Facial landmark detector learned by structured output SVM

Kostiantyn Antoniuk

Augmented Pixels co

July 13, 2017
Outline for section 1

1. TODO
2. Preprocessing
3. Structured output classifier design
Motivation

Face registration

- Essential part of face recognition systems (identity, gender, age, etc)
- Quality of face recognition and further processing depends on the quality of registration.
- Lack of solutions working on a wide yaw range views (i.e. up to profiles).
Goal

- Develop facial landmark detector invariant to scale and rotation.
Motivation: Structured output classifier

- Peaks of patch score trained independently per each landmark coincide with true landmark positions, however there are also peaks in other locations due to local similarity.
- The ambiguity is resolved by the geometric model!
Outline for section 2

1 TODO

2 Preprocessing

3 Structured output classifier design
Face normalization

- Blue box is a detection as provided by the face detector.
- Red box is the detection box enlarged by a defined margin.
- The similarity transformation (removing the possible in-plane rotation and scaling the image to a fixed size) is applied on the red box and the normalized frame is obtained.
Landmark hypotheses

- Sizes of components typically determined experimentally (using dataset)
- The number of landmarks may be modified
- Though such asymmetric hypothesis for landmark do not perform well
We put additional hard constraints to make hypothesis windows vertically symmetric.

We can restrict landmarks positions due to its “natural” geometry, e.g., left eye has to be on the left part of the image, etc.
Outline for section 3

1. TODO

2. Preprocessing

3. Structured output classifier design
Appearence model

\[ q(s_i, \mathcal{I}, w^q_i) = \langle w^q_i, \Psi^q(s_i, \mathcal{I}) \rangle \]

- \( \Psi^q(s_i, \mathcal{I}) \) denotes a feature descriptor of a patch cropped from image \( \mathcal{I} \) around the position \( s_i = (x_i, y_i) \).
- Approach allows to use any feature descriptor
- We found out that pyramid local binary patterns (PLBP) work well
LBP features can be built on different surrounding windows and with different neighbourhood.

Though features described on the pic. give good performance when

- LBP are built on image pyramid (4 scales) and
- normalized in L2 norm.
Deformation Cost

\[ g_{ij}(s_i, s_j; w_{i,j}^g) = \langle w_{i,j}^g, \Psi_{ij}^g(s_i, s_j) \rangle \]

- \( \Psi_{ij}^g(s_i, s_j) \) is defined as quadratic function of the displacement vector, namely

\[ \Psi_{ij}^g(s_i, s_j) = (\delta x, \delta y, \delta x^2, \delta y^2) , \\
(\delta x, \delta y) = s_i - s_j = (x_i - x_j, y_i - y_j) \]
Loss function

\[ L(s, s') = \frac{\kappa}{M} \sum_{i=1}^{M} ||s_i - s'_i|| \]

To make the comparison between landmarks invariant to the changing scale of the detected faces, all measured deviations are normalized to the length of the line connecting the center of eyes with the mouth.
Modern detectors exploit relatively large graph instances!

- Graph must have “simple” structure (e.g., tree) if we want to use it in real-time!
- So the inference can be done very quickly!
Learning Landmark detector

\[ h(\mathcal{I}; \mathbf{w}) = \max_s \left\{ \sum_{i \in \mathcal{V}} q(s_i, \mathcal{I}, \mathbf{w}_i^q) + \sum_{i, j \in \mathcal{E}} g_{ij}(s_i, s_j; \mathbf{w}_{i,j}^g) \right\} \]

\[ \mathbf{w}^* = \arg\min_{\mathbf{w}} \frac{\lambda}{2} \| \mathbf{w} \|^2 + \frac{1}{n} \sum_{i=1}^{n} \mathcal{R}(\mathbf{w}; x^i, s^i), \]

\[ \mathcal{R}(\mathbf{w}; \mathcal{I}^i, s^i) = \max_{s \in \mathcal{Y}} \left\{ L(s, s^i) + \langle \mathbf{w}, \Psi(\mathcal{I}^i, s) \rangle \right\} - \langle \mathbf{w}, \Psi(\mathcal{I}^i, s^i) \rangle \]

Can be learned using Bundle method risk minimization [Uricar, et al.]
Making approach robust to face rotation.

- Discretize rotation angle, e.g. 
  \[0^\circ, 45^\circ\], \((45^\circ, 90^\circ)\], \((90^\circ, 135^\circ)\], \((135^\circ, 180^\circ)\]
- We can consider different graph structures per each angle range.
- Train model for each such graph.
- When making prediction, use each graph to predict and report answer when it is possible.
Learning landmark detector jointly with face rotation angle.

\[ h(I, \phi_1; w) = \max_s \left\{ \sum_{i \in V} q(s_i, I, w^\phi_{1q}) + \sum_{i,j \in E} g_{ij}^\phi(s_i, s_j; w^\phi_{1g}) \right\} \]

\[ h(I, \phi_2; w) = \max_s \left\{ \sum_{i \in V} q(s_i, I, w^\phi_{2q}) + \sum_{i,j \in E} g_{ij}^\phi(s_i, s_j; w^\phi_{2g}) \right\} \]

\[ h(I, \phi_3; w) = \max_s \left\{ \sum_{i \in V} q(s_i, I, w^\phi_{3q}) + \sum_{i,j \in E} g_{ij}^\phi(s_i, s_j; w^\phi_{3g}) \right\} \]

- How to combine into single classifier?
Learning landmark detector jointly with face rotation angle.

\[
\begin{align*}
    h(I, \phi_1; w) &= \max_s \left\{ \sum_{i \in V} q(s_i, I, w_i^{\phi_1 q}) + \sum_{i, j \in E} g_{ij}^{\phi}(s_i, s_j; w_{i,j}^{\phi_1 g}) \right\} \\
    h(I, \phi_2; w) &= \max_s \left\{ \sum_{i \in V} q(s_i, I, w_i^{\phi_2 q}) + \sum_{i, j \in E} g_{ij}^{\phi_2}(s_i, s_j; w_{i,j}^{\phi_2 g}) \right\} \\
    h(I, \phi_3; w) &= \max_s \left\{ \sum_{i \in V} q(s_i, I, w_i^{\phi_3 q}) + \sum_{i, j \in E} g_{ij}^{\phi}(s_i, s_j; w_{i,j}^{\phi_3 g}) \right\}
\end{align*}
\]

- How to combine into single classifier?

\[
    h(I; w) = \max_{k=1, \ldots, 3} \max_s \left\{ \sum_{i \in V} q(s_i, I, w_i^{\phi_k q}) + \sum_{i, j \in E} g_{ij}^{\phi}(s_i, s_j; w_{i,j}^{\phi_k g}) \right\}
\]
Given training data \( \{(I^i, s^i, \phi^i)\} \) that contains ground truth angle in data as well

\[
h(I; w) = \max_{k=1,...,3} \max_s \{ \sum_{i \in V} q(s_i, I, w_i^{\phi_k q}) + \sum_{i,j \in E} g_{ij}(s_i, s_j; w_{i,j}^{\phi_k g}) \}
\]

\[
R(w; I^i, s^i) = \max_{k \in K} \max_{s \in Y} \{ L(s, s^i) + \langle w, \Psi(I^i, s, \phi^k) \rangle \} - \langle w, \Psi(I^i, s^i, \phi^i) \rangle
\]
Surrogate convex risk

Given training data \( \{(I^i, s^i, \phi^i)\} \) that contains ground truth angle in data as well

\[
h(I; w) = \max_{k=1,\ldots,3} \max_s \left\{ \sum_{i \in V} q(s_i, I, w^{\phi^{k,q}}) + \sum_{i,j \in E} g_{ij}(s_i, s_j; w^{\phi^{k,g}}) \right\}
\]

\[
R(w; I^i, s^i) = \max_{k \in K} \max_{s \in Y} \left\{ L(s, s^i) + \langle w, \Psi(I^i, s, \phi^k) \rangle \right\} - \langle w, \Psi(I^i, s^i, \phi^i) \rangle
\]

- Can we learn classifier if our data did not have ground truth of angle but only landmark position
- If angles are missing in training data, check out Latent SVMs!