Eye Detection Using Haar Features and Cascaded SVM

Vishal Agrawal
(Y6541)

Problem Statement

- Given a grayscale or the color image of a person find the center of eyes as accurately as possible

Approach

- Features:
  - Properly selected Over-complete Haar wavelet coefficients
- Training:
  - Two SVMs (SVM1 & SVM2) in cascade, trained on the above selected coefficients

SVM1 Training Set

- Positive set:
  - Eye images, centered at the center of eye and cropped at the size of inter-ocular distance
- Negative set:
  - Other facial features such as nose, mouth, chin etc. and some other background images (8-11 negatives for each positive)

SVM1 Training Set

- Positive Set

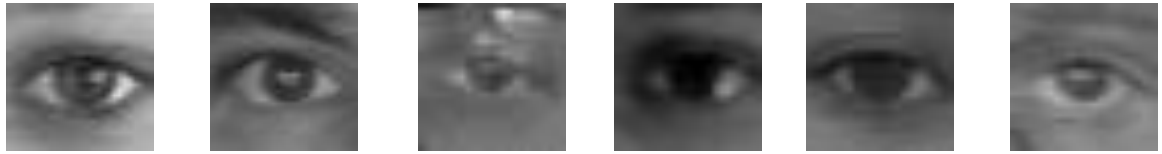
- Negative Set

SVM2 Training Set

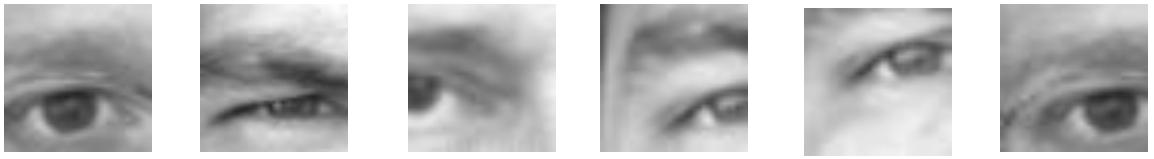
- Positive set:
  - Eye images same as in the SVM1, but this time the cropping is done at half of the size of image in SVM1
- Negative set:
  - Generated by the small random displacement of the sub-images used for the extraction of the positive examples (10 negatives for each positive)

SVM2 Training Set

- Positive Set



- Negative Set



Feature Extraction

- Multi-level over-complete Haar wavelet transform

- Total 852 wavelet coefficients per image

Feature Extraction

- A wavelet coefficient $d^{o}_{j,k1,k2}$ is identified by 4 parameters
  - $o \in \{horizontal, vertical, diagonal\}$
  - *j* is the detail level
  - *(k1, k2)* is the position of the coefficient within the image
- Set of all the wavelet coefficients of a certain level *j* is called the band of level *j(*B_j*)*

Feature Selection

- Relative importance of the coefficients

$$\overline{d^o}_{j,k_1,k_2} = \frac{\sum_{l=1}^{|L|} |d^o_{j,k_1,k_2}(l)|}{|L|}$$

$$\tilde{d^o}_{j,k_1,k_2} = \frac{\overline{d^o}_{j,k_1,k_2}}{m_j}, \quad \text{where } m_j = \frac{\sum_{k_1}\sum_{k_2}\sum_o \overline{d^o}_{j,k_1,k_2}}{|B_j|}$$

Feature Selection

- Expected value of the sum of all $d^{o}_{j,k1,k2}$ in one band is –

$$E\left[\sum_{k_1}\sum_{k_2}\sum_{o}\tilde{d}^{o}{}_{j,k_1,k_2}\right]=\sum_{k_1}\sum_{k_2}\sum_{o}E\left[\frac{\overline{d^{o}}{}_{j,k_1,k_2}}{m_j}\right]\approx\frac{\sum_{k_1}\sum_{k_2}\sum_{o}E\left[\overline{d^{o}}{}_{j,k_1,k_2}\right]\cdot|B_j|}{E\left[\sum_{k_1}\sum_{k_2}\sum_{o}\overline{d^{o}}{}_{j,k_1,k_2}\right]}=|B_j|$$

- Therefore –

$$\tilde{d}^{o}{}_{j,k_1,k_2}\begin{cases}\sim 1 & \Rightarrow \text{no regularity}\\ \ll 1 & \Rightarrow \text{systematic uniformity} \quad (C^{-})\\ \gg 1 & \Rightarrow \text{systematic variation} \quad (C^{+})\end{cases}$$

Feature Selection

- Error function to drive the selection process –

$$w = \arg \min_{\substack{w = w^+ \cup w^-, \\ w^+ \subseteq C^+,\ w^- \subseteq C^-}} ||E - E_w||^2 + \alpha \cdot ||E_w - U||^2$$

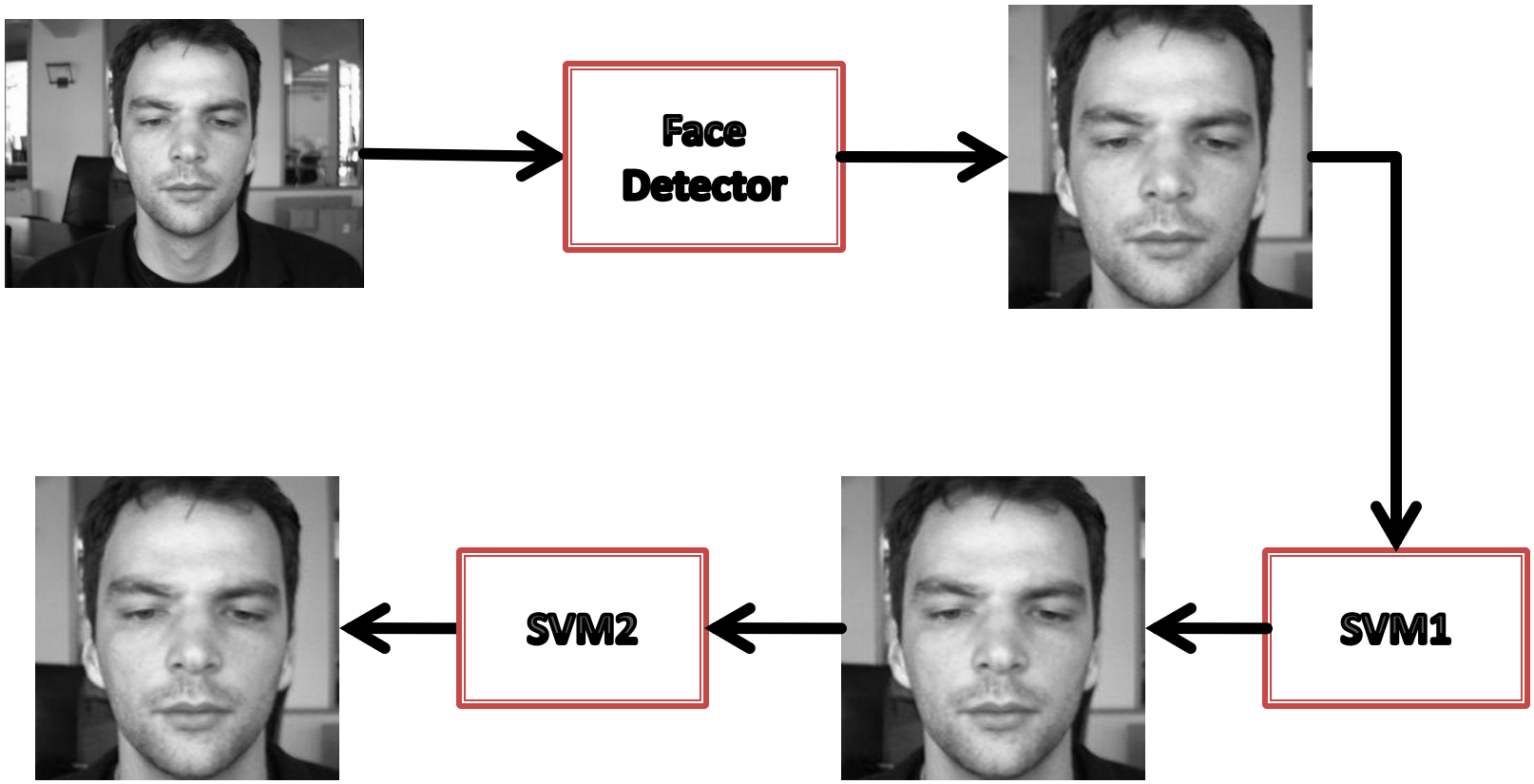
  - E_w is the reconstructed pattern
  - E is Mean eye pattern
  - U is the uniform pattern
  - α ε (0,1) is a constant

- α should be such that the $\frac{|w^-|}{|C^-|} \approx \frac{|w^+|}{|C^+|}$

Cascaded SVM Training

- SVM1 should be general and robust
- First level wavelet coefficients are crucial for the precision
- Therefore first level wavelet coefficients are not used for the SVM1 training

Detection Technique



Challenges

- Scale given by the face detector may not be correct

- Search eyes in the face at three different scale to handle the first uncertainty
  - Inferred scale (X_P)
  - Small underestimation of the scale (X_P^-)
  - Small overestimation of the scale (X_P^+)

Detection Technique

- Evaluate the function –

$$\rho(P) = SVM1(X_P) + SVM1(X_P^-) + SVM1(X_P^+)$$

- Point for which ρ(P) > 0 consider them as point candidates
- Group them based on their proximity in the image
- Each group of point candidates is then represented by its centroid

Detection Technique

- Ideally there should be only two point candidates after the grouping
- If more then 2 points than it can be handled using the verticality of the face –

$$\{C_a, C_b\} = \max \frac{SVM1(C_i) * SVM1(C_j)}{|C_{iy} - C_{jy}|}$$

- Here C_i *and* C_{iy} *are* center and the y coordinate of the i^{th} point candidate

Detection Technique

- Apply SVM2 in the small neighborhood of the found positions to refine results

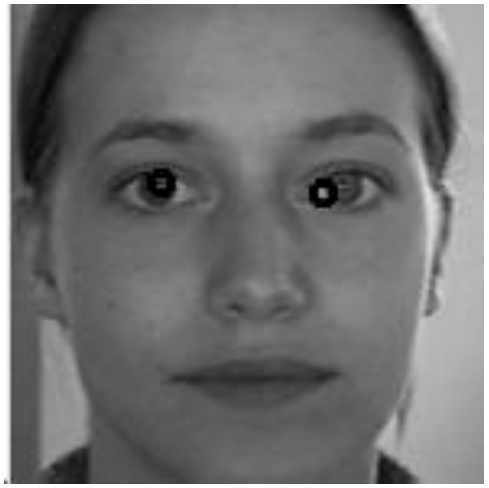
- Scale considered by the SVM-2 is –

$$\frac{1}{2} \times \frac{\sum_{\mathbf{x} \in \{\mathbf{x}_P, \mathbf{x}_P^+, \mathbf{x}_P^-\}} [\Theta(SVM_1(\mathbf{x})) \times scale(\mathbf{x})]}{3} \quad \text{where} \quad \Theta(z) = \begin{cases} z & \text{if } z > 0 \\ 0 & \text{if } z \leq 0 \end{cases}$$

- Search for the candidates that gives the highest response in SVM2

Results





Results



References

- Campadelli P., Lanzarotti R., Lipori G., and Di Milano U. *Precise eye localization through a general-to-specific model definition. In Proceedings of the British Machine Vision Conference* (2006), vol. 1, pp. 187 - 196