Background-Invariant Robust Hand Detection Using One-Class Color Segmentation and Skeleton Description

Andrey Kopylov, Oleg Seredin, Olesia Kushnir, Inessa Gracheva*, Alexandr Larin

Russia, Tula State University

andkopylov@gmail.com, oseredin@yandex.ru, kushnir-olesya@rambler.ru,
gia1509@mail.ru, ekzebox@gmail.com

IDP-2016, Barcelona, Spain

Hand detection problem

IDP-2016

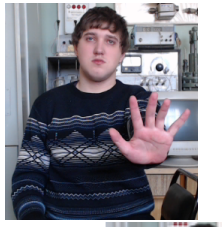
Accurate and robust hand detection in a video stream is crucial stage in construction of easy-to-use human-computer interaction systems as an alternative to touch-based devices in surgery, robot control, bio identification, etc.

Initial data: image or video frame

Result of processing: binary mask, corresponded to the hand region

Main approaches to the solution of hand detection problem

IDP-2016



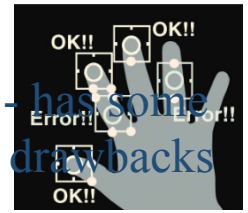
- background must remain unchangeable within the chosen model

The approach on the basis of background subtraction techniques

- geometry of skin color domain inside the color space can be changed dramatically

The approach on the basis of skin-color segmentation

+



- has some drawbacks

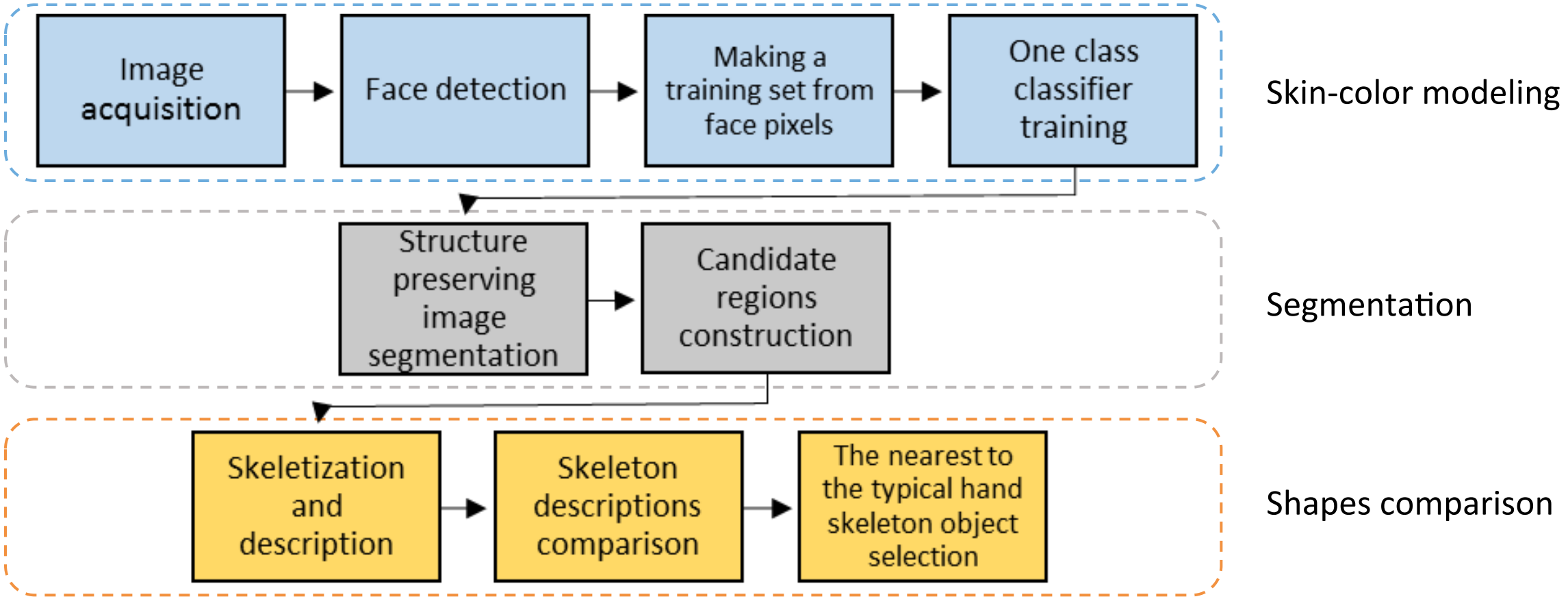
Morphology of hands

- assumes that no other objects is placed between the hand and the camera and requires additional equipment

The approach on the basis of depth data

General flow-chart of hand detection method

IDP-2016



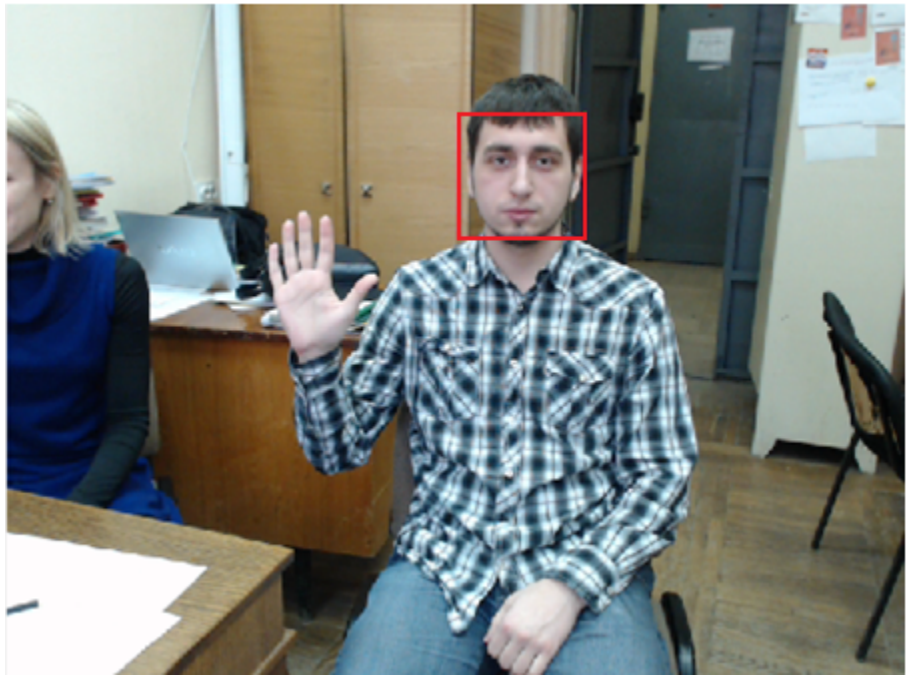
IDP-2016

SKIN-COLOR MODELING

Face detection

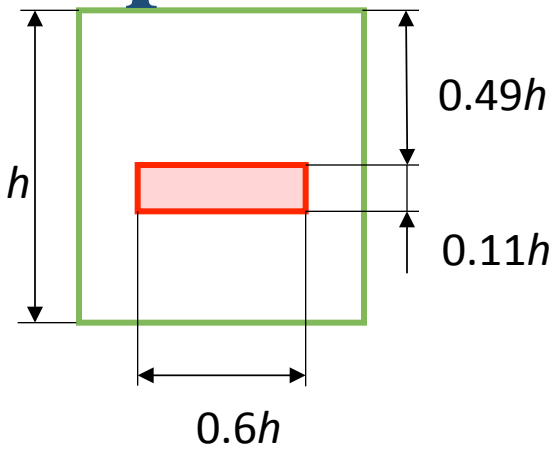
IDP-2016

We use a fragment of a human face on the current video frame for adaptive adjustment of the skin-color model to compensate illumination changes, since human faces can be easily detected by Viola-Jones method [1].



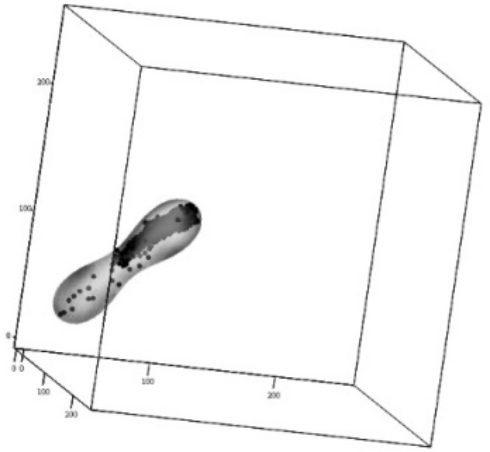
[1] Jones, M. and Viola, P. Fast Multi-view Face Detection. *Mitsubishi Electric Research Lab TR2000396*. July (2003)

Parametric representation of skin color space using one-class classifier



The parameters of training region inside the face rectangle are defined based on [2]. Particularly, just this area lies between the eyes and nose tip and is less deformable and free of beard, mustache, glasses, hair-dress, make-up.

Training Set



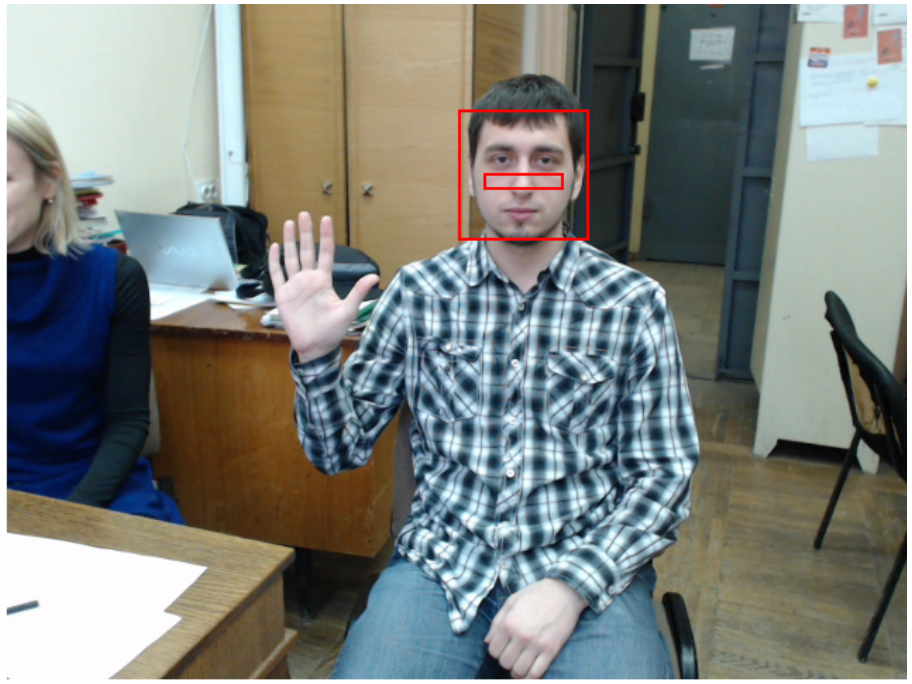
Support Vector Data Description

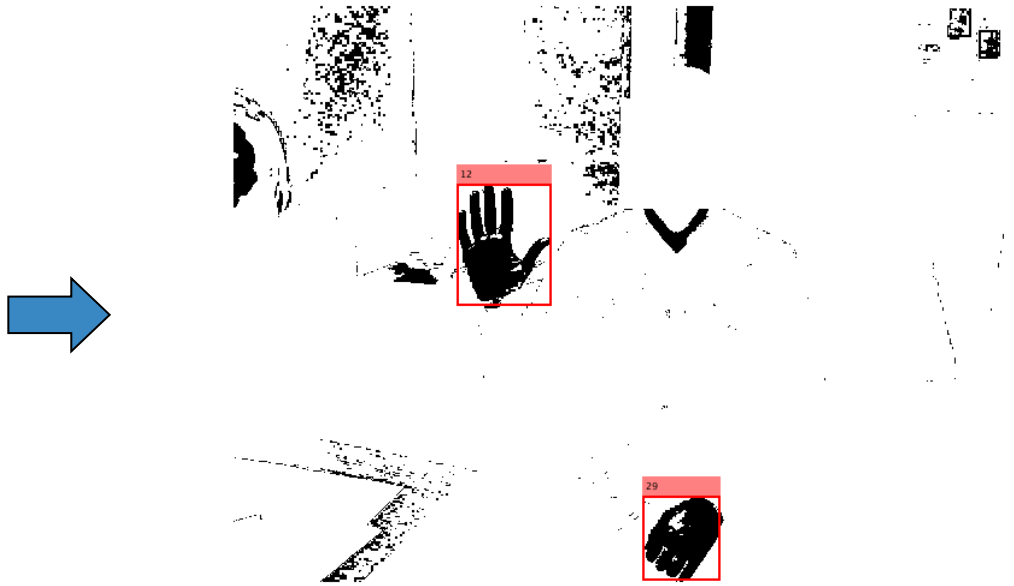
[2] Degtyarev, N. and Seredin, O. Comparative testing of face detection algorithms. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (2010), 200–209.

Pixel-wise segmentation via one-class classifier

IDP-2016

We apply here a modified version of one-class pixel classifier [3], which does not need to collect any training data for background modeling.





[3] Larin, A. et al. Parametric Representation of Objects in Color Space Using One-Class Classifiers. *Machine Learning and Data Mining in Pattern Recognition: 10th International Conference, MLDM 2014, St. Petersburg, Russia, July 21-24, 2014. Proceedings*. P. Perner, ed. Springer International Publishing. 300–314.

IDP-2016

SEGMENTATION

Gamma-Normal Model

IDP-2016

A priori joint distribution: $$\Psi(X \mid \Lambda, \delta) \propto \frac{1}{\left(\prod_{t \in T} \delta\lambda_t\right)^{1/2} (2\pi)^{(N_1 \cdot N_2)/2}} \times \exp\left(-\frac{1}{2} \sum_{t', t'' \in V} \frac{1}{\delta\lambda_{t'}} (x_{t'} - x_{t''})^2 \right),$$

A priori distribution density of coefficients:

$$G(\Lambda \mid \delta, \eta, \mu) = \exp\left[-\frac{1}{2\delta\mu} \sum_{t \in T} \left(\eta \frac{1}{\lambda_t} + \frac{1}{\eta} \ln \lambda_t \right) \right]$$

Joint prior normal gamma-distribution: $H(X, \Lambda \mid \delta, \eta, \mu) = \Psi(X \mid \Lambda, \delta) G(\Lambda \mid \delta, \eta, \mu)$

Bayesian estimate of (X, Λ):

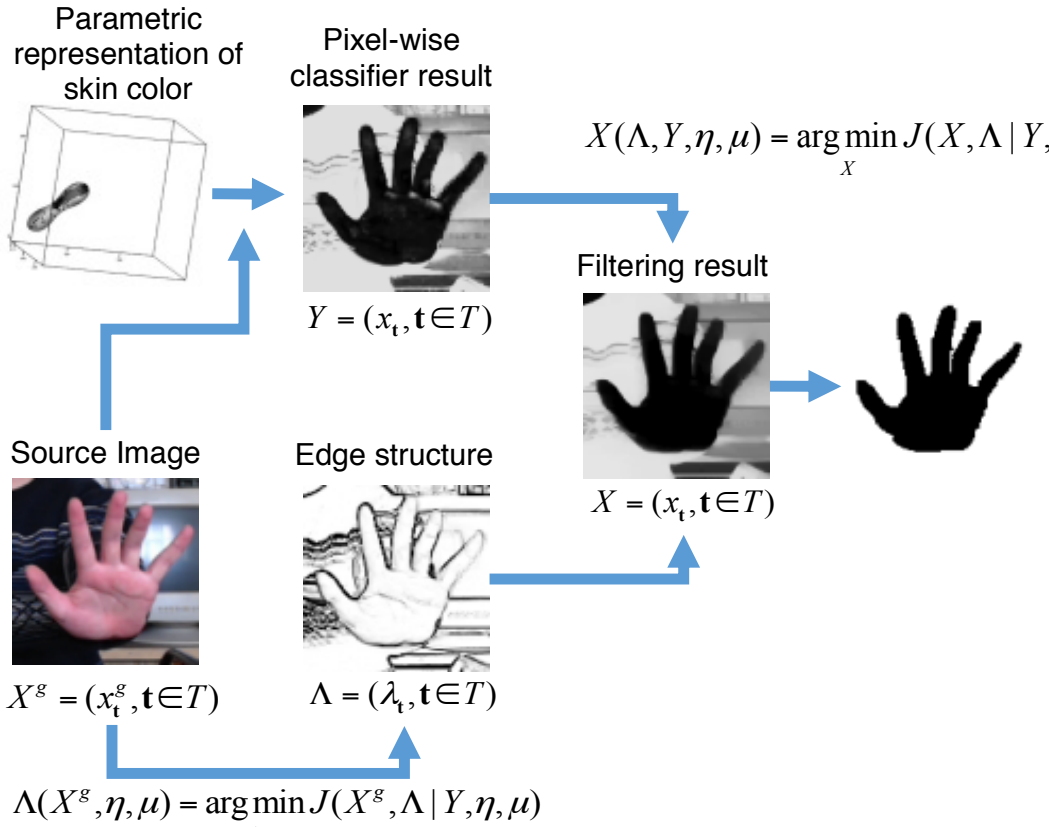
$$\begin{cases} (\hat{X}, \hat{\Lambda} \mid \eta, \mu) = \arg\min_{X, \Lambda} J(X, \Lambda \mid Y, \eta, \mu), \\ J(X, \Lambda \mid Y, \eta, \mu) = \sum_{t \in T} (y_t - x_t)^2 + \sum_{t', t'' \in V} \left\{ \frac{1}{\lambda_t} \left[(x_{t'} - x_{t''})^2 + \frac{\eta}{\mu} \right] + (1 + \frac{1}{\mu}) \ln \lambda_t \right\}. \end{cases}$$

Gracheva I., Kopylov A., Krasotkina O.: Fast global image denoising algorithm on the basis of nonstationary gamma-normal statistical model. Communications in Computer and Information Science. Springer. 542, 71-83 (2015).

Segmentation based on skin color model and image structure information

IDP-2016

To adjust output of the one-class classifier the structure-transferring filter built on probabilistic gamma-normal model [4] is applied. It utilizes additional information about the structure of an image and coordinates local decisions to achieve more robust segmentation results.



$$\begin{cases} (\hat{X}, \hat{\Lambda} \mid \eta, \mu) = \arg\min_{X,\Lambda} J(X, \Lambda \mid Y, \eta, \mu), \\ J(X, \Lambda \mid Y, \eta, \mu) = \sum_{t \in T} (y_t - x_t)^2 + \sum_{t', t'' \in V} \left\{ \frac{1}{\lambda_t} \left[(x_{t'} - x_{t''})^2 + \frac{\eta}{\mu} \right] + (1 + \frac{1}{\mu}) \ln \lambda_t \right\}. \end{cases}$$

$$\hat{\lambda}_{t'}(X^g, \eta, \mu) = \eta \frac{(1/\eta)(x_{t'}^g - x_{t''}^g)^2 + 1/\mu}{1 + 1/\mu}$$

$$\hat{X} = (\hat{x}, t \in T) = \arg\min_X J(X, \Lambda \mid Y, \eta, \mu) =$$

$$= \arg\min_X \left\{ \sum_{t \in T} (y_t - x_t)^2 + \sum_{t', t'' \in V} \frac{1}{\lambda_{t'}} (x_{t'} - x_{t''})^2 \right\}.$$

[4] Gracheva I., Kopylov A. Image Processing Algorithms with Structure Transferring Properties on the Basis of Gamma-normal Model. *Analysis of Images, Social Networks, and Texts April, 7-9th, Yekaterinburg* (2016)

A set of candidates to hand region

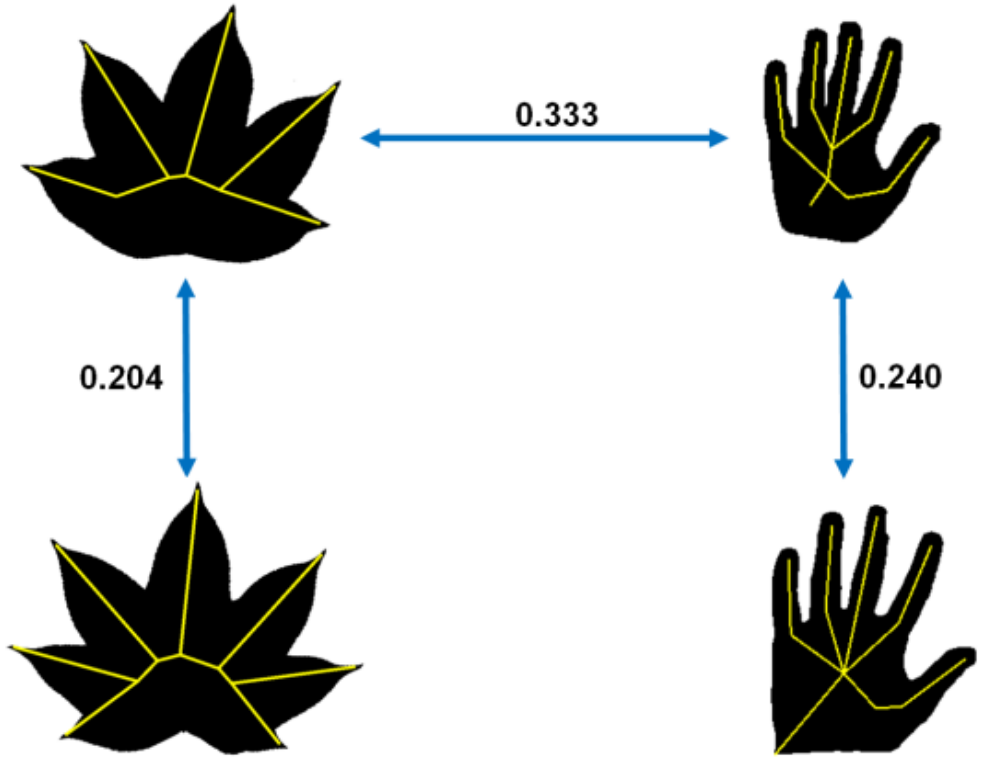
IDP-2016

The color segmentation allows to select a set of candidates according to their empirical features: geometrical characteristics; the size which is proportional to a face size; the filled-in degree which is computed as the ratio of the number of black pixels to the size of the region.

IDP-2016

SHAPES COMPARISON

Skeletons comparison procedure



The comparison procedure of skeleton graphs performs following steps [5]:

1) the comparison procedure calculates the skeleton of candidate region and the skeleton of the etalon palm,

2) code each skeleton by sequence of primitives (primitive chain); each primitive contains information about topological characteristic of the corresponding edges (lengths, inner-angles and radial function) of skeleton,

3) align two primitive chains by dynamic programming,

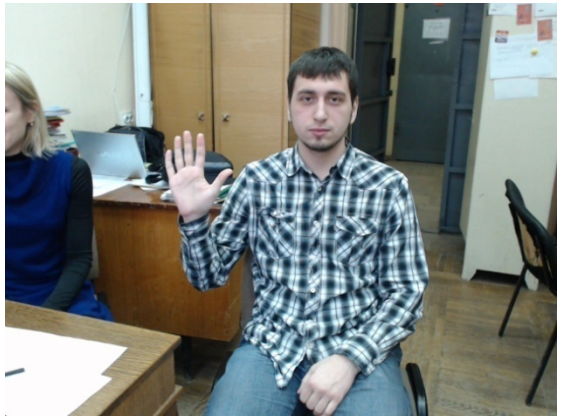
4) calculate dissimilarity measure of primitive chains and, consequently, corresponding skeletons, based on their optimal pair-wise alignment.

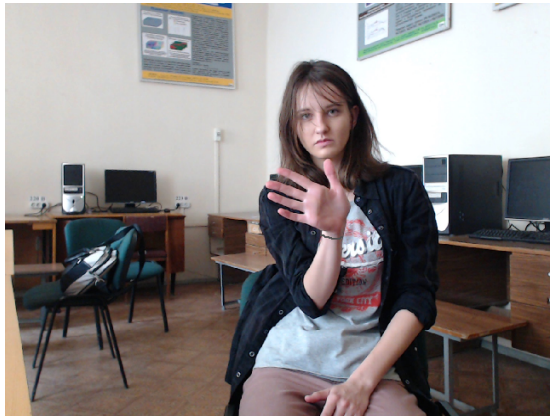
[5] Kushnir, O. and Seredin, O. Shape Matching Based on Skeletonization and Alignment of Primitive Chains. *Analysis of Images, Social Networks and Texts: 4th International Conference, AIST 2015, Yekaterinburg, Russia, April 9--11, 2015, Revised Selected Papers*. Y.M. Khachay et al., eds. Springer International Publishing. 123–136 (2015).

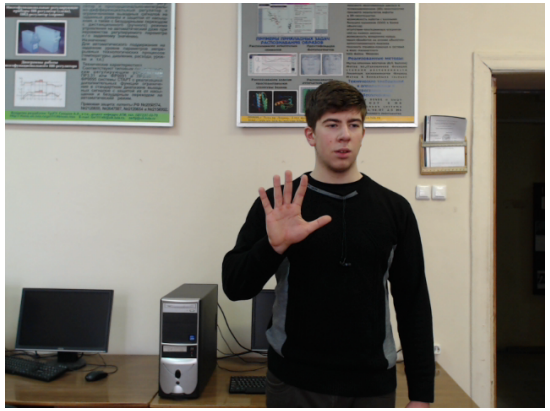
EXPERIMENTAL RESULTS

Evaluation database

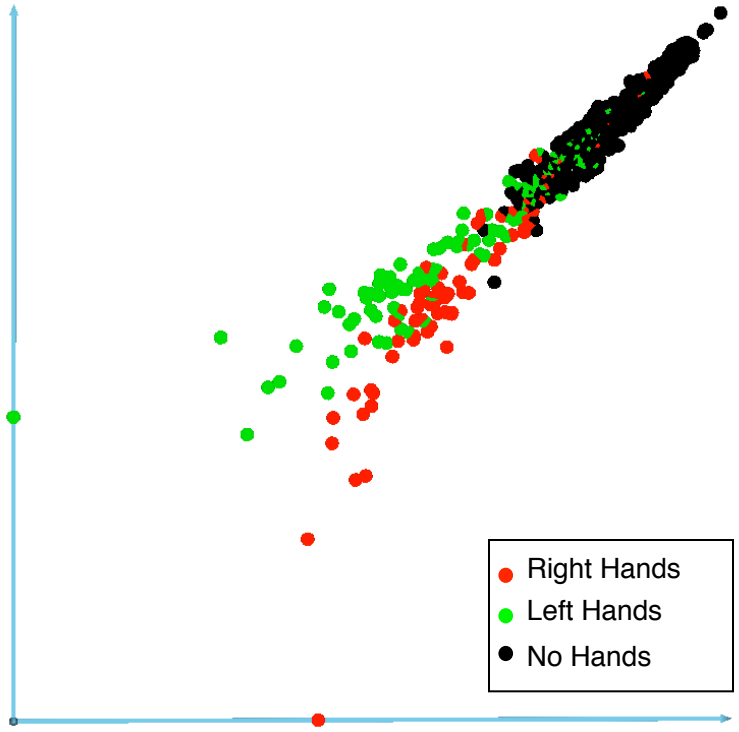
The database is available at http://lda.tsu.tula.ru/papers/TulaSU_HandsDetDB.zip. The total number of images in the database is 302. For the each image the true hand(s) position were determined by expert and marked. This information is used as ground-truth.





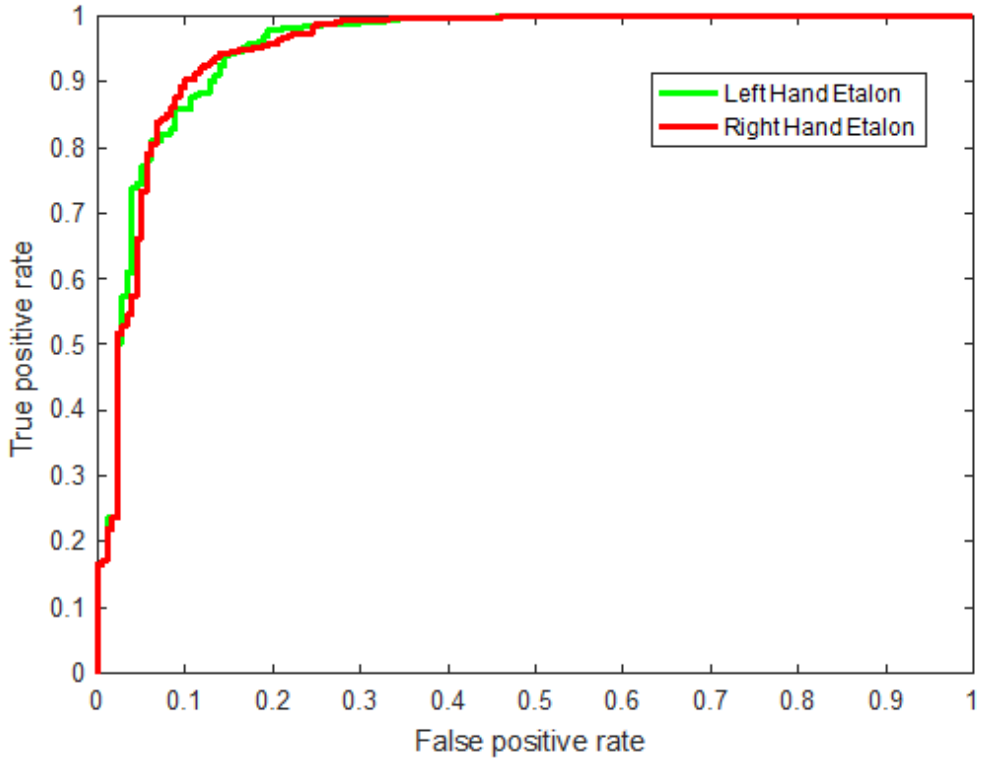


Distances from candidate-regions to hand etalons



We have a projection of whole distance matrix (721 x 721) on two-dimensional space. The first feature is the distance to the left hand etalon (horizontal axis) and the second one is the distance to the right hand etalon (vertical axis). Therefore, the “left hands” (109 instances) are the green points, “right hands” (71) are the red points and the “not hands” (541) are the black points. The chart displays that the compactness hypothesis for the three types of shapes holds true. It allows us to build the hand-detection system using simple threshold rule based on the distance to etalon(s).

Quality of hands separation



The quality of hands separation from the non-hand candidates is shown in the form of ROC-curve. This curve demonstrates the recognition true positive rate versus false positive rate along the increasing of comparison distance from the chosen etalon object. The AUC (area under the curve) for left hands (green curve) is equal 0.9535, and for the right hands (red curve) – 0.9531.

Experimental results

IDP-2016

Parameter	Our method	GrabCut- based method [6]
Number of instances in Database	302	
Number of images where face was found by Viola-Jones method	289	
Average number of candidates per image	3.02	2.39
One of the several candidates corresponds to the ground-truth hand position	233	172
Number of cases, when a single candidate corresponds to the ground-truth	54	36
No candidates found	2	37

Average computational time per image for segmentation algorithms running in MATHLAB environment was 1.3 sec. for graph cut based optimization method and 0.3 sec. in the case of our method on the basis of probabilistic gamma-normal model.

[6] Boykov, Y. and Kolmogorov, V. An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 26, 9 (2004), 1124–1137.

THANK YOU FOR ATTENTION!