Identification Systems
Lecture 10

Piotr Syga

2021-05-18

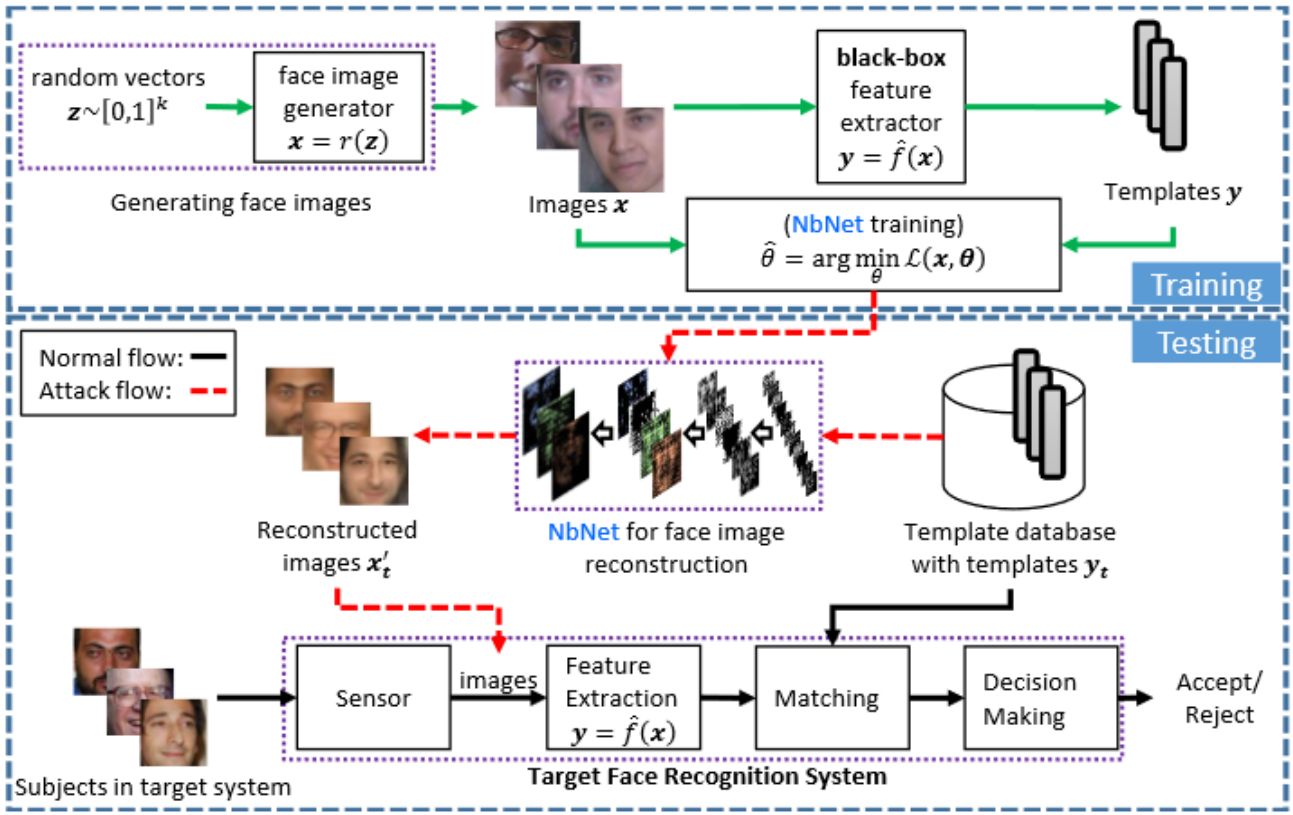
Motivation

- facial samples are easy to acquire without subject's consent or knowledge
- profiles are easy to reverse – landmarks, Eigenface/Fisherface, heatmaps
- even for deep neural network based methods – G. Mai et al.: On the reconstruction of face images from deep face templates
- threats include stalking, presentation attacks, DeepFakes

Reconstruction

- the adversary aims to impersonate subject during authentication
- the adversary has user templates (e.g., database leakage) and a black-box feature extractor of the authenticating system (e.g., buying SDK)
- the adversary has no knowledge on the victim nor the enrollment
- the attack allows only limited number of trials, presenting only fake faces with no return information on similarity score
- solution based on de-convolutions and neighborly de-convlotion blocks
- achieves over 95% TAR@0.1FAR on LFW database

Reconstruction



Idea

Z. Erkin et al.: *Privacy-Preserving Face Recognition*

- protecting facial templates used in Eigenface recognition
- using MPC techniques, jointly running the algorithm so that image provider does not learn anything more than basic parameters and the server does not learn the image
- utilizing homomorphic encryption – Pallier and Damgård, Geisler, Krøigaard
- operating on secured data

Basics

Homomorphic encryption

We call a public-key encryption schemes an additively homomorphic cryptosystem if given ciphertexts $\mathrm{Enc_k}(\mathrm{m_1})$ and $\mathrm{Enc_k}(\mathrm{m_2})$ we can calculate $\mathrm{Enc_k}(\mathrm{m_1} + \mathrm{m_2})$ as $\mathrm{Enc_k}(\mathrm{m_1})\mathrm{Enc_k}(\mathrm{m_2})$. Moreover, for a constant c and ciphertext $\mathrm{Enc_k}(\mathrm{m_1})$ we have $\mathrm{Enc_k}(\mathrm{m_1} \cdot \mathrm{c}) = \mathrm{Enc_k}(\mathrm{m_1})^c$.
Examples: Pallier, DGK, BGV, GSW, CKKS.

Basics

Pallier cryptosystem

Let $n = pq$ of size k, with p, q prime and k from the range 1000-2048. Let $l = \mathrm{lcm}(p-1, q-1)$ and $u = \frac{n}{(t^l \mod n^2)-1}$, with $t \in \mathbb{Z}^*_{n^2}$ be a private key and n and t be a public key. Let $g = n+1$ and $m \in \mathbb{Z}_n$. We calculate ciphertext $c = g^m r^n \mod n^2 = (mn+1)r^n \mod n^2$ with r being a random number from $\mathbb{Z}_n$. We decrypt the message by calculating $\hat{m} = \frac{c^l \mod n^2 - 1}{n} \cdot u \mod n$

Note that the cryptosystem is homomorphic and subject to rerandomization without knowing the private key or m.

Basics

DGK cryptosystem

Let $n = pq$ of size k, with p, q prime and k from the range 1000-2048 and message $m \in \mathbb{Z}_u$ with u being small (8 bits in the paper). We calculate the ciphertext $c = g^m h^r \mod n$ with r as in Pallier and g, h are numbers of fixed order modulo pq.
Note that DGK is also additively homomorphic and it is possible to rerandomize existing ciphertexts. We achieve performance gain due to the substantially smaller ciphertexts and the smaller plaintext space.

The protocol

- two-party setting with semi-honest attacker – learning all possible information form the logs during honest protocol execution
- communication of the participants takes place over secure channel
- we assume facial recognition system with the basis $u_1, \ldots, u_K$ and feature vectors $O_1, \ldots, O_M$
- all coordinates of the Eigenfaces and feature vectors are represented as integers (quantization)
- each feature vector has Id_i – identity of the person the feature vector belongs to, encrypted as a non-zero element of the message space
- during verification an Pallier-encrypted face image $\mathrm{Enc}(\mathrm{I})$ is presented
- the protocol ends with returned information if I matches identity of any Id_i

The protocol

- Coordinate readjustment of the image may be performed on encrypted data, i.e., instead of $H = I - S$, we calculate $\mathrm{Enc}\,(\mathrm{H}) = \mathrm{Enc}\,(\mathrm{I}) \cdot \mathrm{Enc}\,(-\mathrm{S})$.
- Similarly, we calculate eigenface vector $\mathrm{Enc}\,(\mathrm{o_i}) = \mathrm{Enc}\,(\mathrm{s_1\,u_{i1}} + \cdots + \mathrm{s_N u_{iN}}) = \mathrm{Enc}\,(\mathrm{s_1})^{u_{i1}} \cdot \cdots \cdot \mathrm{Enc}\,(\mathrm{s_N})^{u_{iN}}$, for each $i \in [K]$.
- Calculating the distance between two feature vectors O_1, O_2, i.e., $D(O_1, O_2) = \|O_1 - O_2\|^2 = \sum_{i=1}^{K} o_{i,1}^2 + \sum_{i=1}^{K}(-2o_{i,1}o_{i,2}) + \sum_{i=1}^{K} o_{i,2}^2 = \mathcal{S}_1 + \mathcal{S}_2 + \mathcal{S}_3$ also is done in encrypted manner, namely $\mathrm{Enc}\,(\mathrm{D}(\mathrm{O}_1, \mathrm{O}_2)) = \mathrm{Enc}\,(\mathcal{S}_1)\mathrm{Enc}\,(\mathcal{S}_2)\mathrm{Enc}\,(\mathcal{S}_3)$.

The protocol

- Note that $\mathcal{S}_1$ is a sum over components known to the database in plaintext.
- $\mathcal{S}_2$ is a product of elements that the database has as plaintext and has the ciphertexts $\mathrm{Enc}\left(\mathrm{o_{i,2}}\right)$, so after encrypting $-2o_{i,1}$ we can calculate the term as $\mathcal{S}_2 = \Pi_{j=1}^{K} \mathrm{Enc}\left((-2\mathrm{o_{i,1}})\mathrm{o_{i,2}}\right)$.
- Only $\mathcal{S}_3$ cannot be calculated directly due to squaring. So the server additively blinds the values, namely $\mathrm{Enc}\left(\mathrm{x_i}\right) = \mathrm{Enc}\left(\mathrm{o_{1,2}}\right)\mathrm{Enc}\left(\mathrm{r_i}\right)$, with fresh nonces r_i and sends $\mathrm{Enc}\left(\mathrm{x_i}\right)$ to the presenter for decryption.
- x_i^2 is computed in the clear and $\mathrm{Enc}\left(\mathcal{S}_3'\right)$ is sent to the server with $\mathcal{S}_3' = \sum_{j=1}^{K} x_i^2$.
- The server calculates $\mathrm{Enc}\left(\mathcal{S}_3\right) = \mathrm{Enc}\left(\mathcal{S}_3'\right)\Pi_{j=1}^{K}(\mathrm{Enc}\left(\mathrm{o}_{\mathrm{i},2}^{(-2\mathrm{r_i})}\right) \cdot \mathrm{Enc}\left(-\mathrm{r}_\mathrm{i}^2\right))$. The calculation has to be run only once during the process.

The protocol

- The last step of the protocol is finding the closest match.
- The distance between the closest match and the authentication threshold T is calculated to grant or reject the access.
- In order to find the minimal distance a recursive cryptographic protocol that compares two encrypted values is made.
- The smallest distance in each comparison is rerandomized, so that the server cannot determine the result.
- In order to track identities, we work with encrypted pairs $(\mathrm{Enc}\,(\mathrm{D_i}), \mathrm{Enc}\,(\mathrm{Id_i}))$.
- In order to compare with T, we treat it as an additional distance, hence we sort by distances the 'identity vector' $Id \in \{0, Id_1, \ldots, Id_M\}$ and if the threshold is matched we get a proper Id_i, and 0 otherwise.

The protocol

- We compare two encrypted numbers $\mathrm{Enc}\,(\mathrm{a})$ and $\mathrm{Enc}\,(\mathrm{b})$ by calculating $\mathrm{Enc}\,(\mathrm{z}) = \mathrm{Enc}\,\left(2^{\mathrm{l}} + \mathrm{a} - \mathrm{b}\right)$.
- The most significant bit of z, $z_l = 0 \iff a < b$.
- Since $z_l = 2^{-l}(z - (z \mod 2^l))$, if we had $\mathrm{Enc}\left(\mathrm{z} \mod 2^{\mathrm{l}}\right)$, we could calculate the bit.
- The value is calculated by random blinding and sending to the presenter, who calculates the modulus problem and sends the encrypted result to the server. Server can use a variant of Yao's millionaires problem in order to calculate the exact value of $\mathrm{Enc}\left(\mathrm{z} \mod 2^{\mathrm{l}}\right)$.

Idea

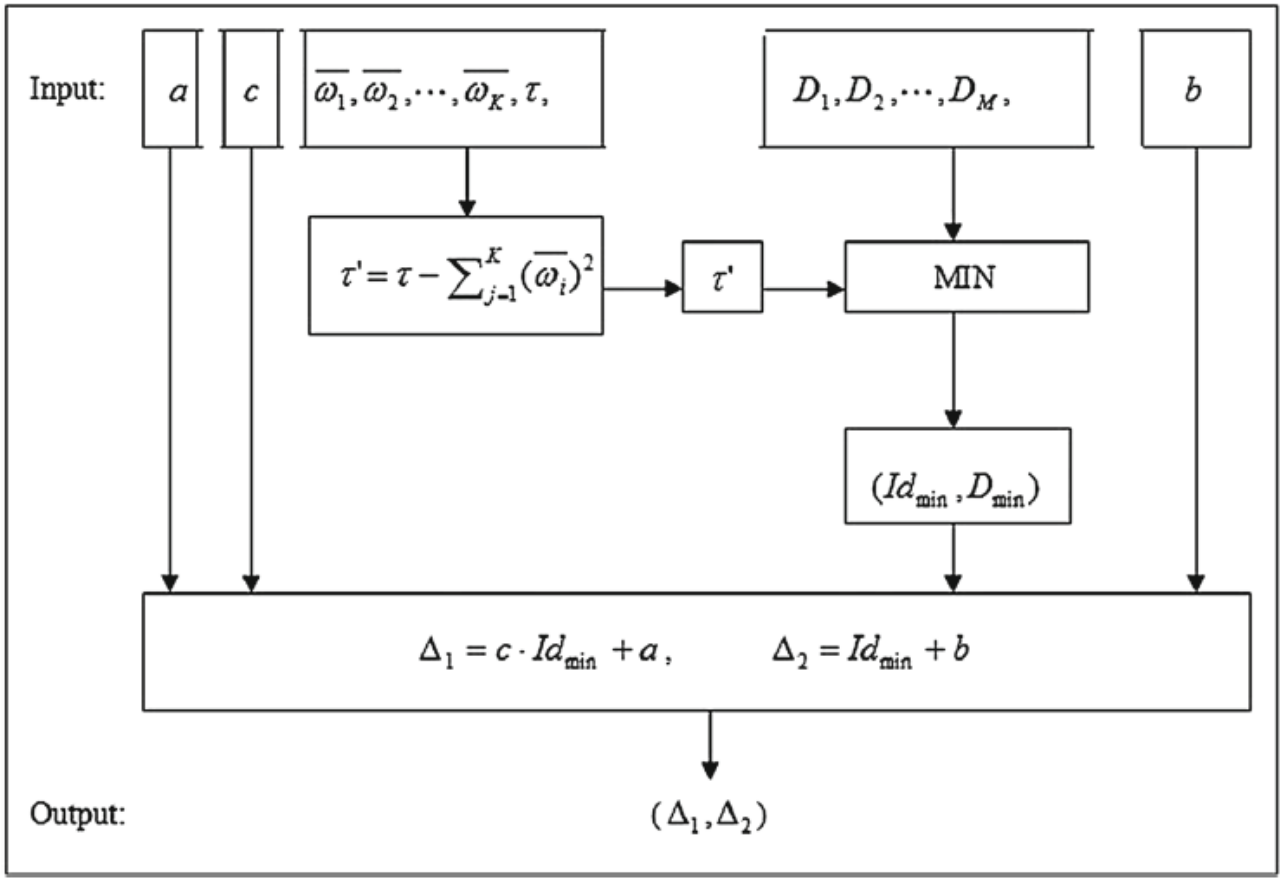
C. Xiang et al.: *Privacy-preserving face recognition with outsourced computation*

- Focuses on minimizing the computation overhead of the presenter and the server.
- Presents a secure way of outsourcing the data – involves 3 parties with semi-honesty assumption.
- Utilizes Eigenface protocol.
- Uses FHE-based hybrid encryption by Li and Lai and Pallier encryption.
- Reduces the overall online computation cost of the presenter and the database server by half.
- The presenter requires $O(1)$ decryptions.

The protocol – sketch

- The setup is similar to Erkin et al.'s protocol.
- The prover sends to the server the face image and 3 random values a, b, c – all encrypted using Pallier cryptosystem and Pallier's secret key, encrypted with FHE.
- The projection to lower dimension (on encrypted data) takes place in the cloud – encrypted normalization is sent to the database alongside user's data.
- The calculation of the distances is made in the cloud, similarly to Erkin et al.'s protocol, yet the values are ecrypted using Pallier's public key.
- Matching is made using reencrypted into FHE-based encryption distances. The comparison is done differently than in Erkin et al. – a specially designed circuit is used. Two values are calculated and decrypted by the prover If $\Delta_1 - a = c(\Delta_2 - b)$ then the identification result is $Id = \Delta_2 - b$.

The protocol – sketch



Idea

M. Plata et al.: *How to Save Your Face: a Facial Recognition Method Robust Against Image Reconstruction*

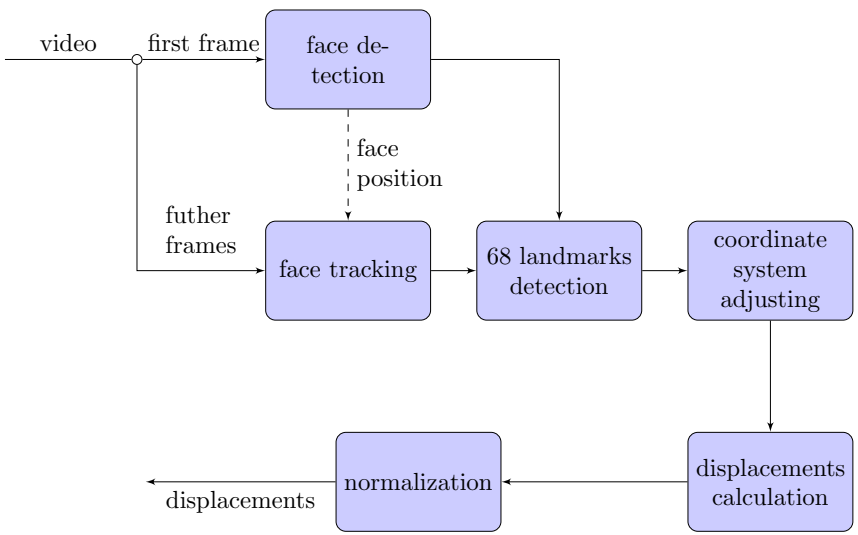
- Designing the protocol to be robust against known reconstruction methods instead of protecting existing recognition system.
- No encryption needed, low resources required.
- Using video samples instead of a single facial image.
- The subjects are reading a presented sentence, hence the method may be used as an auxiliary, anti-presentation method.
- The difference between frames is calculated using DTW and classified with SVM – no deep learning methods or large datasets are needed.
- Instead of facial landmarks, displacement vectors are used.
- The results are significantly lower than state-of-the-art NN-based identification methods that do not provide privacy protection – 70% for 115 features and 69% using 44 features.

The protocol

- We start with Viola-Jones face detection. On subsequent frames motion tracking is used instead of face detection.
- We detect 68 facial landmarks and re-adjust the coordinate system according to the nose line. The size is normalized, each subsequent frame is rescaled according to the nose length, so that head movements do not influence the displacement vectors.
- Displacement of each landmark is calculated between each frame pair, in both directions.

The protocol

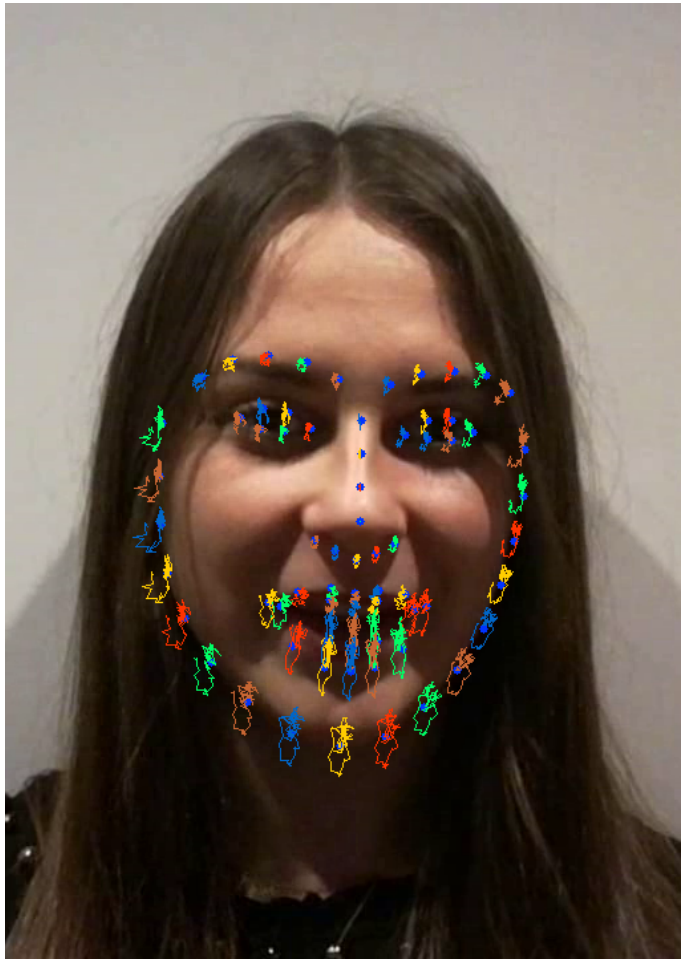
Feature extraction steps:



The protocol

Examples of displacement trace:





The protocol

- The video is represented by a tensor $D \in \mathbb{R}^{N \times F \times 2}$.
- We investigate D_{nf1} a horizontal displacement and D_{nf2} a vertical displacements of n-th point between f-th and $(f+1)$-th frame.
- A variable that is responsible for the distant adjustment – $\eta \in \mathbb{R}^{F+1}$.
- We normalized the tensor by readjusting each point f and $i \in \{1, 2\}$,

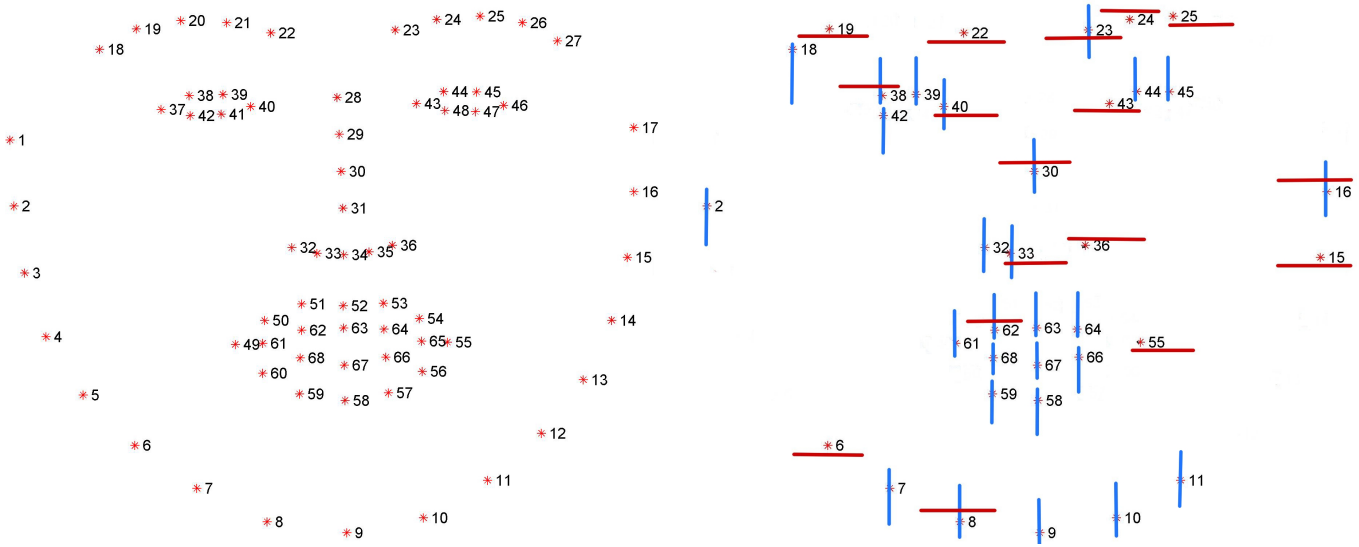
$$D_{nfi} := \frac{D_{nfi}}{\frac{1}{2}(\eta_f + \eta_{f+1})}, \tag{1}$$

where η_f refers to f-th value of the normalized vector.

The protocol

- Each point's displacement is compared separately.
- DTW used to compare two sequences of point's movement – sequences do not have to be of the same length.
- Similarity vector between two videos is obtained from DTW: $s = [s_1, \ldots, s_{136}]^T$.
- Classification was made using SVM using 3 or 4 user videos as a template.
- SVM contained $\binom{102}{2} = 5151$ and $\binom{136}{2} = 9180$ records from 3 and 4 videos of every subject.
- For identification $k = 3$ Nearest Neighbors algorithm is used.
- With vector of probabilities $h(s^{(i)})$, where $s^{(i)}$ is similarity of the given sample and i-th enrolled sample, we select k pairs of samples with the highest probability $h(s^{(i)})$.
- For the given sample we identify the subject with the highest number of chosen i-th enrolled samples.

The protocol



The data

- The idea – identification using commonly available hardware.
- Data acquired using 13M-Pixel CMOS image sensor.
- Resolution of 1280 x 720 pixels with 24fps – AVI with MPEG-4 compression.
- 34 subjects with 191 videos in total.
- The subjects only instructed to gaze into the camera repeat the given sentence.
- The subjects stood during the recording in order to replicate possible in-life scenarios.
- No restrictions upon glasses, jewelry, hairstyles, facial hairstyles nor make-ups were introduced.
- Videos were recorded in a non-laboratory environment – the background, the distance to nor the elevation from the camera was restricted. The environmental noise was not limited.
- Data processing and classification made using mid-shelf laptop.

The results

- For the user vectors derived from 3 samples we obtained: for 115 IR=70.78% , for 44 features IR=68.54%.
- For vectors derived from 4 samples and 44 features IR=69.08%.
- Mean correlation between static facial landmarks and displacement vectors was equal to 0.007546 and the standard deviation was equal to 0.022897.
- When calculating absolute values for displacements, the mean is equal to 0.186707 and the standard deviation 0.046074.
- Identification rate still above 50% even after removing 50 most correlated features.

The results

IR based on number of features:

