Rapid Stereo-Vision
Enhanced Face Detection

Sergey Kosov, Kristina Scherbaum, Kamil Faber,
Thorsten Thormählen, Hans-Peter Seidel

Face detection often serves as a first step in image processing applications, like, face recognition, visual surveillance, or human-machine interaction. Our approach improves state-of-the-art 2D detection techniques, by additionally evaluating a disparity map, which is estimated for the face region.

1. First, faces are detected in the 2D images with a rapid object classifier based on haar-like features.
2. In a second step, we train a PCA-classifier (off-line) that helps removing falsely detected faces by analyzing the disparity map.
3. In the near field of the camera, this classifier is used, which evaluates the Eigenfaces of the normalized disparity map. Thereby, the transformation into Eigenspace is learned off-line using a principal component analysis approach.
4. In the far field, a much simpler approach determines false-positives by evaluating the relationship between the size of the face in the image and its distance to the camera.
5. This novel combination of algorithms runs in real-time and significantly reduces the number of false-positives compared to classical 2D face detection approaches.

Stereo Setup and Algorithm
- Two cameras with a resolution of 384 x 288 pixels
- Near-field stereo case: 20 cm baseline distance and 9.5° convergence-angle, far-field stereo case: 4 cm baseline distance and 0° convergence-angle
- Off-line calibration and rectification of input images
- Estimation of disparity map (variational approach, including regularization)

Example of a disparity map estimation:

2D Face Detection
- Training of a boosted classifier cascade, which uses rotated haar-like features, using discrete AdaBoost
- Input: 4700 positive samples, 3300 negative samples
- We apply the classifier on the left image of all stereo pairs including potential face candidates

Receiver Operator Characteristic of the classifier

Evaluating The Disparity Map
- We add a classification step on the disparity maps
- And use a PCA-classifier that is trained off-line on 30 facial regions

Stereo-image pairs and generated disparity maps that where used in our training set of the PCA

Near Field: 3D-Eigenfaces
- The PCA estimates the probability distribution of facial disparity maps around their average

Average face, and the first 3 Eigenfaces (left to right)

- Online detection: we project disparity map regions of potential candidates (as detected) into PCA space
- We calculate the Mahalanobis distance of current test signals to the average face
- By a threshold of $\sigma = 3.29$ (in terms of standard deviation), we accept all face candidates lying within the range of ~99.9% of all training samples
- Face candidates that lie above this threshold will be rejected as non-faces

Far Field: Evaluating Size-Relation
- In small image patches, the above algorithm fails because a reliable disparity map cannot be estimated
- If the patch size of a face is smaller than 40x60 pixels we estimate the disparity to the right frame with the Kanade-Lucas-Tomasi approach and calculate the median disparity and the corresponding distance to the camera
- We can determine the actual size of the face in 3D and check it against an interval of sizes for veridical faces, as learned from a database of 3D face scans
- If the face size lies not in a range from 15.78 to 24.16 cm, the face is marked as a false-positive

Experimental Results
- Comparison of the detection rates of the monocular detection approach with our stereo enhanced method
- The two shaded columns show that the number of false-positives (FP) decreases using our method, and the number of true-positive (TP) and false-negative (FP) stays the same

<table>
<thead>
<tr>
<th></th>
<th>40 Faces</th>
<th>19 Non-Faces</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FP</td>
<td>TP</td>
</tr>
<tr>
<td>monocular</td>
<td>7</td>
<td>38</td>
</tr>
<tr>
<td>stereo enhanced</td>
<td>0</td>
<td>38</td>
</tr>
</tbody>
</table>

Timings for an Intel® Core™i7 2600 CPU with 2.70 GHz

<table>
<thead>
<tr>
<th>Step</th>
<th>FF [msec]</th>
<th>NF [msec]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Run monocular detector</td>
<td>27</td>
<td>27</td>
</tr>
<tr>
<td>Estimate disparity map</td>
<td>-</td>
<td>78</td>
</tr>
<tr>
<td>Transform into Eigenspace</td>
<td>-</td>
<td>42</td>
</tr>
<tr>
<td>Estimate sparse disparity map</td>
<td>63</td>
<td>-</td>
</tr>
<tr>
<td>Total</td>
<td>90</td>
<td>147</td>
</tr>
</tbody>
</table>