Context-Aware Hand Pose Classifying Algorithm Based on Combination of Viola-Jones Method, Wavelet Transform, PCA and Neural Networks

Ngoc Hoang Phan(✉) and Thi Thu Trang Bui

Faculty of Information Technology, Ba Ria-Vung Tau University,
Truong Van Bang Street 01,
Vung Tau City, Ba Ria-Vung Tau Province, Vietnam
hoangpn285@gmail.com, trangbt.084@gmail.com
{hoangpn,trangbtt}@bvu.edu.vn

Abstract. In this paper we propose a novel context-aware algorithm for hand poses classifying. The proposed algorithm based on Viola-Jones method, wavelet transforms, PCA and neural networks. At first, the Viola-Jones method is used to find the location of hand pose in images. Then the features of hand pose are extracted using combination of wavelet transform and PCA. Finally, these extracted features are classified by multi-layer feedforward neural networks. In this proposed algorithm, for each training hand pose we create one neural network, which will determine whether an input hand pose is training hand pose or not. In order to test the proposed algorithm, we use known Cambridge Gesture database and divide it into 5 parts with difference light contrast conditions. The experimental results show that the proposed algorithm effectively classifies the hand pose in difference light contrast conditions and competes with state-of-the-art algorithms.

Keywords: Hand poses classifying · Method Viola-Jones · Wavelet transform · PCA · Neural networks

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. The main aim of these systems is to transfer information or to manage difference devices, such as computers, televisions, etc. In this paper, the hand pose classifying task, which is one main subtask of hand gesture recognition, is considered.

In order to classify the hand pose in images, we can do these following steps:

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

P. Cong Vinh et al. (Eds.): ICCASA 2016, LNICST 193, pp. 42–51, 2017.
DOI: 10.1007/978-3-319-56357-2_5

To detect and find the location of hand pose in images we use method Viola-Jones. Because of high processing speed and effectiveness, method Viola-Jones becomes one of the most used object detection methods. 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].

After location of hand pose is detected, the next step is its features extraction. In order to extract image features, wavelet transform is one of the most effective methods. Wavelet transform enables to obtain the necessary information about the image and it is also can be 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 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. 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.

2 Proposed Algorithm

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

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

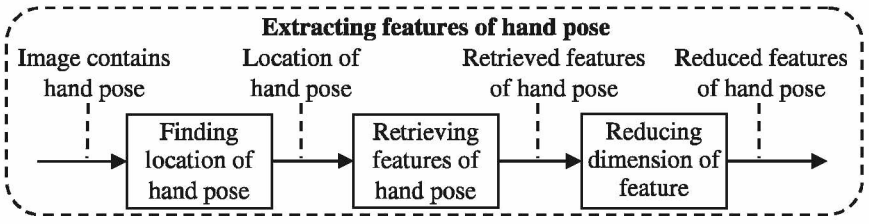

Fig. 1. Process of extracting features of hand poses.

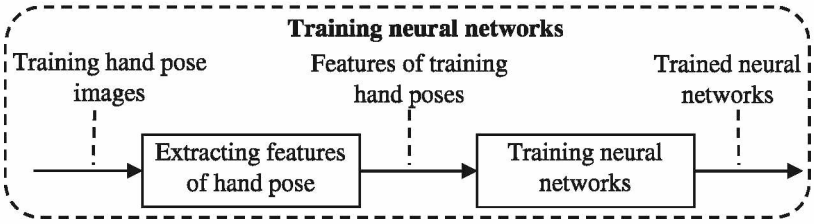

Fig. 2. Process of training neural networks.

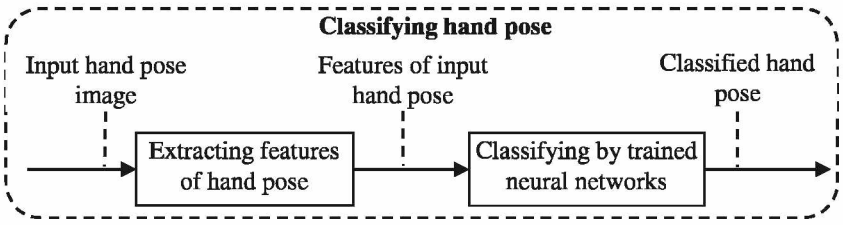

Fig. 3. Process of classifying hand poses.

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).

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

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):

$$I(x,y) = i(x,y) - I(x-1,y-1) + I(x,y-1) + I(x-1,y). \tag{2}$$

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.

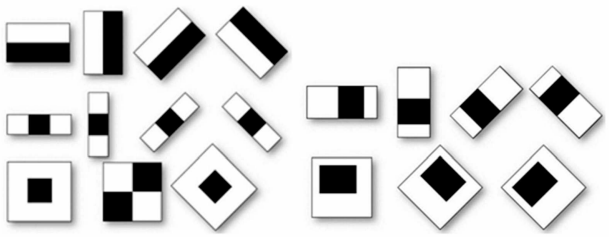

Fig. 4. Examples of Haar-like features.

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 j and a parity pj indicating the direction of the inequality sign (formula 3):

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

where z is a 24 $\times$ 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 favor 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 $32 \times 32 = 1024$ low-frequency wavelet coefficients (Fig. 5).

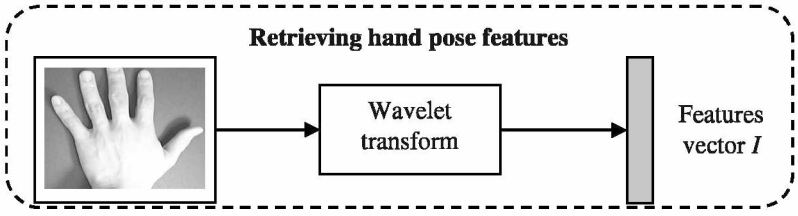

Fig. 5. Retrieving hand pose features using wavelet transform.

2.3 Dimension Reduction Using PCA

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 $\vec{I}_1, \ldots, \vec{I}_M$ feature vectors. Then we form the mean vector, the value of each element of which is calculated by the formula (4):

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

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

$$\vec{\Phi}_n = \vec{I}_n - \vec{I}_{\text{cp}},\ n = 1,\ldots,M. \tag{5}$$

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 ($K < M$).

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

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

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

and an orthogonality condition (8):

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

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):

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

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.

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

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

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

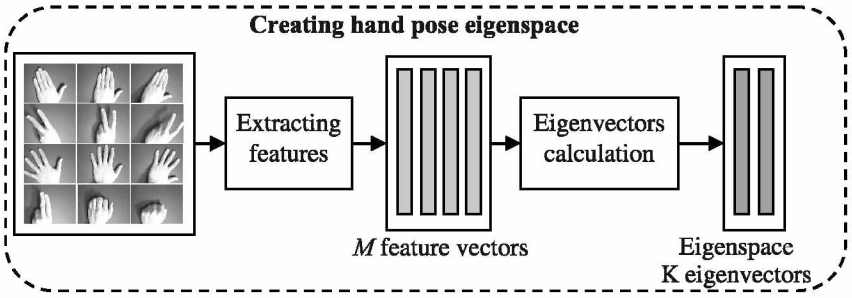

Fig. 6. Creation of hand pose eigenspace.

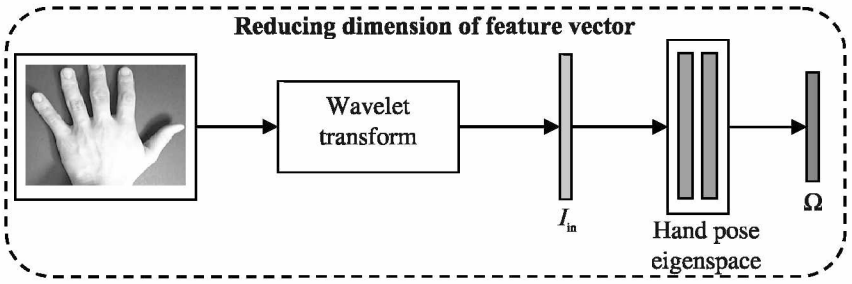

Fig. 7. Reducing dimension of hand pose feature vector.

$$\vec{\Omega}^T = \{w_1, \ldots, w_K\}. \tag{11}$$

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, K is much less than 1024 (Fig. 7).

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 multilayered 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 which returns the largest value (Fig. 8).

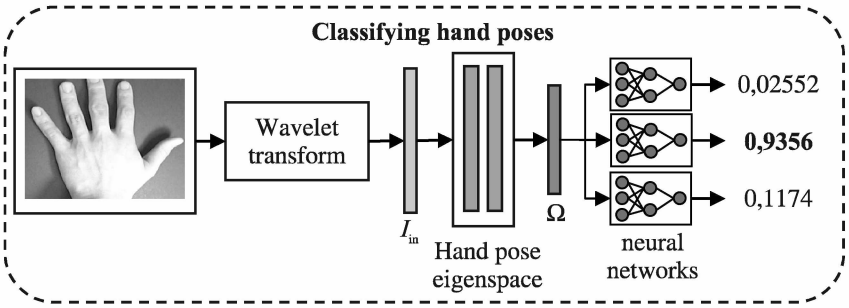

Fig. 8. Classifying hand poses.

3 Experimental Results

The proposed algorithm was tested using a part of the Cambridge Gesture database [25]. All experiments were performed on a laptop with the processor Intel Core Duo P7350 2.0 GHz and 2.0 GB of RAM. This hand pose database consists of 5 difference parts, which contain images in various light contrast conditions (Fig. 9).

Fig. 9. Examples of hand pose images of 5 difference parts

In the part 1 (Fig. 9a), the light is straight ahead the hand pose. The light comes from bottom right corner of the hand pose for part 2, top right corner – part 3 (Fig. 9c), top left corner – part 4 (Fig. 9d) and bottom left corner – part 5 (Fig. 9e).

Fig. 10. Examples of images of 12 classes of hand pose of dataset part 1.

In these experiments hand poses are divided into 12 classes presented on Fig. 10. 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).

The experimental results are presented in Table 1. 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	Part 1, %	Part 2, %	Part 3, %	Part 4, %	Part 5, %	All parts, %
[20] (Haar)	94,63	90,96	89,46	92,33	90,17	93,30
[20] (Daubechies)	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 (Daubechies)	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 Conclusion

In this paper we developed a novel algorithm for hand pose classifying based on method Viola-Jones, wavelet transform, PCA and neural networks. Developed algorithm enables effectively classifying hand pose with difference light contrast conditions.

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

References

1. Viola, P., Jones, M.J.: Rapid object detection using a boosted cascade of simple features. In: IEEE Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii, USA, vol. 1. pp. 511–518 (2001)
2. Viola, P., Jones, M.J.: Robust real-time face detection. Int. J. Comput. Vision **57**(2), 137–154 (2004)
3. Wang, Y.-Q.: An analysis of the Viola-Jones face detection algorithm. Image Process. On Line **4**, 128–148 (2014)
4. Phan, N.H., Bui, T.T.T., SpitsynVladimir, G.: Real-time hand gesture recognition base on Viola-Jones method, algorithm CAMShift, wavelet transform and principal component analysis. Tomsk State Univ. J. Control Comput. Sci. **2**(23), 102–111 (2013)
5. Mehdi, L., Solimani, A., Dargazany, A.: Combining wavelet transforms and neural networks for image classification. In: 41st Southeasten Symposium on System Theory, Tullahoma, TN, USA, pp. 44–48 (2009)
6. Weibao, Z., Li, Y.: Image classification using wavelet coefficients in low-pass bands. In: Proceedings of International Joint Conference on Neural Networks, Orlando, Florida, USA, pp. 114–118 (2007)

7. Chang, T., Jay, K.: Texture analysis and classification with tree-structured wavelet transform. IEEE Trans. Image Process. **2**(4), 429–440 (1993)
8. Daniel, M.R.S., Shanmugam, A.: ANN and SVM based war scene classification using wavelet features: a comparative study. J. Comput. Inf. Syst. **7**, 1402–1411 (2011)
9. Park, S.B., Lee, J.W., Kim, S.K.: Content-based image classification using a neural network. Pattern Recogn. Lett. **25**, 287–300 (2004)
10. Gonzalez, A.C., Sossa, J.H., Riveron, E.M.F., Pogrebnyak, O.: Histograms, wavelets and neural networks applied to image retrieval. In: Gelbukh, A., Reyes-Garcia, C.A. (eds.) MICAI 2006. LNCS (LNAI), vol. 4293, pp. 820–827. Springer, Heidelberg (2006). doi:10.1007/11925231_78
11. Lai, J.H., Yuen, P.C., Feng, G.C.: Face recognition using holistic Fourier invariant features. Pattern Recogn. **34**, 95–109 (2001)
12. Kakarwal, S., Dsehmuhk, R.: Wavelet transform based feature extraction for face recognition. Informatica **15**(2), 243–250 (2004)
13. Zhang, B.-L., Zhang, H.: Face recognition by applying wavelet subband representation and kernel associative memory. IEEE Trans. Image Process. **4**(11), 1549–1560 (1995)
14. Gumus, E., Kilic, N., Sertbas, A., Ucan, O.N.: Evaluation of face recognition techniques using PCA, wavelets and SVM. Expert Syst. Appl. **37**, 6404–6408 (2010)
15. Wadkar, P.D., Wankhade, M.: Face recognition using discrete wavelet transform. Int. J. Adv. Eng. Technol. **3**(1), 239–242 (2012)
16. Mazloom, M., Kasaei, K.: Face recognition using PCA, wavelets and neural networks. In: Proceeding of the First International Conference on Modeling, Simulation and Applied Optimization, Sharjah, UAE, 1–3 February, pp. 1–6 (2005)
17. Phan, N.H., Bui, T.T.T., Spitsyn, V.G., Bolotova, Y.A.: Using a Haar wavelet transform, principal component analysis and neural networks for OCR in the presence of impulse noise. J. Comput. Opt. **40**(2), 249–257 (2016)
18. Phan, N.H., Bui, T.T.T.: Context-aware handwritten and optical character recognition using a combination of wavelet transform, PCA and neural networks. In: Vinh, P.C., Alagar, V. (eds.) ICCASA 2015. LNICSSITE, vol. 165, pp. 254–263. Springer, Cham (2016). doi:10.1007/978-3-319-29236-6_25
19. Phan, N.H., Bui, T.T.T., Spitsyn, V.G., Bolotova Yu, A., Savitsky Yu, V.: Development of algorithms for face and character recognition based on wavelet transforms, PCA and neural networks. In: Proceedings of 2015 International Siberian Conference on Control and Communications (SIBCON). IEEE (2015)
20. Phan, N.H., Bui, T.T.T., Spitsyn, V.G.: Face and hand gesture recognition based on wavelet transforms and principal component analysis. In: 7th International Forum on Strategic Technology IFOST: Proceedings of IFOST 2012. IEEE (2012)
21. Gonzalez, R.C., Woods, R.E.: Digital Image Processing. Addison-Wesley, Reading (2001)
22. Papageorgiou, C., Oren, M., Poggio, T.: A general framework for object detection. In: International Conference on Computer Vision (1998)
23. Kearns, M.: Thoughts on hypothesis boosting. Unpublished manuscript in Machine Learning Class Project (1988)
24. Freund, Y., Schapire, R.E.: A short introduction to boosting. J. Japan. Soc. Artif. Intell. **14**(5), 771–780 (1999)
25. Kim, T.K., Wong, S.F., Cipolla, R.: Cambridge Hand Gesture Data set. http://www.iis.ee.ic.ac.uk/~tkkim/ges_db.htm