EAI Endorsed Transactions
on Context-aware Systems and Applications Research Article EAI.EU

Context-aware hand poses classifying on images and video-sequences using a combination of wavelet transforms, PCA and neural networks

Phan Ngoc Hoang and Bui Thi Thu Trang*

Ba Ria-Vung Tau University, 80 Truong Cong Dinh street, Ward 3, Vung Tau city, Ba Ria-Vung Tau province, Vietnam

Abstract

In this paper we propose novel context-aware algorithms for hand poses classifying on images and video-sequences. The proposed hand poses classifying on images algorithm based on Viola-Jones method, wavelet transform, PCA and neural networks. On the first step, the Viola-Jones method is used to find the location of hand pose on images. Then, on the second step, the features of hand pose are extracted using combination of wavelet transform and PCA. Finally, on the last step, these extracted features are classified by multi-layer feed-forward neural networks. The proposed hand poses classifying on video-sequences algorithm based on the combination of CAMShift algorithm and proposed hand poses classifying on images algorithm. The experimental results show that the proposed algorithms effectively classify the hand pose in difference light contrast conditions and compete with state-of-the-art algorithms.

Keywords: Hand poses classifying, image processing, video processing, method Viola-Jones, CAMShift algorithm, wavelet transform, PCA, neural networks.

Received on $ULO, accepted on 0D\, published on -\

doi: 10.4108/_____

1. Introduction

Hand gesture recognition is one of the most difficult and required task in the field of image processing and computer vision. The hand gesture recognition systems are used to classify specific human hand gesture to transfer information or to manage devices, such as computers, televisions, etc. In this paper, the hand pose classifying on images and on video-sequences, which is main subtask of hand gesture recognition, is considered.

Classification hand pose on images can be done based on these following steps:

1. Detecting the location of hand pose on images;
2. Extracting the features of detected hand pose;
3. Classifying hand pose using extracted features.

Because of high processing speed and effectiveness, method Viola-Jones becomes one of the most used object detection methods. So, to detect the location of hand pose on images we use method Viola-Jones. This method based on three ingredients to enable fast and accurate object detection: the integral image for feature detection, Adaboost for feature selection and an attentional cascade for efficient computational resource allocation. These ingredients allow method can perform the object detection in real time [1–4].

The next step is extracting features of detected hand pose. In order to extract image features, wavelet transform is one of the most effective methods. It enables to obtain the necessary information about the image and it is also can be very quickly calculated. The experimental results of image classification algorithms [5–10] showed that images, features of which extracted by using wavelet transform, were classified with 76–99.7% accuracy rate.

In the algorithms [4, 11–20] wavelet transform is effectively used to solve the task of pattern recognition on

* Email:hoangpn285@gmail.com, trangbt.084@gmail.com

noisy images. In this case, the objects were recognized with 90–98.5% accuracy rate.

Besides the experimental results of algorithms [4, 16–20] showed that using combination of wavelet transform, PCA and neural networks gave more effective performance of object recognition. In these algorithms, neural networks were used to recognize objects based on their features, which extracted by using the combination of wavelet transform and PCA.

Thus, using the combination of Viola-Jones method, wavelet transform, PCA and neural networks is perspective solution for development of novel context-aware hand pose classifying algorithm on images. In this paper we propose a novel context-aware algorithm for hand pose classifying based on combination of Viola-Jones method, wavelet transform, PCA and neural networks. In this case, the context is any information about an image such as: image light condition, contour, noise and so on.

Classification hand pose on video-sequences can be done based on these following steps:

1. Detecting the location of hand pose on video-frame;
2. Tracking hand pose on video frame, used when hand pose is detected on previous frame;
3. Extracting the features of detected (tracked) hand pose;
4. Classifying hand pose using extracted features.

In 1998, Harry Bradsky created the algorithm CAMShift (Continuously Adaptive MeanShift) [26], which based on color information was able to effectively track objects in real time. So in this paper, we propose hand pose classifying algorithm on video-sequences based on combination of CAMShift algorithm and proposed hand pose classifying algorithm on images.

2. Proposed hand pose classifying algorithm on images

The proposed hand pose classifying algorithm on images consists of following main steps:

1. **Finding the hand pose location** on image based on Viola-Jones method (Fig. 1);

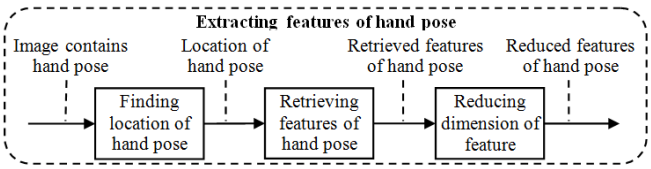

Figure 1. Process of extracting features of hand poses

2. **Retrieving the features** of hand pose using wavelet transform (Fig. 1);
3. **Reducing dimension** of extracted features vector based on PCA (Fig. 1);
4. **Training neural networks** using obtained feature vectors (Fig. 2);
5. **Classifying hand pose** based on obtained feature vectors and trained neural networks (Fig. 3).

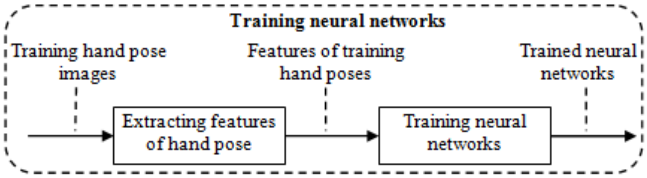

Figure 2. Process of training neural networks

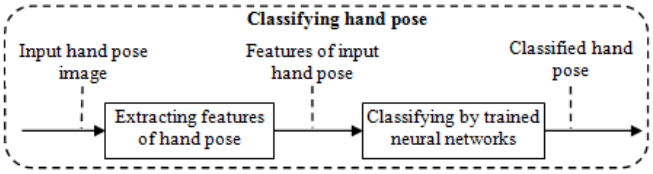

Figure 3. Process of classifying hand poses

2.1. Finding hand pose location using Viola-Jones Method

This method was developed and proposed in 2001 by Paul Viola and Michael Jones, and it is still effective to detect object in digital images and videos in real-time [1, 2]. Using simple cascade classifier, which is the feature detector instead of one complex classifier, is the main idea of this method. Based on this idea, it enables to construct a detector, which can work in real time.

Integral image

In Viola-Jones method, integral image is used to rapidly compute rectangle features. The integral image is widely used in other methods, such as wavelet transforms, SURF, Haar filtering and etc. [21]. Pixel value of the integral image at location (x, y) contains the sum of pixels above and to the left of (x, y) and is computed by formula (1).

$$I(x,y) = \sum_{x' \le x, y' \le y} i(x', y'), \qquad (1)$$

where $I(x, y)$ is value of integral image pixel (x, y); $i(x, y)$ – intensity of original image pixel (x, y). Each pixel value of integral image $I(x, y)$ is sum of the original pixels from $i(0, 0)$ to $i(x, y)$. Time of computation of integral image matrix depends on the number of pixels of original image. Value of each pixel of integral image can be computed by formula (2):

$$I(x,y) = i(x,y) - I(x-1,y-1) + I(x,y-1) + I(x-1,y). \qquad (2)$$

Haar-like features

Haar-like features are image features, which are used in the object recognition task. Viola and Jones adapted the idea of using an alternate feature set based on Haar wavelets instead of the usual image intensities of Papageorgiou et al. [22]. And they developed the new

features called Haar-like features. A Haar-like feature considers adjacent rectangular regions at a specific location in a detection window, sums up the pixel intensities in each region and calculates the difference between these sums.

In the detection phase of the Viola–Jones object detection framework, a window of the target size is moved over the input image, and for each subsection of the image the Haar-like feature is calculated. This difference is then compared to a learned threshold that separates non-objects from objects. Because such a Haar-like feature is only a weak learner or classifier (its detection quality is slightly better than random guessing) a large number of Haar-like features are necessary to describe an object with sufficient accuracy. Examples of Haar-like features are presented in Fig. 4.

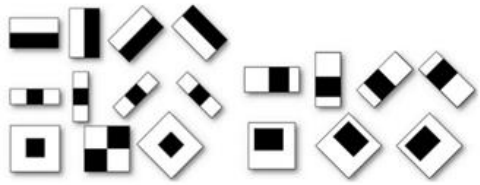

Figure 4. Examples of Haar-like features

Learning classification using Adaboost

Boosting is a machine learning meta-algorithm for performing supervised learning. Boosting is based on the question posed by Kearns [23]: can a set of weak learners create a single strong learner? A weak learner is defined to be a classifier which is only slightly correlated with the true classification (it can label examples better than random guessing). In contrast, a strong learner is a classifier that is arbitrarily well-correlated with the true classification.

Schapire's affirmative answer to Kearns' question has had significant ramifications in machine learning and statistics, most notably leading to the development of boosting [24].

For each feature, the weak learner determines the optimal threshold classification function, such that the minimum number of examples is misclassified. A weak classifier *hj*(*x*) thus consist of a feature *fj*, a threshold *θj* and a parity *pj* indicating the direction of the inequality sign (formula 3):

$$h_j(z) = \begin{cases} 1, & if\ \ p_j f_j(z) < p_j \theta_j \\ 0, & \text{otherwise} \end{cases}, \qquad (3)$$

where *z* is a 24×24 pixel sub-window of an image.

Development of this approach was development more perfect family algorithms of a boosting – AdaBoost, short for Adaptive Boosting, is a machine learning algorithm, formulated by Yoav Freund and Robert Schapire. It is a meta-algorithm, and can be used in conjunction with many other learning algorithms to improve their performance. AdaBoost is adaptive in the sense that subsequent classifiers built are tweaked in favour of those instances misclassified by previous classifiers.

For combining increasingly more complex classifier in a "cascade" which allows background regions of the image to be quickly discarded while spending more computation on promising object-like regions.

2.2. Extracting hand pose features using Wavelet transforms

By using wavelet transform to extract image features, we will obtain the necessary information about the image. Besides we can also quickly calculate the wavelet transform. So wavelet transform becomes one of the most effective methods, which are used to extract image features to classify (recognize) objects [4–20].

In this paper, after hand pose location in image is found by using method Viola-Jones, the Haar and Daubechies wavelet transforms are used to extract hand pose image features. The process of extracting hand pose features by using wavelet transform works as follows. Firstly, the hand pose image is resized to 64×64 pixels. Then we apply wavelet transform to obtained image and extract the low-frequency wavelet coefficients. In the result, we have matrix that consists of 32×32 = 1024 low-frequency wavelet coefficients (Fig. 5).

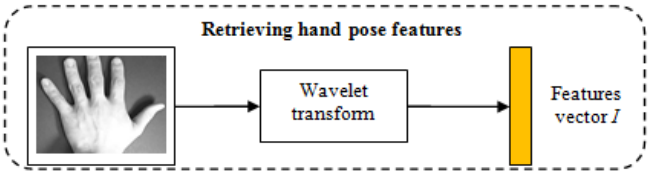

Figure 5. Retrieving hand pose features using wavelet transform

2.3. Extracting hand pose features using Wavelet transforms

Before classifying by neural networks, dimension of hand pose feature vector is reduced. In this paper, PCA is used to solve this task. At first, eigenspace for hand poses (eigenhandpose) will be created using *M* images of hand poses. The process of creating hand pose eigenspace is carried out as follows.

In first step, the process of extracting features is applied to each of *M* images. After that we obtain a set of $\vec{I}_1,...,\vec{I}_M$ feature vectors. Then we form the mean vector, the value of each element of which is calculated by the formula (4):

$$\vec{I}_{avg} = \frac{1}{M}\sum_{n=1}^{M}\vec{I}_n. \ . \qquad (4)$$

In second step, each vector of the *M* feature vectors is subtracted by mean vector using formula (5):

$$\vec{\Phi}_n = \vec{I}_n - \vec{I}_{\text{cp}}, n = 1,...,M. \ . \qquad (5)$$

In third step, an eigenspace, which consists of K eigenvectors of the covariance matrix C (6), is created. It is the best way to describe the distribution of these M feature vectors ($K<M$).

$$C = \frac{1}{M}\sum_{n=1}^{M}\vec{\Phi}_n\vec{\Phi}_n^T = AA^T, \quad A = \{\vec{\Phi}_1,...,\vec{\Phi}_M\}. \quad (6)$$

where k-th vector $\vec{u}_k$ satisfies maximization of the following formula (7):

$$\lambda_k = \frac{1}{M}\sum_{n=1}^{M}(\vec{u}_k^T\vec{\Phi}_n)^2 \quad (7)$$

and an orthogonality condition (8):

$$\vec{u}_l^T\vec{u}_k = \begin{cases}1, & l = k\\ 0, & \text{otherwise}\end{cases}. \quad (8)$$

Vectors $\vec{u}_k$ and values λ_k are eigenvectors and eigenvalues of covariance matrix C. In order to create this eigenspace, firstly, we calculate M eigenvectors $\vec{u}_l$ of covariance matrix C by using eigenvectors of other matrix $L = A^TA$. Each vector $\vec{u}_l$ is calculated by the formula (9):

$$\vec{u}_l = \frac{1}{M}\sum_{k=1}^{M}v_{lk}\Phi_k, l = 1,\ldots,M. \quad (9)$$

After that we select K eigenvectors, which have the largest eigenvalues from M obtained eigenvectors. The eigenspace is the set of K selected eigenvectors (Fig. 6).

When the hand pose eigenspace is created, the process of reducing dimension of hand pose feature vector $\vec{I}_{\text{in}}$ is carried out as follows.

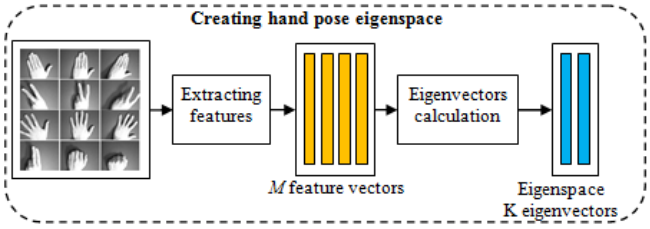

Figure 6. Creation of hand pose eigenspace

Firstly, we decompose the hand pose feature vector on K eigenvectors $\vec{u}_i$ and calculate corresponding decomposition coefficients by the formula (10):

$$w_i = \vec{u}_i^T(\vec{I}_{\text{in}} - \vec{I}_{\text{avg}}), i = 1,...,K. \quad (10)$$

Then we form a novel hand pose feature vector using formula (11):

$$\vec{\Omega}^T = \{w_1,...,w_K\}. \quad (11)$$

This vector describes the distribution of each eigenvectors in presentation of hand pose feature vector. The novel hand pose feature vector is $\vec{\Omega}$, which consists of K elements. In this case, number K is much less than 1024 (Fig. 7).

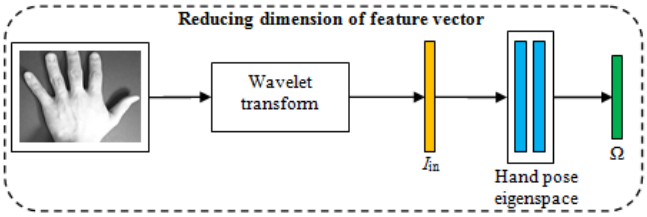

Figure 7. Reducing dimension of hand pose feature vector

2.4. Hand pose classifying using neural networks

In this proposed algorithm paper, we use back-propagation feed-forward neural networks to classify hand poses based on obtained feature vectors. For each hand pose of training set, we create one multi-layered feed-forward neural network, which is trained by back propagation method.

The input of these neural networks is the hand pose feature vector $\vec{\Omega}$ (11), which consists of K elements. These neural networks will return a value from 0 to 1, which determine whether an input hand pose is training hand pose or not.

The neural networks classify the input hand pose as follows. Firstly, feature vector of the input hand pose is extracted. After that the dimension of this vector is reduced. Finally, obtained hand pose feature vector is submitted to the inputs of all trained neural networks. Input hand pose is classified as a hand pose of training set, neural network of this hand pose returns the largest value (Fig. 8.).

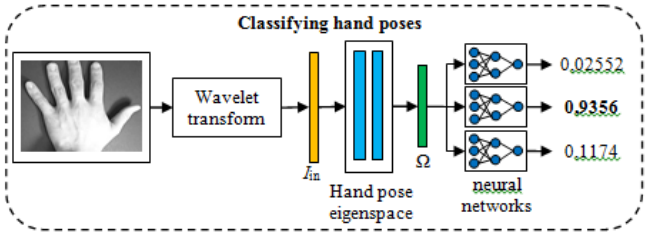

Figure 8. Classifying hand poses

3. Proposed hand pose classifying algorithm on video-sequences

The proposed hand pose classifying algorithm on video-sequences consists of following main steps:

1. **Finding the hand pose location** on video frame based on Viola-Jones method;
2. **Tracking hand pose location** on video frame using CAMShift algorithm if hand pose location is

detected on previous video frame. In another case, go back to step 1 (Fig. 9).

3. **Retrieving the features** of hand pose using wavelet transform;
4. **Reducing dimension** of extracted features vector based on PCA;
5. **Classifying hand pose** based on obtained feature vectors and trained neural networks.

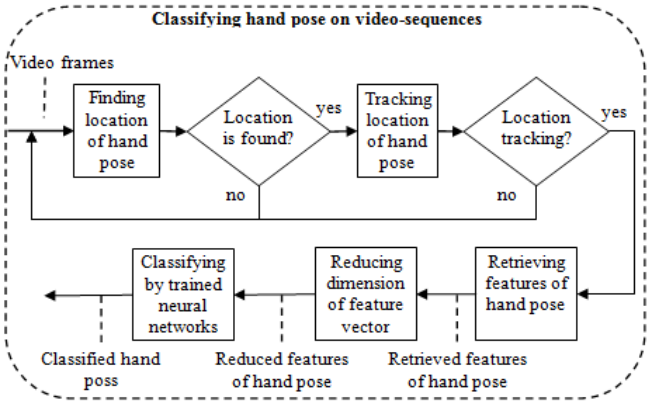

Figure 9. Classifying hand poses on video-sequences

4. Experimental results

All experiments were performed on a laptop with the processor Intel Core Duo P7350 2.0 GHz and 2.0 GB of RAM.

4.1. Classifying hand poses on images

The proposed algorithm of classifying hand poses on images was tested using a part of the Cambridge Gesture database [25]. This hand pose database consists of 5 difference parts, which contain images in various light contrast conditions (Fig. 10).

Figure 10. Examples of hand pose images of 5 difference parts

In the part 1 (Fig. 10a), the light is straight ahead the hand pose. The light comes from bottom right corner of the hand pose for part 2 (Fig 10b), top right corner – part 3 (Fig. 10c), top left corner – part 4 (Fig. 10d) and bottom left corner – part 5 (Fig. 10e).

In these experiments hand poses are divided into 12 classes presented on Fig. 11. For each part, we created one testing dataset, which contains 2400 hand pose images (20 images of each class). And for each part we also created one training dataset, which contains 1200 hand pose images (10 images of each class).

Figure 11. Examples of images of 12 classes of hand pose of dataset part 1

The experimental results are presented in table 1. Column P1 is presented classifying results for dataset part 1 and so on. It is shown that the proposed hand pose classifying algorithm, which based on a combination of wavelet transform, PCA and neural networks, gave more accurate classifying results than algorithm [20].

Table 1. Accuracy rate of hand pose classifying

Wavelet transform type	P1, %	P2, %	P3, %	P4, %	P5, %	All, %
[20] (Haar)	94,63	90,96	89,46	92,33	90,17	93,30
[20] (Db)	93,67	90,17	87,58	90,79	87,63	92,57
Proposed (Haar)	96,75	92,34	90,58	94,15	91,53	94,96
Proposed (Db)	95,49	91,40	88,69	92,32	88,75	93,88

The highest hand pose classifying accuracy was obtained for the dataset part 1, in which the light is straight ahead the hand pose. For other parts, the classifying accuracy is competed with each other. Besides, it is shown that in this case, using wavelet Haar gave more effective classifying results than using wavelet Daubechies.

4.2. Classifying hand poses on video-sequences

The proposed algorithm of classifying hand poses on video-sequences was tested using created data set, consisting of 6 classes of hand poses. Each hand pose is used to present a number from zero to 5 (Fig. 12).

Figure 12. Examples of 6 classes using for hand poses classification on video-sequences

The experimental results showed that proposed algorithm effectively classify hand poses on video-sequences with accuracy rate about 93% and real time processing speed – 30 frames per second. Examples of hand poses classification on video-sequences are presented in Fig. 13.

Figure 13. Examples of hand poses classification on video-sequences

5. Conclusions

In this paper we developed novel algorithms for hand pose classifying on images and on video-sequences based on wavelet transform, PCA and neural networks. Developed algorithms enables effectively classifying hand pose with difference light contrast.

The developed algorithm for classifying hand poses on images gave the highest accuracy rate 96,75%, which was obtained for the dataset part 1. In this part, the light is straight ahead hand pose. The experimental results also showed that using wavelet Haar gave more accuracy rate of hand pose classifying than using wavelet Daubechies.

The developed algorithm for classify hand poses on video-sequences performed with real time processing speed and gave the accuracy rate about 93%.

References

[1] Viola P., Jones M.J. (2001) Rapid object detection using a boosted cascade of simple features // *IEEE Conf. on Computer Vision and Pattern Recognition*. Kauai, Hawaii, USA, V. 1, pp. 511–518.

[2] Viola P., Jones M.J. (2004) Robust real-time face detection // *International Journal of Computer Vision*, V. 57, No. 2. pp. 137–154.

[3] Yi-Qing Wang, (2014) An Analysis of the Viola-Jones Face Detection Algorithm // *Image Processing On Line*, No. 4, pp. 128–148.

[4] Phan N.H., Bui T.T.T., Spitsyn V.G. (2013) Real-time hand gesture recognition base on Viola–Jones method, algorithm CAMShift, wavelet transform and principal component analysis // *Tomsk State University Journal of Control and Computer Science*, No 2(23), pp. 102–111.

[5] Mehdi, L., Solimani, A., Dargazany, A. (2009) Combining wavelet transforms and neural networks for image classification. In: *41st Southeasten Symposium on System Theory*, Tullahoma, TN, USA, pp. 44–48.

[6] Weibao, Z., Li, Y. (2007) Image classification using wavelet coefficients in low-pass bands. In: *Proceedings of International Joint Conference on Neural Networks*, Orlando, Florida, USA, pp. 114–118.

[7] Chang, T. Jay, K. (1993) Texture analysis and classification with tree-structured wavelet transform. In: *IEEE Trans. Image Processing*, vol. 2, no. 4, pp. 429–440.

[8] Daniel, M.R.S., Shanmugam, A. (2011) ANN and SVM based war scence classification using wavelet features: a comparative study. In: *Journal of Computational Information Systems*, pp. 1402–1411.

[9] Park, S.B., Lee, J.W., Kim, S.K. (2004) Content-based image classification using a neural network. *Pattern Recognition Letters*, pp. 287–300.

[10] Gonzalez, A.C., Sossa, J.H., Riveron, E.M.F. (2006) Histograms, wavelets and neural networks applied to image retrieval. In: *Proceedings of the 5th Mexican international conference on Artificial Intelligence: Lecture Notes in Computer Science*, vol. 4293, pp. 820–827.

[11] Lai, J.H., Yuen, P.C., Feng, G.C. (2001) Face recognition using holistic Fourier invariant features. *Pattern Recognition*, vol. 34, pp. 95–109.

[12] Kakarwal, S., Dsehmuhk, R.: Wavelet transform based feature extraction for face recognition. Informatica, vol. 15, no. 2, pp. 243–250 (2004).

[13] Zhang, B.-L., Zhang, H. (1995) Face recognition by applying wavelet subband representation and kernel assosiative memory. *IEEE Trans. Image Process.*, vol. 4, no. 11, pp. 1549–1560.

[14] Gumus, E., Kilic, N., Sertbas, A., Ucan, O.N. (2010) Evulation of face recognition techniques using PCA, wavelets and SVM. *Expert Systems with Application*, vol. 37, pp. 6404–6408.

[15] Wadkar, P.D., Wankhade, M. (2012) Face recognition using discrete wavelet transform. *International Journal of Advanced Egineering Technology*, vol. III, iss. I, pp. 239–242.

[16] Mazloom, M., Kasaei, K. (2005) Face recognition using PCA, wavelets and neural networks. In: *Proceeding of the First International Conference on Modeling*, Simulation and Applied Optimization, Sharjah, UAE, pp. 1–6, February 1–3.

[17] Phan N.H., Bui T.T.T., Spitsyn V. G., Bolotova Y. A. (2016) Using a Haar wavelet transform, principal component analysis and neural networks for OCR in the presence of impulse noise // *Journal Computer Optics*, T 40, No 2, pp. 249–257.

[18] Phan N.H., Bui T.T.T. (2016) Context-aware Handwritten and Optical Character Recognition Using a Combination of Wavelet transform, PCA and Neural Networks // *Context-Aware Systems and Applications, LNICST*, Vol 165, Springer, pp. 254–263.

[19] Phan N.H., Bui T.T.T., Spitsyn V. G., Bolotova Yu. A., Savitsky Yu. V. (2015) Development of algorithms for face and character recognition based on wavelet transforms, PCA and neural networks // *Proceedings of Control and Communications (SIBCON), 2015 International Siberian Conference, IEEE*.

[20] Phan N.H., Bui T.T.T., Spitsyn V. G.: Face and Hand Gesture Recognition based on Wavelet Transforms and Principal Component Analysis // 7th International Forum on Strategic Technology IFOST: Proceedings of IFOST 2012, IEEE, (2012).

[21] Gonzalez R. C., Woods R. E. (2001) Digital image processing. *Reading MA // Addison-Wesley*

[22] Papageorgiou C., Oren M., Poggio T. (1998) A general framework for object detection // *International Conference on Computer Vision*.

[23] Kearns M. (1988) Thoughts on Hypothesis Boosting // *Unpublished manuscript in Machine Learning class project*

[24] Freund Y., Schapire R.E. (1999) A Short Introduction to Boosting // *Journal of Japanese Society for Artificial Intelligence*, vol.14, no. 5, pp. 771–780

[25] Kim T.K., Wong S.F., Cipolla R.: Cambrige Hand Gesture Data set [Online]. Available: http://www.iis.ee.ic.ac.uk/~tkkim/ges_db.htm.

[26] Bradski G. R. (1998) Computer vision face tracking for use in a perceptual user interface // *Intel Technology Journal*. 1998, 2nd Quarter.