Wireless Video Surveillance System Based on Incremental Learning Face Detection

Wenjuan Liao[2], Dingheng Zeng[1], Liguo Zhou[2], Shizheng Wang[3], and Huicai Zhong[1]

[1]Institute of Microelectronics of Chinese Academy of Sciences, Beijing, China 100029
[2]Chinese Academy of Sciences R&D Center for Internet of Things, Wuxi, China 214135
[3]School of Electrical & Electronic Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798
{wenjuan.liao,zhou.liguo}@outlook.com,
{zengdingheng,zhonghuicai}@ime.ac.cn, shizhengwang@ntu.edu.sg

Abstract. As an important supplement to wired video in video surveillance applications, wireless video has taken increasing attentions and has been extensively applied into projects like "Safe City". Despite of in taxis, buses, emergency command vehicles, or temporary monitory point, there will definitely produce massive surveillance videos. In order to retrieve and browse these videos in an efficient way, key video browsing system and technique based on face detection is accepted and put into promotion. Face detection is widely studied and used in many practical applications; however, because of the distinct features of different factors in experiments and applications, such as orientation, pose, illumination, etc., challenges usually obstruct the practical usage. To perform successful application in wireless video browsing system, this paper proposes an incremental learned face detection method based on auto-captured samples. Experiments demonstrate that our proposed incremental learning algorithm has favorable face detection performance and can work in the proposed system.

Keywords: Face detection, incremental learning, wireless video, video browsing.

1 Introduction

As one of the most novel applications in video surveillance, wireless video has attracted increasing concerns from industry and academy. In fact, wireless video is selected as the applicable solution in most "Safe City" projects of China, for example, Shenzhen has invested more than 280 million RMB to install wireless video surveillance system for over 13,000 buses in the city, while Dongguan has invested more than 110 million RMB for over 10,000 buses. Such applications provide an efficient solution of real-time storage and remote browsing for video surveillance under unfixed scene.

After the successful storage of massive surveillance video, how to retrieve and browse these videos in an efficient and fast way has become another challenge in addition to effective storage for surveillance video. For this reason, this paper mainly discusses the face detection based on incremental learning in a wireless video brows-

X. He et al. (Eds.): MMM 2015, Part I, LNCS 8935, pp. 118–127, 2015.

ing system. This system adopts scalable coding technique [22] to encode the video with CIF resolution and 9 FPS into the base layer of surveillance video, which is transferred to remote monitoring center in a wireless manner with low-delay for real-time retrieval and browsing; meanwhile, regular video with D1 resolution and 25 FPS is encoded into complete bitstream of base layer and enhance layer, and then stored in local server of camera terminal. When there is retrieval or browsing task, videos with high resolution and frame rate will be transferred to monitoring center in a high-delayed way for browsing or checking, and the transmission mount can therefore be efficiently controlled when retrieving or browsing wireless videos. Wherein, the face detection technique oriented to wireless video is the basic supporting technique for this system.

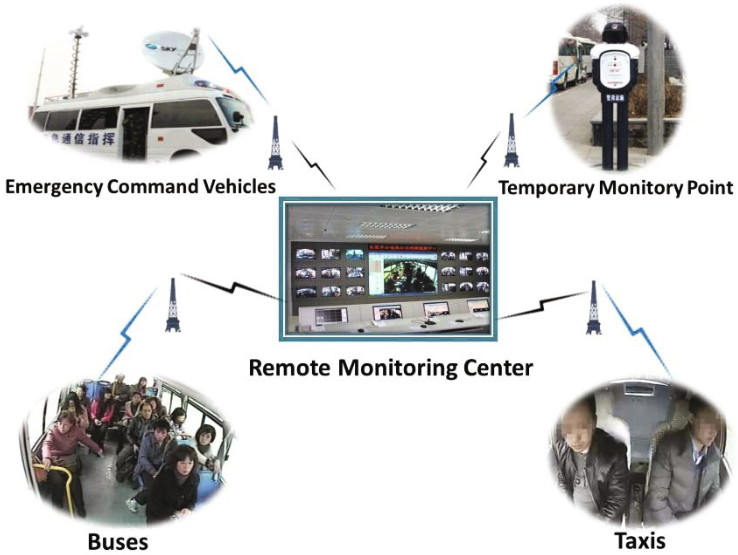

Fig. 1. The schematic diagram of wireless video system

Face detection is one of the hottest research topics in the field of computer vision; its main difficulty lies in many variations of orientation (in-plane rotation), pose (out-of-plane rotation), illumination, etc. In order to solve the problems above and reach the required speed for practical application, researchers have studied in aspects of robust feature selection, classifier improvement, detection speed promotion, and adaption enhancement. The field of face detection has significant development in the last decade. Particularly, the creative face detection algorithm based on Adaboost by Viola and Jones [1] enables the feasibility of face detection for robustness and high-speed in practical application. [2] has proposed a structured rectangle feature, [3], [4], and [5] have introduced LBP, HOG, and Edgelet feature into face detection application, respectively. [6] adds anti-spoofing technique into face detection so as to achieve reliable living face detection under the condition of low illumination as well. Since a single cascade of Haar feature can fully accomplish the frontal or near-frontal face detection task rather than straightforwardly expand to multi-view face detection. Therefore, [7] represents the parallel cascade to learn an individual classifier for each view. [2] also uses a classifiers combination with pyramid structure to solve the multi-pose problem. [8] has proposed a similar tree structure used for face detection with different orientations. In [9], a compressed-domain face detection algorithm based on neural network is proposed. [17] proposes a face image enhancement algorithm for face detection under low resolution. In addition, [10], [11], and [12] have attempted new methods to improve detection speed from the angle of feature, detector, and detection process improvement, respectively. Finally, besides different learning methods based on batch algorithm above, [13] has presented an incremental learning face detection method applicable to collision with respirator and sunglasses as well as extreme illumination change. While a threshold-based incremental learning classifier is proposed for gender recognition, face detection, and human detection [14].

However, as for wireless video, limited by view, pose, and illumination, video quality is usually low. Furthermore, no appropriate training samples can be used offline. Hence, the incremental face detection algorithm for wireless video is one of our focusing works. Our main contributions include: (1) specific to the problem of lacking offline samples with wireless video, we propose an automatic new sample selection algorithm based on region modeling, face detection, and face tracking to automatically add samples for online incremental learning; (2) merging with tree-Adaboost face detection algorithm, we propose an learning method that learns part of typical views (e.g., frontal face and 90 degree profile face) offline and then incrementally learns positive and negative samples of each view online; (3) as for the requirement of retrieval and browsing application for wireless video, we design and realize a wireless video retrieval and browsing system based on face detection, which can store detected faces for retrieving and playbacking corresponding original video according to specified face.

The rest of this paper is organized as follows. The proposed wireless video face detection method and system based on incremental learning are described in details in section 2. Some experimental results on the standard test database and scene-collected video are given to demonstrate the performance of the proposed system in section 3. Finally, section 4 concludes this paper.

2 Proposed Method and System

In this paper, oriented to the practical requirement of retrieval and browsing application for wireless video surveillance, an incremental learning based face detection and browsing system is proposed: first, an initial face detector is trained offline with standard face database; then, by region modeling, face detection, and face tracking, face tubes are extracted and some of face images are automatically selected as positive samples, while negative samples are similarly extracted from background automatically based on background modeling in selected region; next, on the basis of these new samples selected online, we perform incremental learning and online update on offline trained primary classifiers to further improve the detection performance of classifiers; finally, complete the storage of detected face images so as to retrieve and browse surveillance video according to the content and tag of face image. The functional flow diagram of face detection based on incremental learning is shown in Fig. 2.

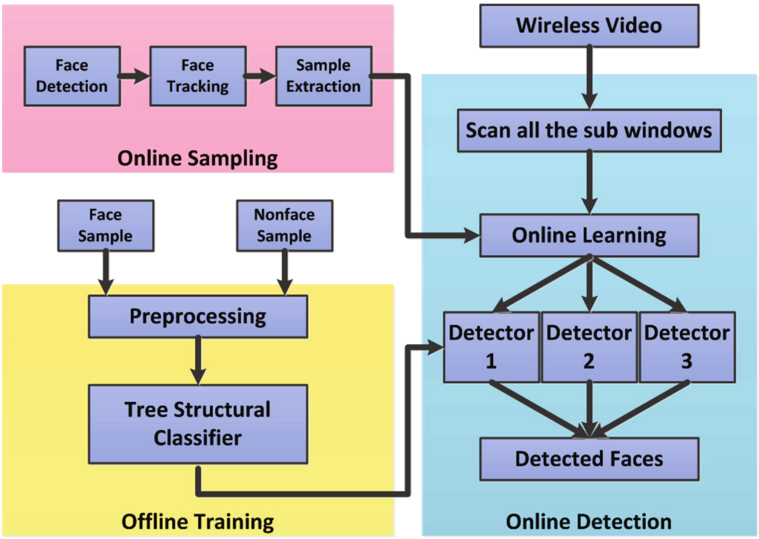

Fig. 2. The framework of incremental learning algorithm

2.1 Sample Auto-selection with Online Face Tracking

An online face tracking method facing to head rotation and pose variation is introduced in this section. As for wireless video camera, compared with fixed camera its position is unfixed, but for a certain time period (e.g., contingency command post) or a particular region (e.g., inside of buses), the background of surveillance video is relatively fixed. Therefore, Gaussian modeling method can be used for background subtraction and face tracking [15]. A gaussian mixture model (GMM) with K component densities at time t can be modeled as follow:

$$P(X_t|Y) = \sum_{k=1}^{K} w_{k,t} \cdot I(X_t, \mu_{k,t}, \sigma_{k,t}) \tag{1}$$

Where X_t denotes the appearances of each modeling pixel in the video sequence and Y denotes the background or tracked foreground tracked. $I(X_t, \mu_{k,t}, \sigma_{k,t})$ denotes the k-th Gaussian component with mean vector $\mu_{k,t}$ and covariance matrix $\sigma_{k,t}$. $w_{k,t}$ is the corresponding weight. The appearance information is the pixel value when the GMM model is applied to background modeling and face tracking.

In the algorithm of this paper, we apply background modeling, face detection and face tracking into wireless video. Through background modeling, we can narrow the range of face detection into foreground, and detect faces with offline trained face detector irrespective of false negatives rate. Then, we track the face and add the face that can be tracked but not detected into positive samples for incremental learning, while add the false alarmed face images in background region and randomly extracted images from background into negative samples for incremental learning as well.

In the process of tracking, new color information of tracked face can be obtained in each frame. Let x_{t+1} be the new color information of tracked face Y obtained at time t + 1. It can be modeled as:

$$P(x_{t+1}|Y) = \sum_{k=1}^{K'} w'_{k,t+1} \times I(x_{t+1}, \mu'_{k,t+1}, \sigma'_{k,t+1}) \tag{2}$$

Model $p(x_{t+1}|Y)$ is used to update $p(X_t|Y)$ into $p(X_{t+1}|Y)$. However, some components of the distribution should not be used for the update, such as those belonging to occluding objects and the background. The distribution distances between these and each component of old model $p(X_t|Y)$ are usually great. The components of $p(x_{t+1}|Y)$, which have big Mahalanobis distances to $p(X_t|Y)$, are dropped in the updating process. Only those components of $p(x_{t+1}|Y)$, which have small Mahalanobis distances to $p(X_t|Y)$, are used to treat the face from the same people, and add them to incremental learning samples when they are not detected by original detector. However, in each of the period that detector cannot detect the faces, only the first 5 samples are used for incremental learning to avoid drifting of the samples.

2.2 Online Update by Selected Sample

This paper exploits Gentle Adaboost algorithm [16] for learning and combination of each weak classifier. The goal of boosting is to combine many weak classifiers h(x) into an additive strong classifier $H(x)$:

$$H(x) = \sum_{k=1}^{K} \alpha_k h_k(x) \tag{3}$$

Where α_k are scalar weights of k classifiers. There have been many boosting algorithms proposed to learn this model; typically, this is done in a greedy algorithm where weak classifiers are trained sequentially. After each weak classifier is trained, the training examples are re-weighted so that examples misclassified previously should receive more weight.

Traditional algorithms are usually batch learning [1], training will be restarted once a new sample incoming. To this end, we propose a new sample learning method to incrementally change the existing models and classifiers. Based on the sample auto-selection method mentioned above, classifiers are capable of fully automatic online learning with new samples.

Since face detection is regarded as the problem of binary classification, in order to clearly introduce the Adaboost method combining both offline and online, we focus on the design of weak classifiers based on Gaussian distribution at first: assuming face image belongs to ω_1 class while nonface image is ω_2 class. When we adopt the same Gaussian distribution used in [14] to model each weak classifier, the distribution of features is determined by only two parameters: mean u and variance σ^2. Therefore, the classifier threshold can also be updated by Algorithm 1.

Algorithm 1: Online Learning

Input: x, one-dimensional feature.

Data: N , feature number; u_i , mean; σ_i^2 , variance; g_i , minimum-error-rate classification; η, learning rate; $x(n)$, value of the n-th inputting sample on one-dimensional feature x.

Output: threshold θ.

1. Given one-dimensional feature x and its classification w_i.

   Get its mean $u_i = \frac{1}{N}\sum\limits_{x \in \omega_i} x$ and variance $\sigma_i^2 = \frac{1}{N}\sum\limits_{x \in \omega_i} (x - u_i)^2$.

2. For $g_i(x) = \ln[p(\omega_i|x)] = \ln p(x|\omega_i) + \ln P(\omega_i)$

   Then $g_i(x) = -\frac{1}{2}(\frac{x-u_i}{\sigma_i})^2 - \frac{1}{2}\ln 2\pi - \frac{1}{2}\ln \sigma^2 + \ln P(\omega_i)$

   For $g(x) = g_1(x) - g_2(x) = 0$

   (a) If $\sigma_1 = \sigma_2$

   $$\theta = x = \frac{(u_1 + u_2)}{2}$$

   (b) If $\sigma_1 \neq \sigma_2$

   $$\theta = \frac{-(\sigma_2^2 u_1 - \sigma_1^2 u_2)^2 \pm \sqrt{(\sigma_2^2 u_1 - \sigma_1^2 u_2)^2 - (\sigma_1^2 - \sigma_2^2)(\sigma_1^2 u_2^2 - \sigma_2^2 u_1^2 + \sigma_1^2 \sigma_2^2 \ln(\sigma_2^2/\sigma_1^2))}}{(\sigma_1^2 - \sigma_2^2)}$$

3. Add samples

   Update for m

   $u(n) = (1-\eta)u(n-1) + \eta x(n)$

   $\sigma^2(n) = (1-\eta)\sigma^2(n-1) + \eta(x(n) - u(t))^2$

The error bound e (Bhattacharyya bound) for the classifier is given as follows [18]:

$$e = \exp(-k(1/2)) \tag{4}$$

$$k(1/2) = \frac{1}{4}\frac{(u_+ - u_-)^2}{\sigma_+ + \sigma_-} + \frac{1}{2}\ln\frac{\sigma_+ + \sigma_-}{2\sqrt{\sigma_+ \sigma_-}} \tag{5}$$

Where (u_+, σ_+) and (u_-, σ_-) are the mean and squared variance of the positive class (ω_1) and negative classes (ω_2), respectively.

If $e_1 < e_2 < ... < e(n) < Th < ...$, we take the weak classifiers when error e is less than Th to update the threshold of corresponding classifier. Where Th is 0.8 in this paper.

3 Experimental Results

To verify the performance of method and system proposed in this paper, we adopt standard testing set MITEx face database [19], CAS-PEAL-R1 standard face database [20], and wireless video captured on buses and taxis in the procedure of project implementation to test the detection performance of incremental learning method and effectiveness of real system when retrieving and browsing wireless video, respectively.

3.1 The Effectiveness Test of Our Method

We use MITEx face database to evaluate the effectiveness and reliability of our incremental learning based face detector compared with traditional offline learned detector [1] and other incrementally learned detector [14]. The training set contains 800 face images and 1,100 nonface images of this database, in which 900 samples are randomly selected as incremental samples to gradually add into training set, while testing set contains 1,906 faces and 3,281 nonfaces with the same size of 20*20. Both positive and negative samples in this experiment are manually added.

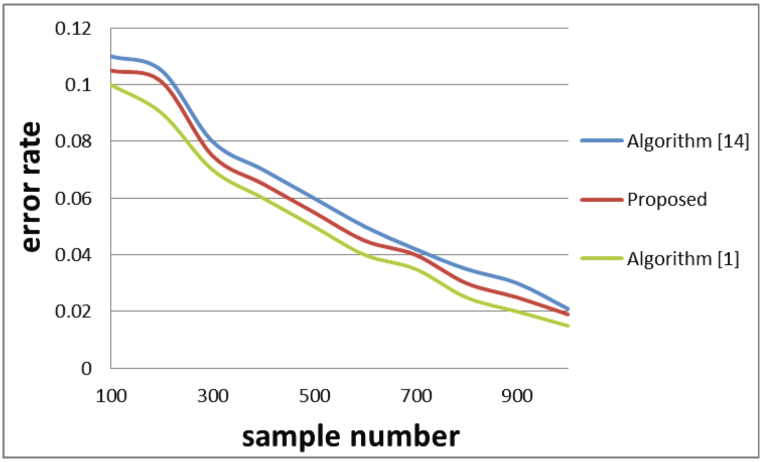

Fig. 3. Incremental performance on the MITEx database

Fig. 3 shows the detection performance improves as samples are increasing in order. Note that there are 563 features adopted in experiments. Roughly speaking, the detection performance will be always enhanced when more samples are incrementally available. Our proposed incremental algorithm performs well nearly the same as the conventional batch algorithm [1], and superior to the training method with complete online learning [14].

It can be known from experiments and data statistics that our incremental algorithm is 15 times faster than traditional algorithm [1] and the storage cost is 130 times less than it, while resources occupancy is fairly the same as other recent incremental learning algorithm [14].

3.2 The Reliability Test of Our Method

In order to verify the reliability of incremental learning algorithm, we also adopt 3500 faces and 7000 nonfaces from internet for offline learning, and employ the MITEx database [19] for testing: collecting 50 face samples from testing database at a time manually, then taking one incremental learning updating for detectors by using these 50 faces as positive samples and other 100 nonfaces as negative samples. The rest images are still reserved for testing and further incremental learning. The total time of incremental learning is 3.

In Fig. 4, the incremental learning on this testing database still can significantly improve the detection performance, where x axis is false alarm rate and y axis is detection rate.

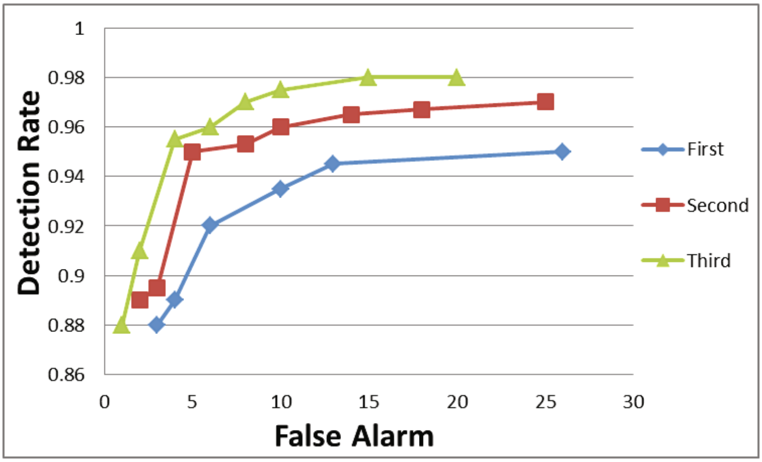

Fig. 4. Test in the MITEx databaset

3.3 The Application Test of Our Auto-incremental Learning in Real System

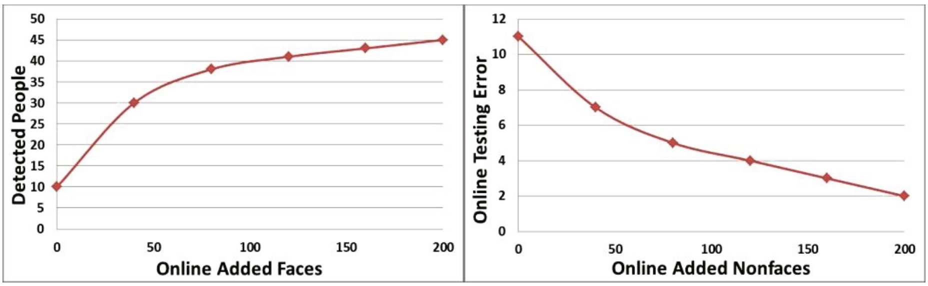

Fig. 5. Improvement of detection performance

In order to test the effectiveness of our auto-incremental learning method applied to wireless video browsing system based on face detection, we further employ the real wireless video captured in projects to perform faces detection and video browsing test. Wherein, face images in CAS-PEAL-R1 standard face database [20] are used for learning training, and then we take the 3-hours wireless video to test our incremental learning method and face detection based video browsing system (total 49 faces). Fig. 5 illustrates that, in real face detection applications with complex poses and views, incremental learning can ensure the promotion of detection performance as incremental learning time increasing by automatically adding learning other face samples, which are those faces extracted by tubes but cannot be detected at that moment. Wherein, those missed people can be detected by unremittingly adding these detected other face samples into incremental learning set.

Fig.7 describes the interface of wirele ss video retrieval and browsing system based on face detection. This system first detects the faces in wireless videos and then stores them. Once playback is required, just clicking the image of the certain person in database, corresponding surveillance video the one appeared can be played.

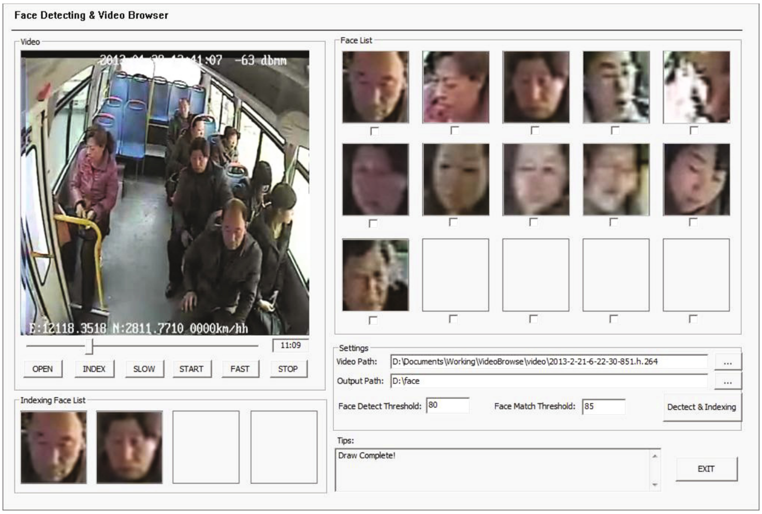

Fig. 6. The interface of wireless video retrieval and browsing system based on face detection with bus scene

4 Conclusions

Oriented to wireless video retrieval and browsing application, this paper has proposed an incremental learned face detection method based on auto-captured samples. Experiments demonstrate that our proposed incremental learning algorithm has better face detection performance and can be adopted in wireless video browsing system. In the future work, we prepare to introduce face recognition into our system to take place of manual face retrieval work, meanwhile, add corresponding feedback module to further improve the reliability of face detection and recognition rate in practical applications.

References

1. Viola, P., Jones, M.: Robust real-time face detection. Int. J. Comput. Vis. 57(2), 137–154 (2004)
2. Li, S.Z., Zhu, L., Zhang, Z., Blake, A., Zhang, H., Shum, H.-Y.: Statistical learning of multi-view face detection. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002, Part IV. LNCS, vol. 2353, pp. 67–81. Springer, Heidelberg (2002)
3. Jin, H., Liu, Q., Lu, H., Tong, X.: Face detection using improved lbp under bayesian framework. In: Third Intl. Conf. on Image and Grahics, ICIG (2004)
4. Wang, X., Han, T.X., Yan, S.: An HOG-LBP human detector with partial occlusion handling. In: Proc. of ICCV (2009)
5. Wu, B., Nevatia, R.: Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors. In: Proc. of ICCV (2005)
6. Michelassi, P., Rocha, A.: Face Liveness Detection under Bad Illumination Conditions. In: IEEE Intl. Conference on Image Processing (2011)
7. Wu, B., Ai, H., Huang, C., Lao, S.: Fast rotation invariant multi-view face detection based on real adaboost. In: Proc. of IEEE Automatic Face and Gesture Recognition (2004)
8. Froba, B., Ernst, A.: Fast frontal-view face detection using a multi-path decision tree. In: Proc. of Audio- and Video based Biometric Person Authentication (2003)
9. Zhuang, S., Lai, S.: Face detection directly from h.264 compressed video with convolutional neural network. In: IEEE Intl. Conference on Image Processing (2009)
10. Porikli, F.: Integral histogram: A fastway to extract histograms in Cartesian spaces. In: Proc. of CVPR (2005)
11. Yan, S., Shan, S., Chen, X., Gao, W.: Locally assembled binary (LAB) feature with feature-centric cascade for fast and accurate face detection. In: Proc. of CVPR (2008)
12. Schneiderman, H.: Feature-centric evaluation for efficient cascaded object detection. In: Proc. of CVPR (2004)
13. Huang, C., Ai, H., Yamashita, T.I., Lao, S., Kawade, M.: Incremental learning of boosted face detector. In: Proceedings of the IEEE Conference on Computer Vision (2007)
14. Pang, Y., Deng, J., Yuan, Y.: Incremental threshold learning for classifier selection. Neurocomputing 89, 89–95 (2012)
15. Liu, R., Gao, X., Chu, R., Zhu, X., Li, S.Z.: Tracking and recognition of multiple faces at distances. In: Lee, S.-W., Li, S.Z. (eds.) ICB 2007. LNCS, vol. 4642, pp. 513–522. Springer, Heidelberg (2007)
16. Friedman, J., Hastie, T., Tibshirani, R.: Additive logistic regression: a statistical view of boosting. Technical report, Dept. of Statistics, Stanford University (1998)
17. Wang, Z., Hu, R., Wang, S., Jiang, J.: Face Hallucination Via Weighted Adaptive Sparse Regularization. IEEE Trans. Circuits Syst. Video Techn. 24(5), 802–813 (2014)
18. Duda, R., Hart, P.: Pattern Classification and Scene Analysis. Wiley, New York (1973)
19. http://www.ecsdn.com/search.php?keyword=MITEx-face-database
20. Zhou, D., Zhang, X., Liu, B., Gao, W.: Introduction of the JDL Large-scale Facial Database. In: The 4th Chinese Conference on Biometrics Recognition (Sinobiometrics 2003), pp. 118–121 (2003)
21. Schafer, R., Schwarz, H., Marpe, D., Schierl, T., Wiegand, T.: MCTF and scalability extension of H.264/AVC and its application to video transmission, storage and surveillance. In: Proc. VCIP, pp. 596 011-1–596 011-12 (2005)