Semi-supervised Adaptive Facial Tracking Method

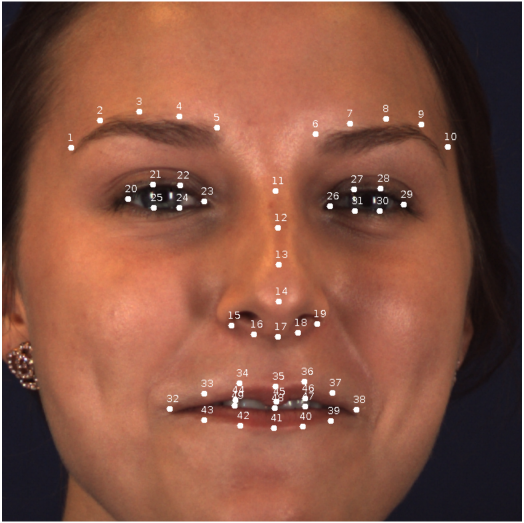
Author: Hy Truong Son
Supervisor: Dr. Lőrincz András

Eötvös Loránd University

Szeged, April 2015

Contents

Facial tracking problem



Computer vision features

Hy T. Son

Chapter I: Introduction

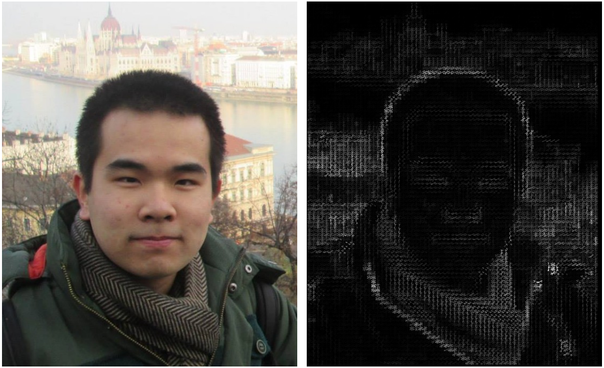
Chapter II: Methods

Chapter III: Experiments and Results

Chapter IV: Future works

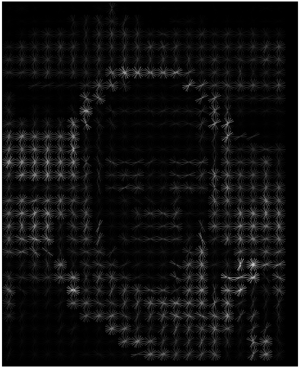
Conclusion

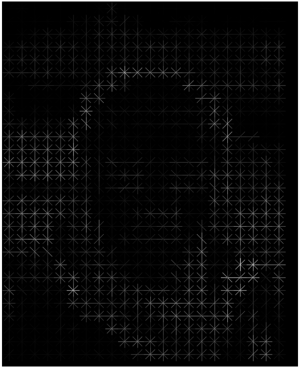
Histogram of Oriented Gradients (HOG) was invented by Navneet Dalal and Bill Triggs. In this case, 9 spatial bins and 8 by 8 patch size:



Computer vision features

Less features are kept with 24 by 24 patch size and 9 spatial bins or 4 spatial bins:



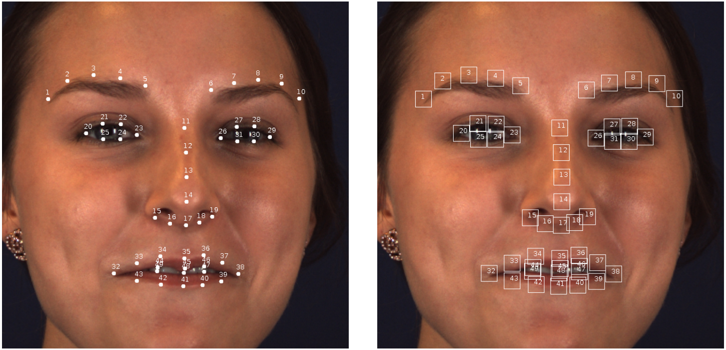


Java source code:
http://people.inf.elte.hu/hytruongson/Download/HOG/

Human face features

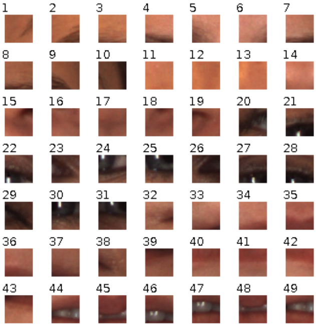
Viola-Jones algorithm

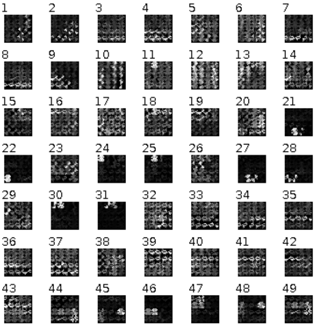
Human face features



Regions of interest around critical landmarks of the face

Human face features





HOG descriptors for each area of interest

Idea - Supervised training

In the supervised training process:

- Supervised descent method (SDM) as sequential linear regressions
- Feature extraction by HOG descriptor

Idea - Unsupervised self-improving algorithm

We expect that the large number of good training samples will compensate for the imprecise positioning of the landmark without human supervision, while the novel textures will improve the database.

- Apply linear regressors trained before to find the best possibly estimated landmarks
- With the new landmark, make a new training sample and adjust the linear regressors by Online Pseudo-Inverse Update Method (OPIUM) algorithm.

SDM

- Supervised Descent Method in facial tracking context was proposed by X. Xiong and F. De la Torre (CMU) in 2013.

Assuming that we are given a set of face images $\{d^i\}$ and their correspondent ground truth landmarks $\{x_*^i\}$. Find $\{R_k\}$ by minimizing:

$$\sum_{d^i} \sum_{x_k^i} \| \Delta x_{k+1}^i - R_{k+1}\phi_k^i \|_2^2 \quad (2)$$

where $\Delta x_{k+1}^i = x_*^i - x_k^i$. The update rule for x_k^i:

$$x_{k+1}^i = x_k^i + R_{k+1}\phi_k^i$$

Pseudo-inverse solution of SDM

- The pseudo-inverse solution of SDM is a stepwise training: linear in each step, but non-linear after more than one steps.

The matrix equation:

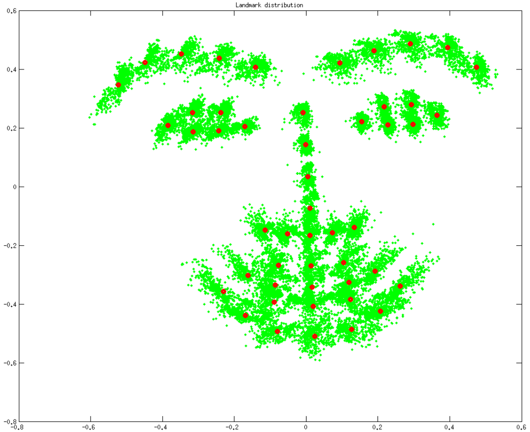
$$Y_{k+1} = R_{k+1} A_k \quad (3)$$

where $Y_{k+1} = [\Delta x_{k+1}^1 \Delta x_{k+1}^2 ... \Delta x_{k+1}^m]$, and $A_k = [\phi_k^1 \phi_k^2 ... \phi_k^m]$ where m is the number of training samples. The solution of R_{k+1} from equation (3) can be given in a closed form:

$$R_{k+1} = Y_{k+1} A_k^T (A_k A_k^T)^{-1} = Y_{k+1} A_k^+ \quad (4)$$

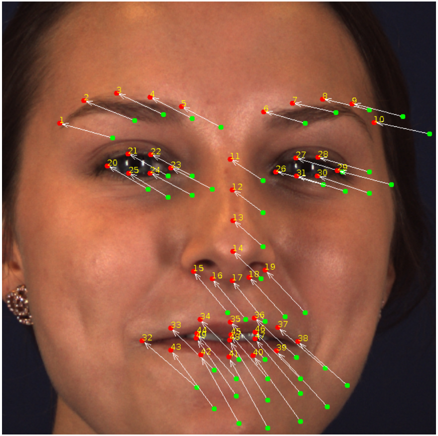
where A_k^+ is the pseudo-inverse of A_k.

Starting landmark for SDM



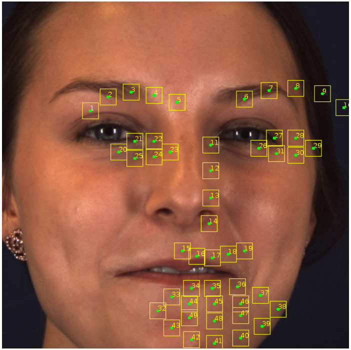
Landmark distribution over the training database

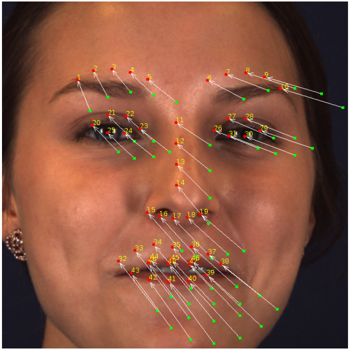
SDM Training



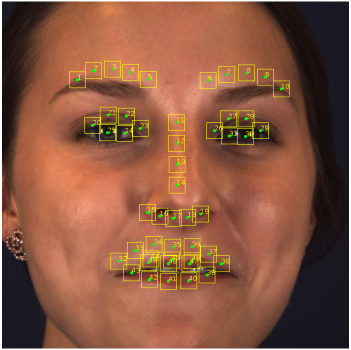
Green dots denote the starting landmark, red dots denote the ground truth landmark and white arrows correspond to vector Δx. SDM learns Δx by sequential linear regressions.

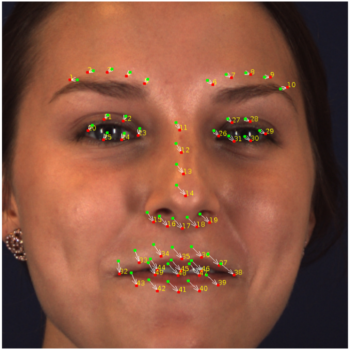
SDM Tracking - Step 1



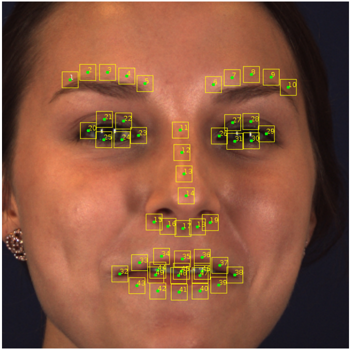


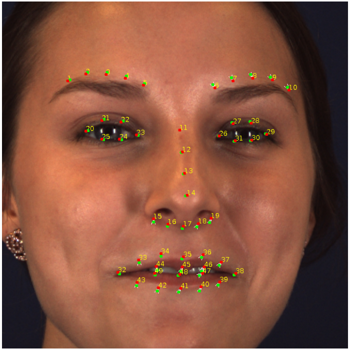
SDM Tracking - Step 2



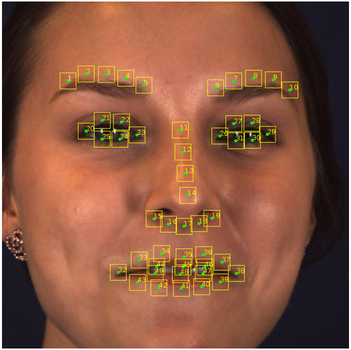


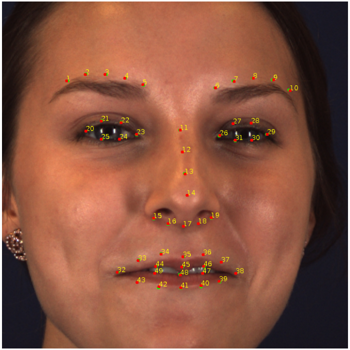
SDM Tracking - Step 3



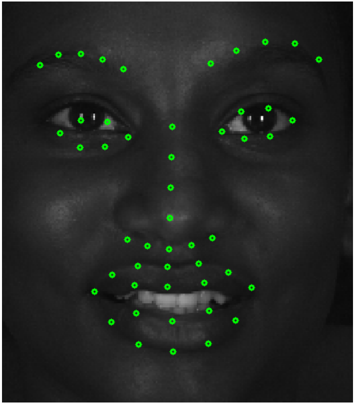


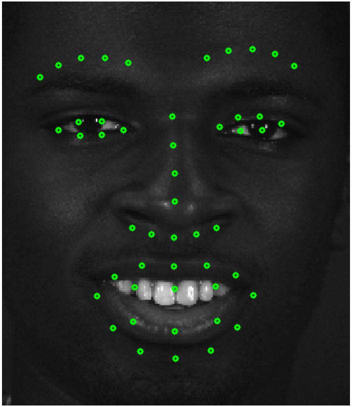
SDM Tracking - Final step





SDM is NOT always perfect!





Online Pseudo Inverse Update Method - OPIUM

- Proposed by A. van Schaik and J. Tapson in the context of Extreme Learning Machines in 2013.
- OPIUM is a quasi-linear approximation of the the pseudoinverse correction.

Given a series of pairs (a^t, y^t) where a^t and y^t is a new column of A and Y while t is a time index. R_t can be updated iteratively:

$$R_t = R_{t-1} + (y^t - R_{t-1}a^t)(b^t)^T \quad (5)$$

where vector b^t is given by:

$$b^t = \frac{a^t}{1/c + (a^t)^T a^t}, \quad c \in \Re$$

Simplification from Greville's method based on Kovanic's matrix inversion lemma.

Adaptive unsupervised tracking - Part 1/3

- The set of original matrices: $\{R_1^{original}, .., R_K^{original}\}$.
- The set of adaptive matrices: $\{R_1^{adaptive}, .., R_K^{adaptive}\}$.
- Initially, $R_k^{adaptive} := R_k^{original}$, $k = 1 \rightarrow K$.
- After a period of τ frames, we need to reset the set of adaptive matrices to the original ones.
- Let J be the number of big-iterations (repetition of the set of matrices) to track a single image.
- No adaptation for R_1, R_2. In our experiment, $K = 4$.

Adaptive unsupervised tracking - Part 2/3


```
For $t := 1 \to \infty$
    - If $t \ mod \ \tau = 1$ then $R_k^{adaptive} := R_k^{original}, \ k := 1 \to K$
    - Get a new image $d^t$, normalize by Viola Jones rectangle.

    - Use SDM to track this image:
    For $j := 1 \to J$
        For $k := 1 \to K$
            - Feature extraction:
                $\phi_{k-1,j} := [h(d^t(x^t_{k-1,j})), -1]^T$
            - Update new landmark:
                $x^t_{k,j} := x^t_{k-1,j} + R_k^{adaptive} \phi_{k-1,j}$
        End loop;
    End loop;
```

Adaptive unsupervised tracking - Part 3/3


```
    - Adaptation by OPIUM-Light:
    For k := 3 → K
        a := φ_{k-1,J}
        b := a / (1/c + a^T a)
        R_k^{adaptive} := R_k^{adaptive} + (x^t_{K,J} − R_k^{adaptive} a) b^T
    End loop;
End loop;
```

$$a := \phi_{k-1,J}$$

$$b := \frac{a}{1/c + a^T a}$$

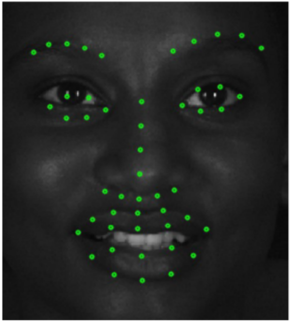
$$R_k^{adaptive} := R_k^{adaptive} + (x^t_{K,J} - R_k^{adaptive} a) b^T$$

Technical information:

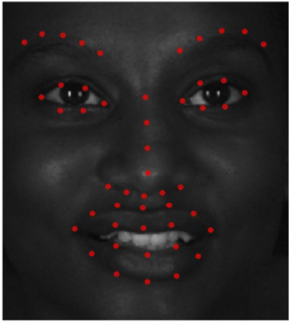
- The constant c is set to be 1 in our experiment.

Adaptive unsupervised tracking improves

Green: Normal SDM



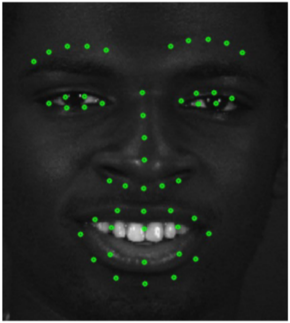
Red: Adaptive tracking

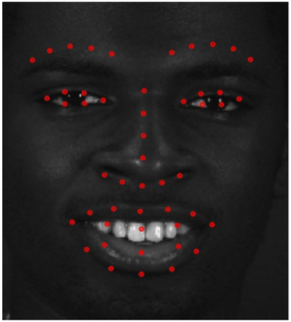


RMSEs	Face	Eye corners	Mouth corners
SDM	5.4193	1.4488	0.6497
Adaptive	4.4806	0.7266	0.3907

Adaptive unsupervised tracking improves

Green: Normal SDM   Red: Adaptive tracking





RMSEs	Face	Eye corners	Mouth corners
SDM	7.5251	1.0476	1.6018
Adaptive	4.8759	0.3805	0.4137

Databases

Hy T. Son

Training dataset:

- Multi-PIE
- CK+
- BU-3DFE

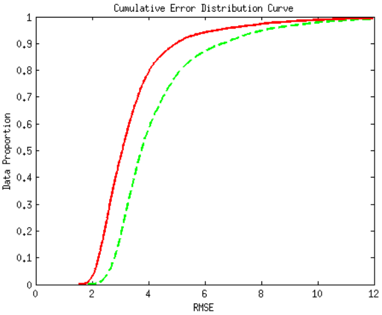
Testing dataset:

- BP-4DSFE
- Multi-view gaze

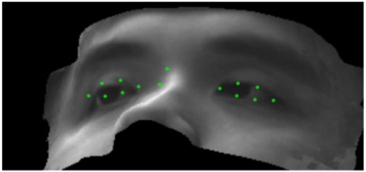
Testing on BP-4DSFE Database

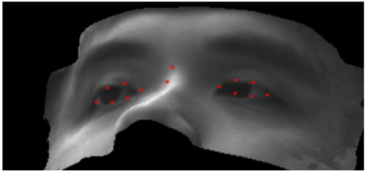
τ/J	1	2	4	8
0	4.3363	3.1580	3.1839	3.1957
1	4.1406	3.0392	3.0791	3.0878
2	4.0100	2.9660	3.0016	3.0046
4	3.8289	2.9010	2.9150	2.9192
8	3.6574	**2.8463**	**2.8385**	**2.8591**
16	**3.4974**	2.8841	2.8951	2.9209
32	3.6167	3.2403	3.2196	3.2221
64	4.6297	4.1705	4.1673	4.1607

Table 1. Average RMSEs of 49 facial landmark points



Testing on Multi-view gaze Database





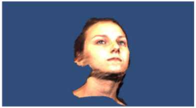
Green: Normal SDM Red: Unsupervised adaptive tracking

Future research

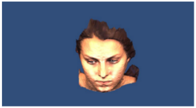
- ELM creates new features from output layer with HOG descriptor as signal for input layer going through a linear mapping and sigmoid function.
- Facial tracking with larger angles of head pose:

a. $yaw = 0°$ and $pitch = 0°$

b. $yaw = 45°$ and $pitch = 0°$



c. $yaw = 22.5°$ and $pitch = 20°$



d. $yaw = 0°$ and $pitch = -30°$

Conclusion

- More precise results at corners and their slight improvements make the key for the successes of the methods.
- Fast, efficient unsupervised tracking.
- Quickly adapt to environment conditions (light, background) during real time application.
- Personalize subjects successfully without additional data.

Finally, I would like to express my deep gratitude to my supervisor Dr. Lőrincz András, members of Neural Information Processing Group (NIPG) and other people who helped me complete this thesis.

Thank you very much for your listening!