Background-Invariant Robust Hand Detection based on Probabilistic One-Class Color Segmentation and Skeleton Matching

Andrey Kopylov, Oleg Seredin, Olesia Kushnir, Inessa Gracheva and Aleksandr Larin
Institute of Applied Mathematics and Computer Science, Tula State University, Tula, Russian Federation

Keywords: Hand Detection, One-class Classification, Pixel Color Classifier, Support Vector Data Description, Structure Transferring Filter, Skeleton Matching.

Abstract: In this paper we present a new method of hand detection in cluttered background for video stream processing. At first, skin segmentation is performed by one-class color pixel classifier which is trained using just a face image fragment without any background training sample. The modified version of one-class classifier is proposed. For each pixel it returns the grade (probability) of its belonging to the skin category instead of common binary decision. To adjust output of the one-class classifier the structure-transferring filter built on probabilistic gamma-normal model is applied. It utilizes additional information about the structure of an image and coordinates local decisions in order to achieve more robust segmentation results. To make a final decision whether an image fragment is the image of human hand or not, the method of binary image matching based on skeletonization is employed. The experimental study on segmentation and detection quality of the proposed method shows promising results.

1 INTRODUCTION

Accurate and robust hand detection in a video stream is necessary and 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. In spite of noticeable progress in this field there are still several challenges to be addressed, for instance, background clutter, illumination change, hand shape flexibility. It is especially important to tackle them in hand detection systems working in real-life environment.

It is possible to specify three main approaches to the solution to this problem: 1) the approach based on background subtraction techniques (Benezeth et al., 2008; Piccardi, 2004; Shiravandi et al., 2013), 2) the approach focused on skin-color segmentation (Jones and Rehg, 2002; Kakumanu et al., 2007; Phung et al., 2005; Vezhnevets et al., 2003; Wimmer and Munchen, 2005) and further hand detection with possible usage of additional information about hand shape peculiarities (Junqiu and Yagi, 2008), 3) the approach utilizing depth information (Suarez and Murphy, 2012).

Background subtraction techniques require unchangeable background within the chosen model. Such a serious drawback significantly limits a number of their applications in video stream processing.

Segmentation methods using in the second approach are based on parametric representation of color space domain (RGB, HSV, YCbCr) which corresponds to the color of skin. In particular, simple thresholding (Francke et al., 2007; Kakumanu et al., 2007; Vezhnevets et al., 2003), PCA-based ellipsoid descriptions (Hikal and Kountchev, 2011; Wimmer and Munchen, 2005) and Gaussian mixture models (Hassanpour et al., 2008; Jones and Rehg, 2002) could be applied. However, in real environment the geometry of skin color domain inside the color space could be changed dramatically. In addition, individual skin color characteristics vary from person to person which leads to the necessity of utilizing adaptive color models. Geometric features of hand shape are often used to improve the quality of segmentation. The contour of hand can be obtained by applying edge detection operators. The combination of shape, texture and color features may produce better performance (Junqiu and Yagi, 2008). There are several methods focusing on the hand morphology features, such as fingertips (Oka et al., 2002). One first common clue to fingertips detection is the curvature (Argyros and Lourakis, 2006), and another is template matching. Images of fingertips (Crowley et al., 1995) or fingers (Rehg and Kanade, 1995) could be used as templates.

Kopylov, A., Seredin, O., Kushnir, O., Gracheva, I. and Larin, A.
Background-Invariant Robust Hand Detection based on Probabilistic One-Class Color Segmentation and Skeleton Matching.
DOI: 10.5220/0006649805030510
In *Proceedings of the 7th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2018)*, pages 503-510
ISBN: 978-989-758-276-9

However, template matching has some inherent drawbacks: 1) it is computationally expensive; 2) it cannot cope with scaling or rotation changes of the target hand; 3) hand shape can vary dramatically due to its 22 degrees of freedom, so it could be difficult to choose the right template. See the comprehensive review of above-mentioned techniques in (Zhu et al., 2013).

Taking into account additional information about scene depth partly overcomes these disadvantages (Suarez and Murphy, 2012). Nevertheless, such approach assumes that no other objects are placed between the hand and the camera and requires additional equipment.

We propose to 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 (Jones and Viola, 2003). Similar idea is described in (Francke et al., 2007), but a simple single Gaussian model is used there for skin color description. In contrast, we apply here a modified version of the one-class pixel classifier (Larin et al., 2014) which does not need to collect any training data for background modeling. The advantage of one-class classifier over statistical threshold is that the region of interest in color space is described by more complex geometrical shape than just cuboids or ellipsoids, by means of Support Vector Data Description method (SVDD) (Tax and Duin, 2004). This fact lets minimizing the misclassification of skin-colored pixels. Re-training of the classifier could be performed in near real-time within almost every single frame of video stream.

The proposed segmentation method implies the further modification of one-class classifier. We use Bayesian approach on the basis of gamma-normal probabilistic model (Gracheva et al., 2015; Gracheva and Kopylov, 2017). Such a model lets us define and correct probabilistic relations between rough classification results of each individual element on a frame based on the structure of a source image.

Thereby, the color segmentation allows selecting a set of candidates (binary image fragments) for subsequent decision about their possible belonging to the class of hands, or hand detection.

It is worth noting that one of the most popular approaches to hand detection today is the method of Viola and Jones (Bowden, 2004; Fang et al., 2007; Jones and Viola, 2003). In general, this method is utilized for real-time detection of different kinds of objects in images. For example, it is implemented in OpenCV computer vision library for human face detection. However, it is necessary to collect large high-quality training set to apply Viola-Jones method to the hand detection problem. Taking into account the number of a human hand degrees of freedom, the generation of the proper training dataset becomes rather sophisticated task. In (Bowden, 2004) the dataset includes 5013 images of hands, divided into the training set consisting of 2504 images, and the test set consisting of 2509 images.

Our solution to the hand detection problem includes the shape skeleton matching method based on the pair-wise alignment of skeleton primitive chains (Kushnir and Seredin, 2015). The main advantage of the method is its invariance to the scaling and rotation changes of the target hand. The skeleton of a segmented binary object is described by the primitive chain, then the chain is compared with the skeleton primitive chain of the typical hand shape. If the dissimilarity measure is less then some predefined threshold, the object is classified as a hand. For the more robust classification result several typical hand shapes could be used.

Overall result of our hand detection method will be the binary mask corresponded to the hand region on the current video frame, coupled with the skeleton description of the hand shape. It allows solving gesture recognition or bio-identification tasks afterwards. The general flow-chart of the proposed hand detection method is shown in Figure 1.

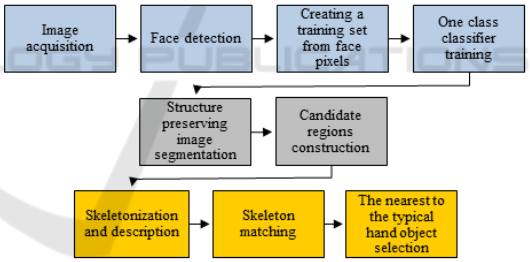

Figure 1: General flow-chart of the proposed hand detection method.

The paper is organized in the following way. In Section 2, the parametric representation of skin color by one-class classifier is described. The probabilistic method of skin segmentation based on such a parametric representation is introduced in Section 3. Section 4 describes the approach to hand detection based on the shape skeleton matching. Finally Section 5 is devoted to the experimental results and quality evaluation.

2 PARAMETRIC REPRESENTATION OF SKIN IN COLOR SPACE USING ONE-CLASS CLASSIFIER

The approach to segmentation based on skin color representation requires a pixel distribution model in the proper color space. In the real-life environment the individual peculiarities of a persons skin and illumination changes vary the color set of skin pixels completely.

To avoid these obstacles we use the modification of one-class pixel classifier (Larin et al., 2014) which does not use the background model. In contrast to static threshold decision, the one-class classifier allows describing more complex surface than cuboid, cylindrical or ellipsoid (Fritsch et al., 2002; Hsieh et al., 2012; Kim et al., 2005; Sabeti and Wu, 2007). This advantage is provided by using kernel trick in the SVDD (Tax and Duin, 2004) and allows minimizing skin-color pixels misclassification. The second benefit is the ability of real-time classifier re-training; we can adjust the color model in every single frame.

As a training set for the one-class color classifier we use a set of pixels in a rectangle located in the central part of face (see Figure 2(a)). Face detection is a well-developed technique (Degtyarev and Seredin, 2010) so this stage is performed in a rather simple way – we use the OpenCV realization of Viola-Jones method. The parameters of training region inside the face rectangle are defined as in (Degtyarev and Seredin, 2010). Particularly, the shift from the top of the face rectangle is equal to 0.49 of the face rectangle height, and the height and the width of training fragment are equal to 0.11 and 0.6 of the height and the width of the face rectangle correspondingly. This area lies between the eyes and nose tip, it is less deformable and free of beard, mustache, glasses, headdress, make-up.

This idea allows us to provide on-line estimation of skin color changes at every new frame, but requires a face presented in the scene (or reliable face detection). Otherwise, we can use a preliminary trained model.

The training set for the one-class classifier is a cloud of points in color-space. Flexibility of the describing surface in color space is defined by the parameter of Gaussian kernel in SVDD (see Figure 2(b)). After the training it is possible to make a fast decision about similarity between each pixel of an image and the training fragment checking whether a pixel belongs to the hypersphere in the Hilbert space.

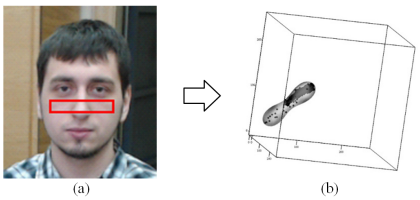

Figure 2: Parametric representation of skin: (a) training region; (b) training set of pixels (dark points) shown in RGB color space and the describing decision surface (light gray) built by the one-class classifier.

3 PROBABILISTIC IMAGE SEGMENTATION BASED ON PARAMETRIC REPRESENTATION OF SKIN

Parametric representation of skin by itself cannot provide robust and accurate segmentation (see Figure 3) since it is based on pixel properties in color space only and does not take into account any spatial relations between neighboring pixels as well as the structure of homogeneous regions inside the image.

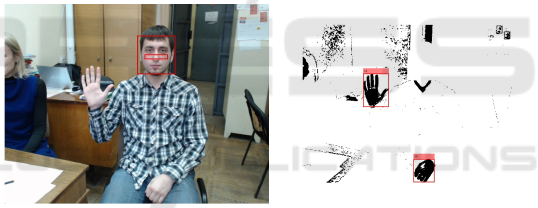

Figure 3: Results of color segmentation via one-class pixel color classifier. The segmentation inside the face bounding box was omitted.

Instead of well-known segmentation methods (e.g. described in (Bali and Singh, 2015)) which, unfortunately, needs too long processing time for online hand detection, we propose here to use the new class of filters, called structure-transferring filters which arise recently in the literature (He et al., 2013; Zhang et al., 2014). The main idea of structure-transferring filters is to extract the structure from the so-called guided image and make the filtering result of the source image consistent with this structure. Joint Bilateral Filter (Petschnigg et al., 2004) and Guided Filter (He et al., 2013) currently occupies a leading position among the filters of this class. The main disadvantage of the Joint Bilateral Filter and the Guided Filtration is the presence of artifacts, which visually manifested in the form of halos around the edges of the objects. The appearance of such artifacts is a characteristic of all filters with finite impulse response which is explained by the Gibbs effect. Recent attempts to overcome this

drawback (Zhang et al., 2014) using weighted global average parameters of corresponded model increase computational complexity beyond the real-time limits.

In this paper we use alternative Bayesian approach described in (Gracheva et al., 2015; Gracheva and Kopylov, 2017) and based on the special model of the Markov random field, called gamma-normal model (Krasotkina et al., 2010). It makes possible to take into account the structure which extracted from the "guide" image through setting an appropriate probabilistic relationships between elements of the sought for filtering result.

Let $Y = (y_t, t \in T)$ be an initial data array defined on a subset of the two-dimensional discrete space $T = \{t = (t_1, t_2) : t_1 = 1, ..., N_1, t_2 = 1, ..., N_2\}$ and let $X = (x_t, t \in T)$ defined on the same argument set plays the role of desirable result of processing. We will consider Y and X as the observed and hidden components of the two-component random field (X, Y). Probabilistic properties of two-component random field (X, Y) are completely determined by the joint conditional probability density $\Phi(Y|X, \delta)$ of original data $Y = (y_t, t \in T)$ with respect to the secondary data $X = (x_t, t \in T)$, and the a prior joint distribution $\Psi(X|\Lambda, \delta)$ of hidden component $X = (x_t, t \in T)$.

Let the joint conditional probability density $\Phi(Y|X, \delta)$ be in the form of Guassian distribution:

$$\Phi(Y|X,\delta) = \frac{1}{\delta^{(N_1N_2)/2}(2\pi)^{(N_1N_2)/2}} \times \times \exp(-\frac{1}{2\delta}\sum_{t\in T}(y_t - x_t)^2), \quad (1)$$

where δ is the variance of the observation noise, which is assumed to be unknown.

The a priori joint distribution $\Psi(X|\Lambda, \delta)$ of the hidden component $X = (x_t, t \in T)$ is also assumed Gaussian. But the variance r_t of hidden variables is assumed to be different at different points $t \in T$ of the hidden field X. It is convenient to make $r_t, t \in T$ proportional to the variance of observation noise $r_t = \lambda_t\delta$. Coefficients of proportionality $\Lambda = (\lambda_t, t \in T)$ can serve as a means to flexibly define the structure of probabilistic relationships between elements of the hidden field $X = (x_t, t \in T)$.

Under this assumption, we come to the improper a priori density:

$$\Psi(X|\Lambda,\delta) \propto \frac{1}{\left(\prod_{t\in T}\delta\lambda_t\right)^{1/2}(2\pi)^{(N_1N_2)/2}} \times \times \exp\left(-\frac{1}{2}\sum_{t',t''\in V}\frac{1}{\delta\lambda_t}(x_{t'} - x_{t''})^2\right), \quad (2)$$

where V is the neighborhood graph of the image elements having the form of a lattice.

Finally, we assume the inverse coefficients $1/\lambda_t$ to be a priori independent and identically gamma-distributed on the positive half-axis $\lambda_t \geq 0$.

$$G(\Lambda|\delta,\eta,\mu) = \exp\left[-\frac{1}{2\delta\mu}\sum_{t\in T}\left(\eta\frac{1}{\lambda_t} + \frac{1}{\eta}\ln\lambda_t\right)\right], \quad (3)$$

where η and μ is, respectively, the basic average factors of variability.

If $\mu \to 0$, then $1/\lambda_t$ is almost completely concentrated around the mathematical expectation $1/\eta$, and with $\mu \to \infty$, $1/\lambda$ tends to have the almost uniform distribution.

The joint a posteriori distribution of hidden field X and Λ is completely defined by (1), (2) and (3):

$$P(X,\Lambda|Y,\delta,\eta,\mu) = = \frac{\Psi(X|\Lambda,\delta)G(\Lambda|\eta,\mu)\Phi(Y|X,\delta)}{\int\int\Psi(X'|\Lambda',\delta)G(\Lambda'|\eta,\mu)\Phi(Y|X',\delta)dX'd\Lambda'}. \quad (4)$$

It is easy to see that the maximum a posteriori probability (MAP) estimate leads to the minimization of the following goal function.

$$J(X,\Lambda|Y,\eta,\mu) = \sum_{t\in T}(y_t - x_t)^2 + + \sum_{t',t''\in V}\left\{\frac{1}{\lambda_{t'}}\left[(x_{t'} - x_{t''})^2 + \eta/\mu\right] + (1 + 1/\mu)\ln\lambda_{t'}\right\}. \quad (5)$$

If the field of coefficients is fixed, optimal MAP estimation of X can be obtained by solving the following simple quadratic optimization task:

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

by using the extremely fast procedure on the basis of tree-serial dynamic programming (Mottl and Blinov, 1998).

If $X = (x_t, t \in T)$ is fixed $X = X^g$, criterion (5) gives the following equation for optimal Λ with fixed structural parameters η and μ:

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

The additional guided image can serve here as X^g to objectify its structure by $\hat{\Lambda} = (\hat{\lambda}_t, t \in T)$. As mentioned above, the field $\Lambda = (\lambda_t, t \in T)$ serves as a measure of local variability of a hidden field $X = (x_t, t \in T)$. As it can be seen from the criterion (5), $\hat{\lambda}_{t'}, t' \in T$ actually plays the role of a penalty for the difference between values of two corresponding neighboring variables $x^g_{t'}$ and $x^g_{t''}$, $(t', t'') \in T$. Thus, $\Lambda = (\lambda_t, t \in T)$, being estimated with the help of an additional guided image, can be used to transfer the structure of local relations between elements of the guided image to the result of processing.

The general scheme of segmentation is shown in Figure 4.

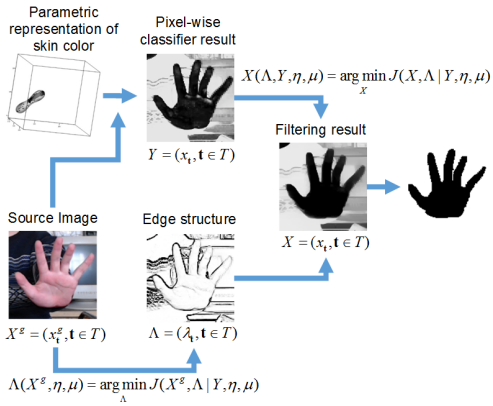

Figure 4: General scheme of probabilistic color segmentation.

Proposed procedure has approximately the same computational time as Fast Guided Filter (He et al., 2013) and has linear computational complexity with respect to the number of pixels in the source image.

For the image segmentation task the initial image plays the role of a "guide" image X^g and the rough output from the one-class classifier plays the role of initial data Y. In addition, the concept of SVDD gives the possibility to get fuzzy classification instead of binary one (see Figure 3). We use the distance from the center of hypersphere in Hilbert space to the object as the grade of membership in the class of interest. Fuzzy classification lets us obtain more accurate decision with the help of structure transferring filter, described above.

4 HAND DETECTION BY THE SKELETON MATCHING

Some regions of a binary image which are potentially can be referred to as hands according to their empirical features (such as geometrical characteristics, the size which is proportional to a face size, and the filled-in degree which is computed as the ratio of the number of black pixels to the size of the region) are presented for the skeleton matching procedure described in (Kushnir and Seredin, 2015). This procedure compares the skeleton of a candidate binary region with the skeleton of the typical hand shape. So far as the procedure is not invariant under reflection, we need to choose two typical hand shapes – for the left and the right hand (see Figure 5).

The skeletons of the typical left and right hands are needed to calculate once before detection will

Figure 5: Shapes of the most typical left and right hands.

start. Then, the skeleton of a candidate region is calculated. Each skeleton is transformed to the appropriate for the matching form (so called "base skeleton") by using pruning and approximation procedures (see Figure 6). The matching procedure performs three following steps:

1. each skeleton is coded by the sequence of primitives (primitive chain); each primitive contains information about topological characteristics of the corresponding skeleton edge (length, inner-angle and radial function),
2. two primitive chains (belonging to the one of typical hands and candidate region respectively) are aligned by dynamic programming procedure,
3. dissimilarity measure of primitive chains and, consequently, corresponding skeletons, is calculated based on their optimal pair-wise alignment.

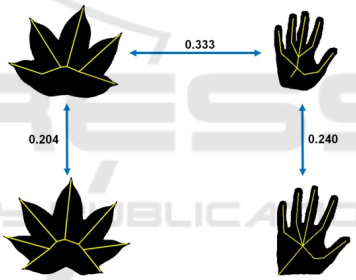

Figure 6: Examples of two classes of shapes, their base skeletons and pair-wise dissimilarity measures.

Thus, the matching procedure produces the distance function of two variables which returns non-negative dissimilarity measure between shapes. It is expected that classification decision (left/right hand or non-hand) could be made based on a simple threshold decision rule applied to dissimilarity value between this region and the image of typical hand. Therefore, it is important to choose typical images relevant to the application. In our preliminary studies, the procedure demonstrated a good processing time, near 3-5 ms per one matching in ordinary PC.

5 EXPERIMENTAL STUDY

For experimental study of the proposed method we collected a database of images gotten under varying illumination conditions and in different interiors. There is only one person with clearly

visible hand(s) in each image. The hand and the face bounding boxes have no intersections (see examples in Figure 7). Total number of images in the database is 549. For the each image ground-truth hand(s) location was determined by expert and marked with four points which are the corners of corresponding bounding box. The database is available on http://lda.tsu.tula.ru/papers/TulaSU_HandsDetDB.zip.

Figure 7: Examples of images used in experimental study.

On the first stage of experimental study we evaluated the segmentation quality of the algorithms proposed in Sections 2 and 3. The main goal was to figure out whether the segmentation method capable to detect candidates for the further comparison with the most typical hands shapes. As shown on Table 1, for the most of images, in which a face was detected (we used parameter NeighborFaces=9 in OpenCV realization of Viola-Jones algorithm), our segmentation method had found image fragment which corresponds to the ground-truth hand position; in 58 cases the only one candidate corresponds to the hand position. Decision about correspondence of candidate fragment to the ground-truth fragment was made if intersection of their bounding boxes was greater than 50%.

We have compared our segmentation method with well-known GrabCut method (Boykov and Kolmogorov, 2004) based on graph cuts and MRF optimization. To make a comparison more correct we have replaced Gaussian Mixture Models which were utilized in the paper for background and foreground representation, by our one-class classifier.

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

The examples of candidate fragments are shown in Figure 8. All objects-candidates were separated by expert into three categories: “left hands” (173 instances), “right hands” (117), and “non-hands” (593).

Table 1: Segmentation quality.

Parameter	Our Method	GrabCut-based Method
Number of images in Database	549	
Number of images in which face was found by Viola-Jones method	521	
Total number of candidates	883	
Average number of candidates per image	2.82	3.45
One of the several candidates corresponds to the ground-truth hand position	454	323
Number of cases in which a single candidate corresponds to the ground-truth	58	32
No candidates found	10	29

Moreover, two most typical hands (the left and the right) were chosen (see Figure 5). We estimated quality of skeleton matching algorithm (Section 4) by analyzing distances between objects-candidates obtained by segmentation procedure and two most typical hands (the left and the right).

Figure 8: Examples of objects-candidates, which are similar to human hand (top) and are not similar (bottom).

In Figure 9 we have a projection of whole distance matrix (883 x 883) on two-dimensional space. The first feature is the distance to the most typical left hand (horizontal axis) and the second one is the distance to the most typical right hand (vertical axis). Therefore, the “left hands” are the green rhombuses, “right hands” are the red squares and the non-hands are the black triangles. The chart displays that the compactness hypothesis for the three classes holds true. It allows us to build hand detection system using simple threshold rule based on the distance to the most typical objects.

Quality of separation of hands from non-hands in the form of ROC-curves is shown in Figure 10. These curves demonstrate the true positive against false positive rate at the increasing of distance function from the chosen typical objects. The AUC (area under the curve) for left hands (green curve) is equal to 0.9535, and for right hands (red curve) is 0.9531.

6 CONCLUSION

Hand detection is rather complex problem and requires a combination of different methods and algorithms for reliable solution. Proposed method consists in three major stages: face detection and one-class

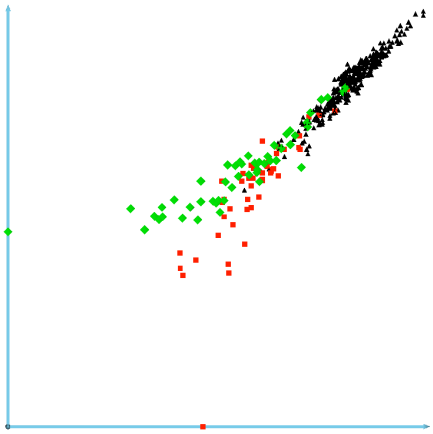

Figure 9: The distances to the most typical left hand (horizontal axis) and right hand (vertical axis).

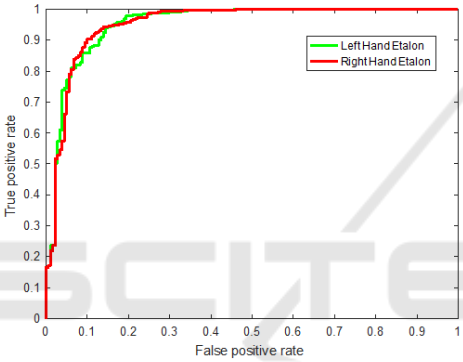

Figure 10: ROC-curves built at the distance function from the most typical left (green) and right hand (red).

classifier training, segmentation of skin-colored regions and comparison of hand-candidate shapes with the most typical hand objects.

The main advantage of the method is invariance to illumination conditions, personal skin color specificity, hand rotation and scale. Additionally, no image background model is used which makes our method performing robust in complex or varying scenes. Our method allows fast recalculating of the skin color model by using information just from the small fragment of face. The new method of probabilistic segmentation based on one-class color classifier and the structure-transferring filter was built on probabilistic gamma-normal model. It allows reducing the number of candidates for the shape matching procedure. Rotating/scaling-invariant matching algorithm based on skeleton primitive chains alignment performs very fast and obtains reliable classification results.

We have paid special attention to low computational time of constituent algorithms. Thereby, our method could be implemented in real-time and applied to hand detection in video streams. However, no temporal inter-frame relations are taking into account so far, that could be a subject for future work. The experimental study demonstrates robustness and precision of proposed hand detection method.

ACKNOWLEDGEMENTS

This work is supported by the Russian Fund for Basic Research, grants 16-57-52042, 16-07-01039. The results of the research project are published with the financial support of Tula State University within the framework of the scientific project No 2017-20PUBL.

REFERENCES

Argyros, A. A. and Lourakis, M. I. (2006). Vision-based interpretation of hand gestures for remote control of a computer mouse. *Computer Vision in Human-Computer Interaction*, 3979 LNCS:40–51.

Bali, A. and Singh, S. N. (2015). A review on the strategies and techniques of image segmentation. *2015 Fifth International Conference on Advanced Computing and Communication Technologies*, pages 113–120.

Benezeth, Y., Jodoin, P., Emile, B., Laurent, H., and Rosenberger, C. (2008). Review and evaluation of commonly-implemented background subtraction algorithms. *19th ICPR*, pages 1–4.

Bowden, R. (2004). A boosted classifier tree for hand shape detection. *Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004. Proceedings*, pages 889–894.

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

Crowley, J., Berard, F., and Coutaz, J. (1995). Finger tracking as an input device for augmented reality. *International Workshop on ldots*, (June):1–8.

Degtyarev, N. and Seredin, O. (2010). Comparative testing of face detection algorithms. *International Conference on Image and Signal Processing*, 6134 LNCS:200–209.

Fang, Y., Wang, K., Cheng, J., and Lu, H. (2007). A real-time hand gesture recognition method. *Multimedia and Expo, 2007 IEEE International Conference on*, pages 995–998.

Francke, H., Ruiz-del Solar, J., and Verschae, R. (2007). Real-time hand gesture detection and recognition using boosted classifiers and active learning. *Advances in Image and Video Technology*, pages 533–547.

Fritsch, J., Lang, S., Kleinehagenbrock, M., Fink, G. A., and Sagerer, G. (2002). Improving adaptive skin color

segmentation by incorporating results from face detection. *IEEE International Workshop on Robot and Human Interactive Communication*, pages 337–343.

Gracheva, I. and Kopylov, A. (2017). Image processing algorithms with structure transferring properties on the basis of gamma-normal model. *Communications in Computer and Information Science*, 661:257–268.

Gracheva, I., Kopylov, A., and Krasotkina, O. (2015). Fast global image denoising algorithm on the basis of non-stationary gamma-normal statistical model. *Communications in Computer and Information Science*, 542:71–82.

Hassanpour, R., Shahbahrami, A., and Wong, S. (2008). Adaptive gaussian mixture model for skin color segmentation. *World Academy of Science, Engineering and Technology*, 31(July):1–6.

He, K., Sun, J., and Tang, X. (2013). Guided image filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(6):1397–1409.

Hikal, N. A. and Kountchev, R. (2011). Skin color segmentation using adaptive pca and modified elliptic boundary model. *Advanced Computer Science and Information System (ICACSIS), 2011 International Conference*, pages 407–412.

Hsieh, C. C., Liou, D. H., and Lai, W. R. (2012). Enhanced face-based adaptive skin color model. *Journal of Applied Science and Engineering*, 15(2):167–176.

Jones, M. and Viola, P. (2003). Fast multi-view face detection. *Mitsubishi Electric Research Lab TR2000396*, (July).

Jones, M. J. and Rehg, J. M. (2002). Statistical color models with application to skin detection. *International Journal of Computer Vision*, 46(1):81–96.

Junqiu, W. and Yagi, Y. (2008). Integrating color and shape-texture features for adaptive real-time object tracking. *IEEE Transactions on Image Processing*, 2(17):235–240.

Kakumanu, P., Makrogiannis, S., and Bourbakis, N. (2007). A survey of skin-color modeling and detection methods. *Pattern Recognition*, 40(3):1106–1122.

Kim, K., Chalidabhongse, T. H., Harwood, D., and Davis, L. (2005). Real-time foreground-background segmentation using codebook model. *Real-Time Imaging*, 11(3):172–185.

Krasotkina, O., Kopylov, A., Mottl, V., and Markov, M. (2010). Bayesian estimation of time-varying regression with changing time-volatility for detection of hidden events in non-stationary signals. *7th IASTED International Conference on Signal Processing, Pattern Recognition and Applications*, pages 8–15.

Kushnir, O. and Seredin, O. (2015). Shape matching based on skeletonization and alignment of primitive chains. *Communications in Computer and Information Science*, 542:123–136.

Larin, A., Seredin, O., Kopylov, A., Kuo, S. Y., Huang, S. C., and Chen, B. H. (2014). Parametric representation of ob jects in color space using oneclass classifiers. *International Workshop on Machine Learning and Data Mining in Pattern Recognition. Springer, Cham.*, pages 300–314.

Mottl, V. and Blinov, A. (1998). Optimization techniques on pixel neighborhood graphs for image processing. *Graph-Based Representations in Pattern Recognition*, 12(Computing. Supplement, 0344-8029):135–145.

Oka, K., Sato, Y., and Koike, H. (2002). Real-time fingertip tracking and gesture recognition. *IEEE Computer Graphics and Applications*, 22(6):64–71.

Petschnigg, G., Szeliski, R., Agrawala, M., Cohen, M., Hoppe, H., and Toyama, K. (2004). Digital photography with flash and no-flash image pairs. *ACM Transactions on Graphics*, 23(3):664.

Phung, S., Bouzerdoum, A., and Chai, D. (2005). Skin segmentation using color pixel classification: analysis and comparison. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(1):148–154.

Piccardi, M. (2004). Background subtraction techniques: a review. *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583)*, 4:3099–3104.

Rehg, J. and Kanade, T. (1995). Model-based tracking of self-occluding articulated objects. *Proceedings of IEEE International Conference on Computer Vision*, pages 612–617.

Sabeti, L. and Wu, Q. M. J. (2007). High-speed skin color segmentation for real-time human tracking. *2007 IEEE International Conference on Systems, Man and Cybernetics*, pages 2378–2382.

Shiravandi, S., Rahmati, M., and Mahmoudi, F. (2013). Hand gestures recognition using dynamic bayesian networks. *2013 3rd Joint Conference of AI and Robotics and 5th RoboCup Iran Open International Symposium*, pages 1–6.

Suarez, J. and Murphy, R. R. (2012). Hand gesture recognition with depth images: A review. *Ro-Man, 2012 Ieee*, pages 411–417.

Tax, D. M. J. and Duin, R. P. W. (2004). Support vector data description. *Machine Learning*, 54(1):45–66.

Vezhnevets, V., Sazonov, V., and Andreeva, A. (2003). A survey on pixel-based skin color detection techniques. *Proceedings of GraphiCon 2003*, 85(0896-6273 SB - IM):85–92.

Wimmer, F. and Munchen (2005). Adaptive skin color classificator. *Proc. of the first ICGST International Conference on Graphics Vision and Image Processing GVIP-05*, (December):324–327.

Zhang, J., Cao, Y., and Wang, Z. (2014). A new image filtering method: Nonlocal image guided averaging. *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (2012):2479–2483.

Zhu, Y., Yang, Z., and Yuan, B. (2013). Vision based hand gesture recognition. *Service Sciences (ICSS), 2013 International Conference on*, 3(1):260–265.