Hand Detection Based on Skin Color Segmentation and Classification of Image Local Features

Ahunzyanov Rasim[1], Tropchenko Alexander [1]

[1]*Saint-Petersburg National Research University of Information Technologies, Mechanics and Optics, Kronverkskiy pr. 49, Saint-Petersburg, Russia*

***Abstract* –This paper presents a hand localization method. It built using adaptive skin tone segmentation and identification of image keypoints specific for a hand. Details about skin detector implementations provided.Application of different feature detectors and descriptors for hand detection is considered. Experimental results show the performance of a method, and demonstrate promising results for hand detection.**

***Keywords* – Hand Detection, Skin Detection, Hand Posture Recognition, Local Features, Artificial Neural Networks.**

1. Introduction

Recognition of hand gestures is one of the many parts of human–computer interaction (HCI) and it attracts the attention of many researches. First widely available peripherals designed especially for gesture recognition were unveiled in 2012. Generally such devices use an infrared projector and camera and a special algorithms to track the movement of objects in three dimensions. But there is still no robust solution based on using of usual color camera only.

First task of any hand gestures recognition system is to determine the region of image where the hand is located. Some methods of hand detection are based only on detection of skin tone color [1, 2] without using any textural features. These assumptions have a negative impact on the quality of detection.Usage of more complex approaches [3, 4] improves detection quality up to 90%. On the other hand these sophisticated methods are inapplicable for real time applications.

We propose simple texture analysis method in supplement to skin detection that allows increasing of the confidence that found piece of skin is really a hand. The paper organized as follows. First we shortly review several skin detection approaches and propose our one in section 2. Also, we provide evaluation of our skin detection method and its comparison to non-parametric approach. Section 3 brieflylists some keypoints detection techniques and then describes our approach to classification of descriptors for detection of hand postures.Also, we give descriptions of actual detector construction and overall classification framework as well. Comparison of different approaches to local feature detection was made.In section 4 results that were obtained are discussed. Also, results for particular cases are represented.

2. Skin colorsegmentation

2.1. Skin color models

There are big amount of skin color modeling techniques. This is incomplete list of them[5]:

1. Explicit skin-color space thresholding;
2. Histogram model with naive Bayes classifiers;
3. Gaussian model classifiers;
4. Elliptical boundary models;
5. Multi-layer perceptron (MLP) classifiers;
6. Others (Maximum entropy classifiers, Bayesian network classifiers, boosted classifiers, etc.).

Explicit thresholding and similar approaches typically considers only a few skin types and a few possible illumination conditions, thus they are not best choice for qualitative classification. Gaussian and elliptical color models of skin being trained once and later there are no ability to adjust their parameters.Such models are not adaptive to environment changes, because they do not support incremental learning.Also, paper [5] shows that Bayesian approach show good performance in skin color detection. Typically TPR of the Bayes classifier lies within a range [0.88; 0.98] and FPR does not exceed 30%. Bayesian classifier might be incrementally trained during operational process.Thereforethis approach was used as a base for our skin detector.

2.2. Skin classification using Bayesian approach

Let's consider skin detection as probabilistic task. Then for each color c we should to know probabilities $P(skin \mid c)$ and $P(\neg skin \mid c)$.

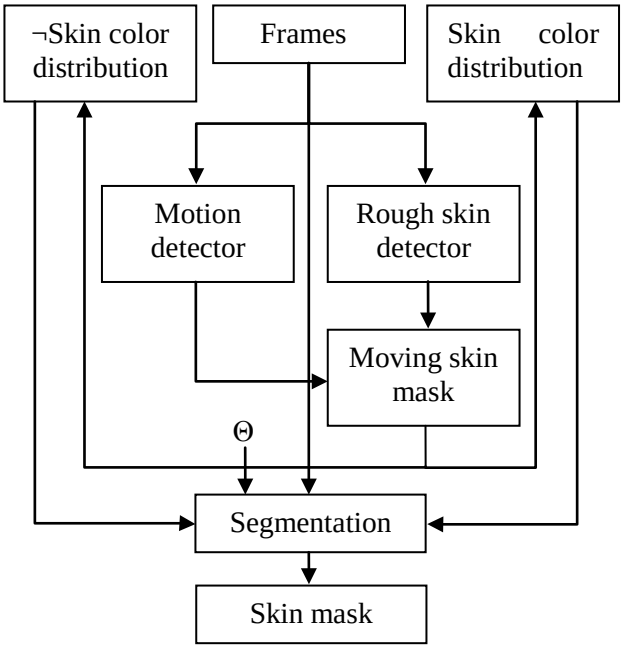


Figure 1. Skin segmentation framework.

$P(skin \mid c)$ is a probability of observing skin, knowing a concrete color value. $P(\neg skin \mid c)$ is the same thing for non-skin case. Then therule for the skin classifier decision might be written as follows:

$$I(x,y) = \begin{cases} skin, \dfrac{P(skin \mid c(x,y))}{P(\neg skin \mid c(x,y))} > \Theta \\ \neg \text{skin}, \text{otherwise} \end{cases}. \quad (1)$$

Threshold value Θ might be chosen on the stage of learning of skin detector to obtainreasonable values of TPR and FPR.

Unfortunately it is difficult to estimate $P(skin \mid c(x,y))$ and $P(\neg skin \mid c(x,y))$ directly. Instead, we can evaluate probabilities ratiousing the Bayes rule:

$$P(skin \mid c) = \frac{P(c \mid skin)P(skin)}{P(c)}, \quad (2)$$

$$P(\neg skin \mid c) = \frac{P(c \mid \neg skin)P(\neg skin)}{P(c)}. \quad (3)$$

Then the decision rule can be reduced and rewritten in the following form:

$$\Theta < \frac{P(c \mid skin)P(skin)}{P(c \mid \neg skin)P(\neg skin)}. \quad (4)$$

And finally constant skin to non-skin prior probabilities ratio can be moved to the left part of inequality:

$$\Theta \times \frac{1 - P(skin)}{P(skin)} < \frac{P(c \mid skin)}{P(c \mid \neg skin)}. \quad (5)$$

The left part of inequality (5) is constant which might be chosen during evaluation of the classifier.Also, this shows that the $P(skin)$ prior probability does not affect the overall detector behavior due to nature of the Bayes model.

Thus, task of color classification can be reduced to estimation of skin likelihood functions and threshold value.

2.3. Histogram-based skin classification

To estimate skin and non-skin likelihood we use the following scheme. Input frames passed through motion detector and rough non-parametric skin detector. We use background detector described in [6] to extract foreground from frames. And we use skin tone detector suggested in [7] which uses explicitly defined skin regions.Of course, any other skin detector might be used instead. Further we use foreground which is marked as skin to update skin color model P_t^S of current frame and background used to update non-skin color model P_t^{Bg} of current frame.Skin color models are built in HS and CrCb color spaces with ignoring illumination component. Updates to the histogram bins are made via the following model:

$$P(c \mid skin)_t = (1-\alpha)P(c \mid skin)_{t-1} + \alpha P_t^S, \quad (6)$$

$$P(c \mid \neg skin)_t = (1-\alpha)P(c \mid \neg skin)_{t-1} + \alpha P_t^{Bg}. \quad (7)$$

Value α is a scalar between 0 and 1 that allows to control the learning speed of concrete color model. Since the histograms P_t^S and P_t^{Bg} obtained from a single frame, they might be bad sampled. Especially, this problem concerns the skin color distribution, since skin regions are small in comparison to the background. Smoothing of the histogram of skin-color with Gaussian kernel with $\sigma = (0.1..0.6)$ helps to solve this issue.

The last stage of skin color detection is applying of decision rule (5). Overall scheme of the process of classification is shown on figure 1.The segmentation part ofthe method issimple and computationally fast as we need only two look-up tables to store the probability of the skin.Estimation part of the method might be efficiently implemented in parallel program threads. Color conversions between different color spaces are not shown on the scheme.

2.4. Skin classifier evaluation

To evaluate the performance of our skin detector, we got a set of 21 video sequences, which used by authors of [8] to perform their experiments. Only 10

sequences where camera is not moving were used in evaluation. Camera should be statically mounted and this limitation arises because we use motion detector.People of several nationalities with various skin tones are represented. Scene contains hands and faces. Also there are hand labeled ground truth data for performance evaluation. For each frame two masks created for skin and non-skin regions respectively.

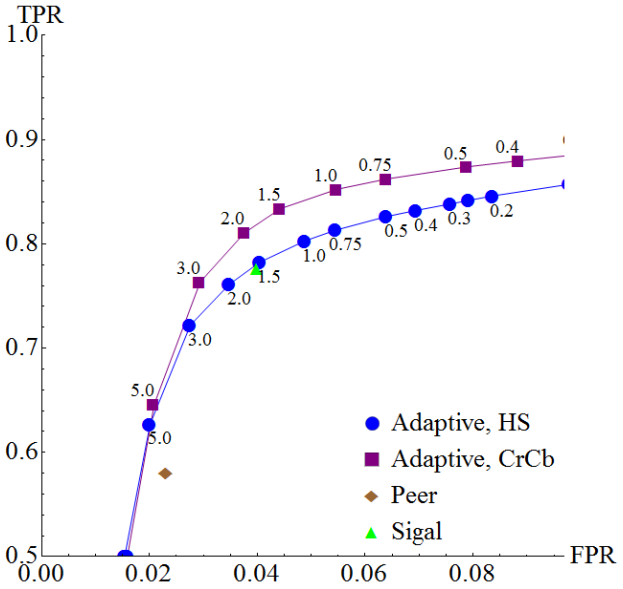
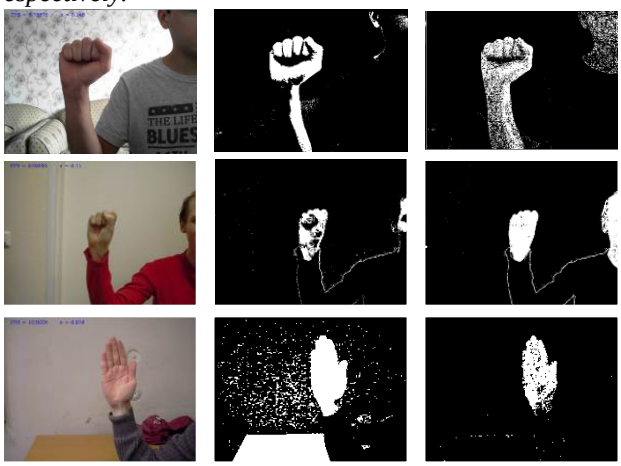


Figure 2. ROC curves of proposed skin detector.

(a) (b) (c)

Figure 3. Comparison between Peer skin detector and our skin detector on frames from [8]. Column (a) shows the input frame. Images in column (b) and (c) were processed by Peer skin detector and by our skin detector respectively.



(a) (b) (c)

Figure 4. Comparison between non-parametric Peer skin detector and our skin detector on our video. Column (a) presents the input images. Images in column (b) were processed by non-parametric Peer skin detector. Images in column (c) were processed by our skin detector.

Performance of skin detector was measured using the confusion matrix. Confusion matrix is a table layout widely used for binary classifiers evaluation. For every hand-labeled frame of the sequence confusion matrixwas computed.To obtain aggregate measures, confusion matrices for different frames were summarized.

To compare classification characteristics of our skin detectors we plot ROC curves with using HS and CrCb color spaces as skin/non-skin models. As can be seen from the graph in Figure 2, the adaptive approach was entirely better than the non-parametric Peer method. At the same time our detector shows performance slightly better thanthe adaptive approach from [8].

Proposed method is less dependent on the shape of skin locus in color spacebecause illumination component is not used for building of skin tone model. It takes into account overlap of skin and non-skin colors. Also, exclusion of the illumination component from the classification process helps to generalize sparse training data.

3. Characteristic local features and hand detection

3.1. Image local features

Image local feature is a pattern which differs from its immediate neighborhood. It is usually associated with a change of an image property or several properties simultaneously, although it is not necessarilylocalized exactly on this change. Typically, some measurements are taken from a regioncentered on a local feature and computed into descriptors. The descriptors can then be used for various applications.

There are a lot of techniques for local features detection. Here is anincomplete list of them: SURF [9], SIFT [10], ORB [11] and Star [12]. They take into account different image feature types such as blobs,edges, corners and T-junctions.Types of feature detectors and descriptors which are used in our experiments summarized in Table 1.

Hand posture forms a texture which consists of

Table 1. Combinations of detectors and descriptors which are used in experiments and types of features which can be detected. Size of the corresponding descriptor listed in the last column.

Detector	Extractor	Blob	Edge	Corner	T-Jct.	Descr. size
Star	SURF	•				128
SIFT	SIFT	•		•	•	128
SURF	SURF	•		•	•	128
ORB	ORB		•	•	•	256
ORB	SURF					128

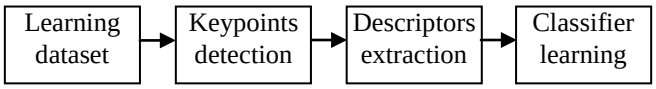


Figure 5. Learning framework.

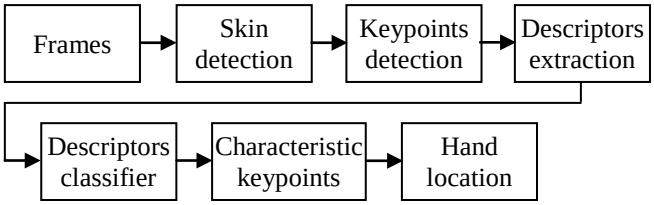


Figure 6. Detection of characteristic keypoints.

specific to this posture local image patterns. These local image patterns might be invariant under hand owner. Thus, the task is to find a way to separate these specific for hand local patterns from others.

Robust feature detection method should to localize similar keypoints on different hands. Robust feature extraction method should to compute descriptors which are invariant to image transformations (e.g., scaling and rotation). A pair of detector and extractor has good joint discriminative abilities if it is possible to train a classifier which is able to distinguish local features that are placed on the hands from the others. In other words, if classifier has good performance with some method of feature detection/extraction then the last one has good discriminative abilities.

3.2. Descriptors classification framework

First, we prepared dataset, which contains about 70 photos. A hand that has the fingers curled into the palm and the thumb retracted, displaying the knuckles shown on the each photo.For each photo we created mask of a region where the hand is located.Examples of a posture and ground truth mask are shown on the Figure 7.

Further, keypoints placed on hands were extracted from each picture and their descriptors were computed. We added big amount of examples of negative descriptors to the dataset. Negative samples collected from random indoor photos. We tuned feature detectors parameters so that they able to find about 40-50 keypoints on the hand in average.Each descriptor's variable treated as independent and values from one variable scaled to range [-1; 1].Data randomly partitioned to three subsets for performing of cross validation. We set the partition ratio as 65% for training subset, 15% of descriptors for validation subset and 20% for testing subset. The training dataset is used to adjust the weight of theclassifier. Validation subset used to minimize overfitting. Since we trained several classifiers with different amount of neurons in hidden layer we need to choose one best from them all. For this purpose separate test subset is used.

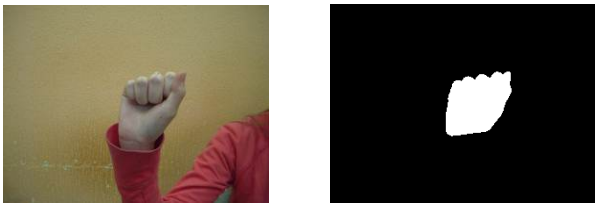


Figure 7. Photo of a hand and its mask.

Multilayer perceptron (MLP) with one hidden layer waschosen as classifier of the descriptors. Symmetric sigmoid was used as activation function. The steepness of activation function was set to 0.5. Initial weights assigned by Widrow-Nguen randomization method. MLP was learned through backpropagation with learning rate set to 0.7.We learned several networks for each type of descriptors. Size of the hidden layer varied from 32 up to 512 neurons. The learning process halted when the mean squared error on validation subset was starting togrow.

3.3. Performance measurement

Since neural network returns a real value within a range from -1 to 1, then ROC and PR curves might be plotted to compare performance of obtained neural networks.

PR curve [13], is similar to ROC curve and it shows the performance of a binary classifier as its discrimination threshold is varied. The Recall (horizontal axis) measure is the fraction of positive examples that are correctly labeled and it is the same as TPR. Precision (vertical axis) measures the fraction of examples classified as positive that are really true positive. PR curves can expose differences between algorithms that are not apparent in ROC space.

It can be seen from the Figures 8 and 9 that MLP performs better on descriptors which found and

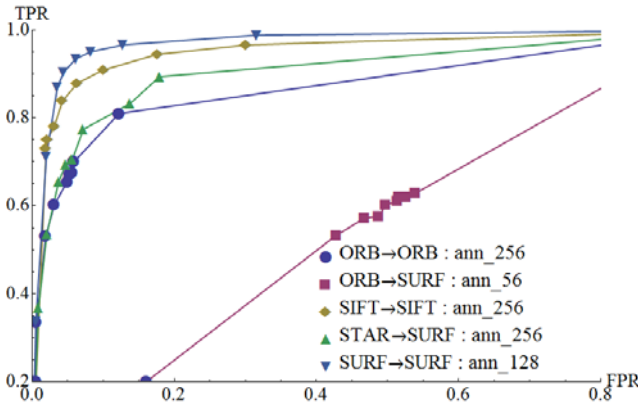


Figure 8. ROC curves for five types of classifiers of descriptors.

Table 2. Area under ROC curve (AUC) for different combination of detectors and descriptors.

Det.	ORB	ORB	SIFT	STAR	SURF
Ext.	ORB	SURF	SIFT	SURF	SURF
AUC	0.89625	0.56456	0.95615	0.9185	0.97158

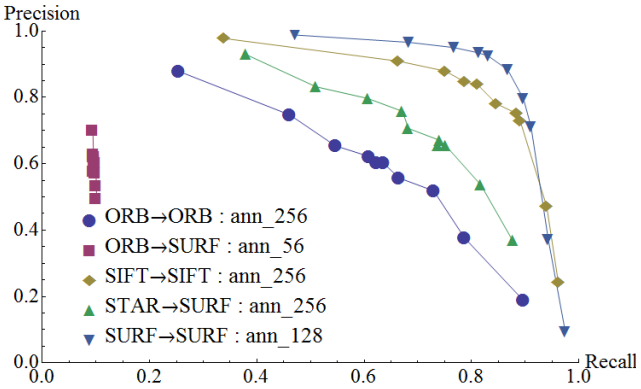


Figure 9. PR curves for five types of classifiers of descriptors.

extracted by SURF and SIFT methods.The results of the “ORB-ORB” and the “STAR-SURF” are less stable. Finally, it’s clearly seen that “ORB-SURF” combination is completely unsuitable for usage.

Local features classified as specific for hand using different detection approaches shown on the Figure 10. It can be seen that false detection occurs frequently with "ORB-SURF" feature detection approach. Local patterns found by “ORB-ORB” feature detector classified as positives in very rare cases. Other methods show better discriminative abilities. Nevertheless false detections happen with each type of feature detection, therefore boosting of such classifier with using skin tone detector will not be superfluous.

3.4. Hand detection

Hand detection is made simply by combination of our adaptive skin detection technique with described above local image features classifier.MLP classifier boosted by skin detector greatly decreases amount of false detections.We consider that the skin region is a hand if it contains more than 7 characteristic local features. This value is enough to discard most of the false positive descriptors.

4. Discussion

This study was primarily limited by small amount of tested combinations of local feature detectors and extractors of the descriptors. Observation of other local feature detectors can give higher quality of the results.

Superiority of SIFT and SURF methods can be explained by the fact that these detectors able to find much more keypoints than the other ones.Similar results of SIFT and SURF probably arises from the fact that they both use determinant of hessian matrix as measure for feature detection.

Applying Star and ORB detectors on real videos gives disappointing results.Perhaps these feature detectorsare not intendedfor such type of application.On the other hand such result might be interesting because these detectors weren't used before for purposes of hand detection or description.

Despite that threshold value for hand detection chosen high enough above the mean noise level sometimes we get false detections. It is good as first sketch but we need to look for more robust measure of hand appearance.

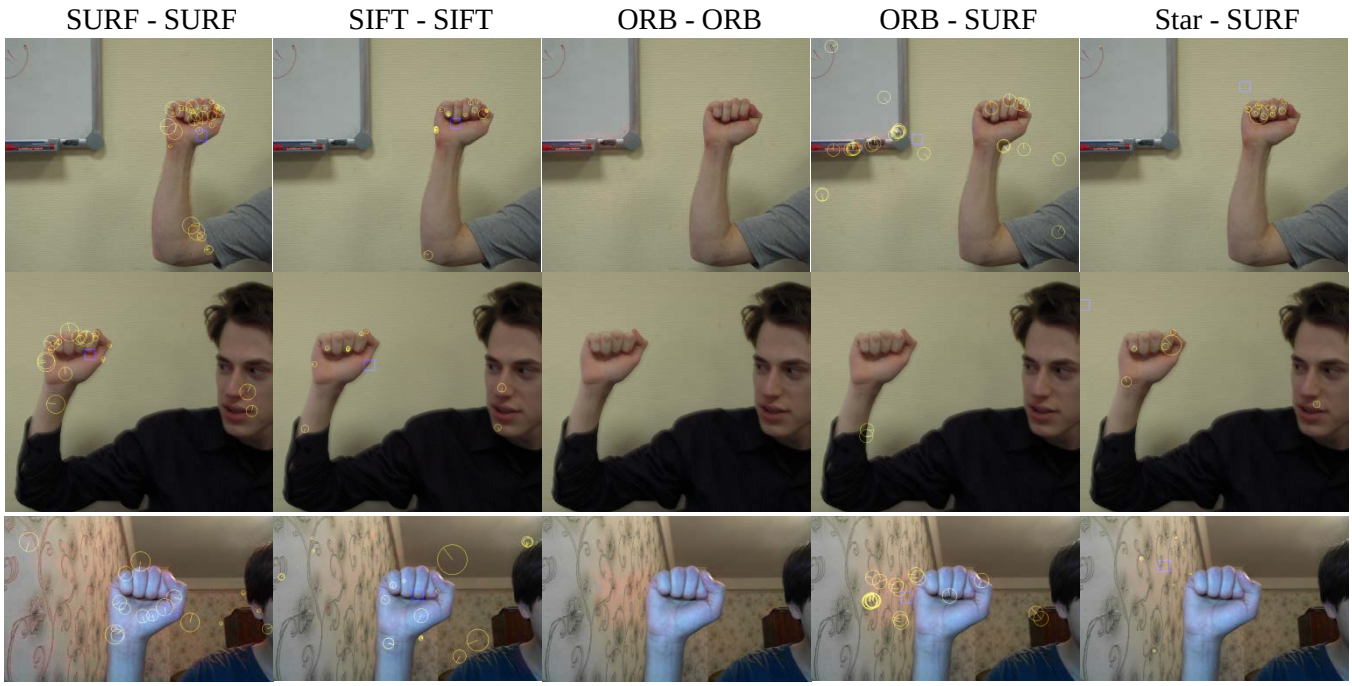


Figure 10. Results of classification of keypoints for three different input pictures. The yellow circles shows keypoints which were classified as characteristic for a hand. Photos in top row shooted in normal light conditions. Photos in the middle row were blurred during exposure. Photos in the bottom row are over-illuminatated.

5. Conclusions

From the results it is concluded that it's possible to build hand detection method based on classification of local features. It may be noted that SURF and SIFT approaches shows good discriminative abilities in sense of description of specific for a hands local features.

As a secondary result adaptive skin detection method was implemented. And it was shown that incrementally learning Bayesian classifier able to outperform non parametric skin detector.

References

[1]. ShipengXie, Jing Pan Hand Detection Using Robust Color Correction and Gaussian Mixture Model // 2011 Sixth International Conference on Image and Graphics (ICIG), 2011, pp.553-557.

[2]. *YingenXiong, Bing Fang, Francis Quek* Extraction of Hand Gestures with Adaptive Skin Color Models and Its Applications to Meeting Analysis // Eighth IEEE International Symposium on Multimedia (ISM'06), 2006, pp.647-651.

[3]. A. Mittal, A. Zisserman, P. H. S. Torr, Hand detection using multiple proposals British Machine Vision Conference, 2011

[4]. E. Ong and R. Bowden, "A boosted classifier tree for hand shape detection," in International Conference on

[5]. P. Kakumanu, S. Makrogiannis, N. BourbakisAsurveyofskin-colormodelinganddetectionmethods.

[6]. Z.Zivkovic, Improved adaptive Gausian mixture model for background subtraction, International Conference Pattern Recognition, UK, August, 2004

[7]. J. Kovac, P. Peer, and F. Solina, "Human skin color clustering for face detection," in EUROCON 2003. Computer as a Tool. The IEEE Region 8, 2003, vol. 2, pp. 144 – 148 vol.2.

[8]. L. Sigal, S. Sclaroff and V. Athitsos,Skin Color-Based Video Segmentation under Time-Varying Illumination, IEEE Transactions on Pattern Analysis and Machine Intelligence, 26(7), pp. 862-877, July 2004.

[9]. Herbert Bay, Andreas Ess, TinneTuytelaars, Luc Van Gool "SURF: Speeded Up Robust Features", Computer Vision and Image Understanding (CVIU), Vol. 110, No. 3, pp. 346--359, 2008

[10]. Lowe, David G. (1999). "Object recognition from local scale-invariant features". Proceedings of the International Conference on Computer Vision. 2. pp. 1150–1157

[11]. Ethan Rublee, Vincent Rabaud, Kurt Konolige, Gary R. Bradski: ORB: An efficient alternative to SIFT or SURF. ICCV 2011: 2564-2571.

[12]. AgrawalMotilal, Konolige Kurt, Blas Morten Rufus, "Censure: Center surround extremas for realtime feature detection and matching", Computer Vision–ECCV 2008, 102-115,2008,Springer Berlin Heidelberg.

[13]. Davis J., Goadrich M. The Relationship between Precision-Recall and ROC curves // Proc. Of 23 International Conference on Machine Learning, Pittsburgh, PA, 2006

Corresponding author: *Rasim Ahunzyanov*

Institution: *Saint-Petersburg State University of Information Technologies, Mechanics and Optics, S-Petersburg, Russia*

E-mail: *brotherofken@gmail.com*