead>
<tr>
<th>Face Detection</th>
<th>Detection Rate</th>
<th>FRR</th>
<th># Non-faces</th>
<th>Automatic</th>
</tr>
</thead>
<tbody>
<tr>
<td>FD</td>
<td>95.7%</td>
<td>4.3% (41)</td>
<td>94</td>
<td>Yes</td>
</tr>
<tr>
<td>SD + FD</td>
<td>95.1%</td>
<td>4.9% (46)</td>
<td>35</td>
<td>Yes</td>
</tr>
<tr>
<td>En + FD</td>
<td>97.7%</td>
<td>2.3% (22)</td>
<td>77</td>
<td>Yes</td>
</tr>
</tbody>
</table>

FD=Face Detection, SD=Skin Detection, En=Enhancement
FRR=False Rejection Rate
Difficult to Detect Images From FRGC Database

Original (2272 x 1704)  En +FD
Difficult to Detect Images From FRGC Database

Original (2272 x 1704)  En +FD
Difficult to Detect Images From FRGC Database

Original (2272 x 1704)  En +FD
Real Time Face Detection

Frame Size: 320 X 240 Pixels

1.48GHz CPU, 1.5GB RAM
Conclusion

- Nonlinear enhancement provides a uniform lighting environment for all images.
- Face detection after skin region segmentation decreases the number of non-faces.
- Face detection after enhancement decreases the FRR.
- Research is in progress to perform skin specific enhancement, skin region extraction and face detection to decrease the number of non-faces and FRR.