Eyeblink-based Anti-Spoofing in Face Recognition from a Generic Webcamera

Gang Pan*, Lin Sun, Zhaohui Wu
Department of Computer Science
Zhejiang University, China
{gpan, sunlin, wzh}@zju.edu.cn

Shihong Lao
Sensing&Control Technology Lab.
OMRON Corporation, Japan
lao@ari.ncl.omron.co.jp

Abstract

We present a real-time liveness detection approach against photograph spoofing in face recognition, by recognizing spontaneous eyeblinks, which is a non-intrusive manner. The approach requires no extra hardware except for a generic webcamera. Eyeblink sequences often have a complex underlying structure. We formulate blink detection as inference in an undirected conditional graphical framework, and are able to learn a compact and efficient observation and transition potentials from data. For purpose of quick and accurate recognition of the blink behavior, eye closity*, an easily-computed discriminative measure derived from the adaptive boosting algorithm, is developed, and then smoothly embedded into the conditional model. An extensive set of experiments are presented to show effectiveness of our approach and how it outperforms the cascaded Adaboost and HMM in task of eyeblink detection.*

1. Introduction

Biometrics is an emerging technology that recognizes human identities based upon one or more intrinsic physiological or behavioral characteristics, e.g. faces, fingerprints, irises, voice [1]. However, spoofing attack (or copy attack) is still a fatal threat for biometric authentication systems [2]. Liveness detection, which aims at recognition of human physiological activities as the liveness indicator to prevent spoofing attack, is becoming very active in fields of fingerprint recognition and iris recognition [2, 3, 4, 5].

In the face recognition community, numerous recognition approaches have been presented, but the efforts on anti-spoofing are still very limited [6]. The most common faking way is to use a facial photograph of the valid user to spoof the face recognition system, since usually one's facial image is very easily available for the public, for example, downloaded from the web, captured unknowingly by the camera. Photo attack is one of the cheapest and easiest spoofing approaches. The imposter can rotate, shift, and bend the valid user's photo before the camera like a live person to fool the authentication system. It is a challenging task to detect whether the input face image is from a live person or from a photograph.

Most of the current face recognition systems are based on intensity images and equipped with a generic camera. An anti-spoofing method without additional device is more preferable. It could be easily integrated into the existing face recognition systems.

The goal of this paper is to develop a real-time liveness testing approach to resist photograph-spoofing in a non-intrusive manner for face recognition, which does not require any additional hardware except for a generic webcamera.

1.1. Analysis of Face Liveness

In general, a human can distinguish a live face or a photograph without much effort, since a human can easily recognize many physiological clues of liveness, for example, facial expression variation, mouth movement, head rotation, eye change. However, sensing these clues is very difficult for a computer, even under an unconstrained environment. We hope to find some easily computational and hardly spoofed clue for the photo-spoofing prevention.

From the static view, the essential difference between the live face and photograph is that a live face is a fully three dimensional object while a photograph could be considered as a two dimensional planar structure. With this natural trait, Choudhary et al [7] employed the structure from motion yielding the depth information of the face to distinguish live person and still photo. The disadvantages of depth information are that , it is hard to estimate depth information when head is still, and the estimate is very sensitive to noise and lighting condition, becoming unreliable.

Compared with the photograph, another prominent characteristic of live face is the occurrence of the non-rigid deformation and appearance change, such as mouth motion and expression variation. Accurate and reliable detection of these changes usually needs either high-quality input data

*Corresponding author.

978-1-4244-1631-8/07/$25.00 ©2007 IEEE

Clues	Data Quality	Additional Hardware	User Collaboration
Facial expression	High	No	Middle
Depth information	High	No	Low
Mouth movement	Middle	No	Middle
Head movement	High	No	Middle
Eye blinking	Low	No	Low
Degradation	High	No	Low
Multi-modal	-	Yes	Middle/High
Facial thermogram	-	Yes	Low
Facial vein map	-	Yes	Middle
Interactive response	-	No	High

Table 1. Comparison of anti-spoofing clues for face recognition

or user collaboration. Kollreider et al [8] applied optical flow to the input video to obtain the information of face motion for liveness judgment, but it is vulnerable to photo motion in depth and photo bending. Some researchers used the multi-modal approaches of face-voice against spoofing [9, 10], exploiting the lip movement during speech. This kind of method needs voice recorder and user collaboration. An interactive approach was tried by Frischholz et al [11], requiring user to act an obvious response of head movement.

Besides, Li et al [12] presented Fourier spectra to classify live faces or the faked images, based on the assumption that the high frequency components of the photo is less than those of live face images. With additional hardware, the vein map of the face by near infrared imaging, face thermogram [13] also could be applied in to liveness detection.

Table 1 summaries these anti-spoofing clues, in terms of data quality, hardware and user collaboration, for comparison.

1.2. Eyeblink for Liveness Detection

Eyeblink is a physiological activity of rapid closing and opening of the eyelid, which is an essential function of the eye that helps to spread tears across and remove irritants from the surface of the cornea and conjunctiva. Although blink speed can vary with elements such as fatigue, emotional stress, behavior category, amount of sleep, eye injury, medication, and disease, researchers report that [14, 15], the spontaneous resting blink rate of a human being is nearly from 15 to 30 eyeblinks per minute. That is, a person blinks approximately once every 2 to 4 seconds, and the average blink lasts about 250 milliseconds. The current generic camera can easily capture the face video with not less than 15 fps (frames per second), i.e. the frame interval is not more than 70 milliseconds. Thus, it is easy for the generic camera to capture two or more frames for each blink when the face looks into the camera. It is feasible to adopt eyeblink as a clue for anti-spoofing. The advantages of eyeblink based approach lie in: 1) it can complete in a non-intrusive manner, generally without user collaboration, 2) no extra hardware is required, 3) the eyeblink behavior is the prominently distinguishing character of a live face from a facial photo, which would be very helpful for liveness detection only from a generic camera.

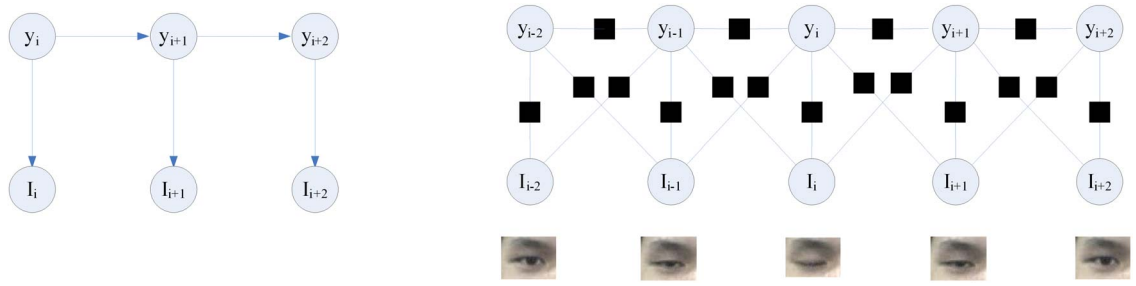
There is little work addressing vision-based detection of eyeblink in the literature. Most of previous efforts need highly controlled conditions and high-quality input data, for example, the system of automatic recognition of human facial action units [16]. Moriyama's eyeblink detection method[17] is based on variation of average intensity in the eye region, sensitive to lighting conditions and noise. Ji et al [18] have attempted to use an active IR camera to detect eyeblinks for prediction of driver fatigue.

This paper poses the photo spoofing problem as detection of eyeblink behaviors. To tackle this problem, we first model blink detection as inference in a Conditional Random Field[19] framework, which enables the long-range dependencies among observations and states. *Eye closity*, a discriminative measure derived from the adaptive boosting algorithm, is introduced and embedded into the contextual model, for computational efficiency and detection accuracy consideration. The extensive experiments are conducted to demonstrate effectiveness of the proposed approach.

2. The Approach

The eyeblink behavior could be represented as a temporal image sequence after being digitally captured by the camera. One typical method to detect blink is to classify each image in the sequence independently as one state of either closed eye or opened eye, for example, using the Viola's cascaded Adaboost approach for face detection [21]. The problem with this method is that it assumes all of the images in the temporal sequence are independent. Actually, the neighboring images of blinking are dependent, since the blink is a procedure of eye going from open to close, and back to open. The temporal information is ignored for this method, which may be very helpful for recognition.

This independence assumption can be relaxed by disposing the state variables in a linear chain. For instance, an HMM (the hidden Markov model) [22] models a sequence of observations by assuming that there is an underlying sequence of states drawn from a finite state set. Features of images can be regarded as the observations, and the eye state label is for the underlying states. HMM makes two independence assumptions to model the joint probability tractably. It assumes that each state depends only on its immediate predecessor, and that each observation variable depends only on the current state, depicted in Fig. 1(a). However, on one hand, the generative-model-based approaches should compute a model of $p(x)$, which is not needed for



(a) (b)

Figure 1. Illustration of graphical structures. (a) Hidden Markov Model, (b) graphical model of a linear-chain CRF, where the circles are variable nodes and the black square boxes are factor nodes, in this example the state is conditioned on contexts of 3 neighboring observations, that is, W=1.

classification anyway. On the other hand, for our task of eyeblink recognition, the two independence assumptions are too restrictive, because there exist dependencies among observations and states, which will benefit blink detection, particularly when the current observation is disturbed by noise such as highlight in eye region, variation of glasses' reflection.

We model eyeblink behaviors in an undirected Conditional Random Field framework, incorporated with a discriminative measure of eye states for simplifying the complex of inference and simultaneously improving the performance. One advantage of the proposed method is that it allows us to relax the assumption of conditional independence of the observed data.

2.1. Conditional Modeling of Blinking Behaviors

An eyeblink activity can be represented by an image sequence $\mathbb{S}$ consisting of T images, where $\mathbb{S} = \{I_i, i = 1, \ldots, T\}$. The typical eye states are *opening* and *closing*. In addition, there is an ambiguous state when blinking from open state to close or from close state to open. We define a three-state set for eyes, $\mathcal{Q} = \{\alpha : open, \gamma : close, \beta : ambiguous\}$. Thus, a typical blink activity can be described as a state change pattern of $\alpha \rightarrow \beta \rightarrow \gamma \rightarrow \beta \rightarrow \alpha$.

Suppose that $\mathbb{S}$ is a random variable over observation sequences to be labeled, and Y is a random variable over the corresponding label sequences to be predicted, all of components y_i of Y are assumed to range over a finite label set $\mathcal{Q}$. Let $G = (V, E)$ be a graph and Y is indexed by the vertices of G. Then $(Y, \mathbb{S})$ is called a *conditional random field (CRF)* [19], when conditioned on $\mathbb{S}$, the random variables Y and $\mathbb{S}$ obey the Markov property w.r.t. the graph:

$$p(y_v|\mathbb{S}, y_u, u \neq v) = p(y_v|\mathbb{S}, y_u, u \sim v) \tag{1}$$

where $u \sim v$ means that u and v are neighbors in G.

We yield a linear chain structure, shown in Fig. 1(b). In this graphical model, a parameter of observation window size W is introduced to describe the conditional relationship between the current state and (2W+1) temporal observations around the current one, in other words, it introduces the long-range dependencies in the model. Using the Hammersley and Clifford theorem [20], the joint distribution over the label sequence Y given the observation $\mathbb{S}$ can be written as the following form:

$$p_\theta(Y|\mathbb{S}) = \frac{1}{Z_\theta(\mathbb{S})} exp(\sum_{t=1}^{T} \Psi_\theta(y_t, y_{t-1}, \mathbb{S})) \tag{2}$$

where $Z_\theta(\mathbb{S})$ is a normalized factor summing over all state sequences, an exponentially large number of terms,

$$Z_\theta(\mathbb{S}) = \sum_{Y} exp(\sum_{t=1}^{T} \Psi_\theta(y_t, y_{t-1}, \mathbb{S})). \tag{3}$$

The potential function $\Psi_\theta(y_t, y_{t-1}, \mathbb{S})$ is the sum of CRF features at time t:

$$\Psi_\theta(y_t, y_{t-1}, \mathbb{S}) = \sum_i \lambda_i f_i(y_t, y_{t-1}, \mathbb{S}) + \sum_j \mu_j g_j(y_t, \mathbb{S}), \tag{4}$$

with parameters $\theta = \{\lambda_1, \ldots, \lambda_A; \mu_1, \ldots, \mu_B\}$, to be estimated from training data.

The f_i and g_j are *within-label* and *between-observation-label* feature functions, respectively. λ_i and μ_j are the feature weights associated with f_i and g_j. We define the *within-label* feature functions f_j as:

$$f_i(y_t, y_{t-1}, \mathbb{S}) = \mathbf{1}_{\{y_t = l\}} \mathbf{1}_{\{y_{t-1} = l'\}} \tag{5}$$

where $l, l' \in \mathcal{Q}$, and $\mathbf{1}_{\{x=x'\}}$ denotes an indictor function of x which takes the value 1 when $x = x'$ and 0 otherwise. Given a temporal window size W around the current observation, the *between-observation-label* feature functions g_j are defined as:

$$g_j(y_t, \mathbb{S}) = \mathbf{1}_{\{y_t = l\}} \mathcal{U}(I_{t-w}) \tag{6}$$

where $l \in \mathcal{Q}$, $w \in [-W, W]$, $\mathcal{U}(.)$ is the eye closity, described in the next section. Here, feature functions f_i and g_j are based on conjunctions of simple rules.

Parameter estimation of $\theta = \{\lambda_1, \ldots, \lambda_A; \mu_1, \ldots, \mu_B\}$ is typically performed by the penalized maximum likelihood. Given a labeled training set $\{Y^{(i)}, \mathbb{S}^{(i)}\}_{i=1,\ldots,N}$, the conditional log likelihood is appropriate:

$$\begin{aligned} L_\theta &= \sum_{i=1}^{N} log(p_\theta(Y^{(i)}|\mathbb{S}^{(i)})) \qquad (7) \\ &= \sum_{i=1}^{N}(\sum_{t=1}^{T} \Psi_\theta(y_t^{(i)}, y_{t-1}^{(i)}, \mathbb{S}^{(i)}) - logZ_\theta(\mathbb{S}^{(i)})) \end{aligned}$$

In order to avoid over-fitting of a large number of parameters, the regularization technique is used, which is a penalty on weight vectors whose norm is too large. For the function L_θ, every local optimum is also a global optimum because the function is concave. Regularization will ensure that L_θ is strictly concave. Finally, the optimization is solved by a limited-memory version of BFGS [23], a kind of quasi-Newton methods. The normalization factor $Z_\theta(\mathbb{S})$ can be computed by the idea of forward-backward.

The inference tasks, for instance, to label an unknown instance $Y^* = \arg\max_Y p(Y|\mathbb{S})$, can be performed efficiently and exactly by variants of the standard dynamic programming methods for HMM [22].

2.2. Eye Closity

From the theoretical view, the original image data could be directly incorporated into the conditional model framework described above. However, it would dramatically increase the complexity and make the problem hard to solve. We hope to take advantage of the features extracted from the image for defining the intermediate observation. For example, silhouette features are commonly used in human motion recognition [24, 25]. Our goal is to develop a real-time approach, thus, we try to use as little feature as possible to reduce the computational cost, meanwhile the features should convey as much discriminative information for eye states as possible to improve the prediction accuracy.

Motivated by the idea of the adaptive boosting algorithm [27], we define a real-value discriminative feature for the eye image, called *eye closity*, $\mathcal{U}(I)$, measuring the degree of eye's closeness, which is constructed by a linear ensemble of a series of weak binary classifiers and computed by an iterative procedure.

$$\mathcal{U}_M(I) = \sum_{i=1}^{M}(\log\frac{1}{\beta_i})h_i(I) - \frac{1}{2}\sum_{i=1}^{M}\log\frac{1}{\beta_i} \qquad (8)$$

where,

$$\beta_i = \epsilon_i/(1-\epsilon_i) \qquad (9)$$

and, $\{h_i(I) : R^{Dim(I)} \rightarrow \{0,1\}, i = 1, \ldots, M\}$ is a set of binary weak classifiers. Each classifier h_i is for classifying the input I as the open eye: $\{0\}$, or the close eye: $\{1\}$. Given a set of labeled training data, the efficient selection of h_i and the calculation of ϵ_i can be performed by an iterative procedure similar to adaptive boosting algorithm [27].

The eye closity can be considered as a sense of the ensemble of effective features. From insight into the training procedure of Adaboost algorithm, we know that the positive value of *closity* indicates that the Adaboost classifier will classify the input as the close eye, and the negative value as the open eye. The bigger the value of *closity*, the higher the degree of eye closeness. A blinking activity sequence is shown in Fig. 2, where the value is closity of the corresponding image, computed after training nearly by 1,000 samples of open eyes and 1,000 samples of close eyes. The closity value of zero is exactly the threshold for the Adaboost classifier.

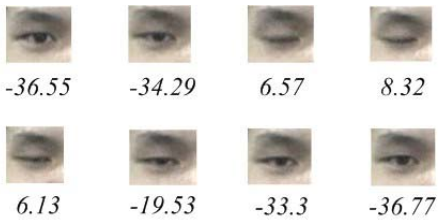


Figure 2. Illustration of the closity for a blinking activity sequence. The closity value of each frame is below the corresponding frame. Bigger the value, higher the degree of closeness. The closity value of zero is exactly the threshold of the Adaboost classifier.

3. Experiments

3.1. Database

To evaluate our approach, we built a publicly available *blinking video database*[26], collected by Logitech Pro5000, a generic webcamera. There is a total of 80 video clips in the blinking video database from 20 individuals, four clips for each individual: one clip for frontal view without glasses, one clip with frontal view and wearing thin rim glasses, one clip for frontal view and black frame glasses, and the last clip with upward view without glasses. Each individual is required to perform blinking spontaneously in normal speed with the above four configurations. One video clip is captured with 30 fps and size of 320×240 for each configuration, lasting about five seconds. The blink number in a video clip varies from 1 to 6 times. There is a total of 255 blinks in the database. All the data are collected indoor without any special lighting condition. Table 2 is statistics of the blinking video database. Some samples are shown in Fig. 3.

Person#	Four clips for each person			Blinks#
	clip#	view	glasses	
20	1	frontal	none	255
	1	frontal	thin rim	
	1	frontal	black frame	
	1	upward	none	

Table 2. Demography of the blinking database. Totally 80 clips and approximately 1 to 6 blinks for each clip.

To test the ability against photo imposters, we also collect a *photo-imposter video database* with 20 persons. A high-quality photo of front view is taken for each person, then five categories of the photo-attacks are simulated before the camera: 1) keep the photo still, 2) move the photo vertically, horizontally, back and front 3) rotate the photo in depth along the vertical axis, 4) rotate the photo in plane, 5) bend the photo inward and outward along the central line. For each attack, one video clip is captured with length of about 10 to 15 seconds and size of 320×240.

3.2. Setting

To compute *eye closity*, we need to train a series of efficient weak classifiers. A total of 1,016 labeled images of close eyes (positive samples) and 1,200 images of open eyes (negative samples) are used in the training stage. We do not differentiate between the left and right eyes. All the samples are scaled to a base resolution of 24×24 pixels. Some positive samples of closed eyes are shown in Fig. 4. Eventually 50 weak classifiers are selected for computing the *eye closity* (Equ.8).

In both the testing stage and the training stage of parameter estimation of blinking conditional model, the center of left and right eyes is automatically localized for each frame by a facial keypoints localization system developed by OMRON's face group. The eye images are extracted and normalized for training, whose size is determined by the distance between the two eyes.

We adopt the *leave-one-out rule* to test the blinking video database. In other words, one clip is selected from 80 clips for test and the remainders act as the training data, then this test procedure is repeated 80 times over the 80 clips, finally get the detection rate. Each pattern of eye state variation $\alpha \rightarrow \beta \rightarrow \gamma \rightarrow \beta \rightarrow \alpha$ is accounted as one blink for this eye.

3.3. Performance Measures

Three types of detection rates are for measuring the approach performance of liveness detection. *One-eye detection rate* is the ratio of number of correctly detected blinks to the total blinks number in test data, where left and right eyes are calculated respectively.

Figure 4. Part of the positive samples for *eye closity*, note that it includes glasses-wearing.

In fact, for each natural blink activity, both left and right eyes will blink. We can determine a live face if we correctly detect the blink of either left or right eye for each blink activity. Thus, *two-eye detection rate* is defined for this case as the ratio of number of correctly detected blink activities to the total blink activities in test data, where the simultaneous blinks of two eyes are accounted for one blink activity.

The third measure is *clip detection rate*, in which case, the clip is considered as live face if any blink of single eye in the clip is detected. In this paper, the length of each blinking clip is five seconds.

3.4. Benefits of Conditioned on Observations

To investigate the benefits of the conditioned on the context of the current observation, an experiment with various windows size setting of $W = \{0, 1, 2, 3, 4\}$ (in Equ.6) is carried out. The results are shown in Fig. 5, from which we can find that the one-eye detection rate significantly increases when the windows size goes from zero to three, demonstrating there exists a strong dependency between the current state and the neighboring observations. Either one-eye detection rate or two-eye detection rate of performance is very close for $W = 3$ and $W = 4$, which shows the dependency becomes weak between the current state and the observations far from its corresponding observation. The window size of $W = 3$ means the contextual observations of 7 frames used for the conditional modeling. A blink activity average 7-8 frames (lasting nearly 250 ms), it can explain that the observations out range of a blink activity have little contribution to the blink detection.

Figure 6 shows three frames' results with W=3. In each frame, there are two bar graphs on the bottom depicting temporal variation of eye closity for both eyes respectively, where the closity of horizontal axis is equal to zero. The red bars indicate the temporal positions that have been labeled as blinking by our method. The temporal variation of closity in Fig. 6(a) is a typical blinking. The closity values of both

Figure 3. Samples from the blinking database. The first row is for no glasses, the second row is with thin rim glasses, the third row for wearing black frame glasses, and the fourth row with upward view. The shown images are sampled every two frames.

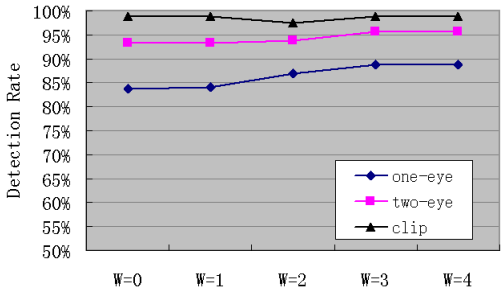


Figure 5. Results of various window size: $W = \{0, 1, 2, 3, 4\}$.

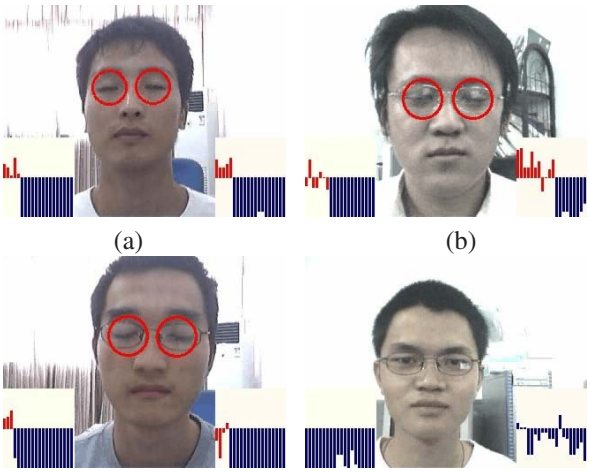


Figure 6. Illustration of temporal variation of closity and blinking detection results. A bar graph shows the temporal variation of closity for each eye. In the bar graph, the vertical axis means eye closity, and the horizontal axis is for time steps. The closity of horizontal axis is equal to zero. The current time step is always located at the leftest of the bar graph. The time steps *in red* indicate these frames have been predicted as a part of a blink activity. The eye is circled in red if its blink is detected by our approach.

eyes are greater than zero during blinking. The left eye in Fig. 6(b) and the right eye of Fig. 6(c) are two samples in which some closity values during blinking are below zero, where Adaboost will fail, while our approach still detects the blinking activities correctly. The right eye in Fig. 6(d) shows another example, where it will be classified as closed eye since the closity values of several neighboring frames are above zero, but our approach "knows" it is open.

The computational cost of online test is very low, averagely 25 ms for one frame of 320-by-240 on P4 2.0GHz, 1GB RAM. Combining with the facial localization system, the whole system could achieve an online processing speed of nearly 20fps, which is reasonable for practical applications.

3.5. Comparison with cascaded Adaboost, HMM

The comparison experiments with cascaded Adaboost and HMM are also conducted.

The labeled training samples for the cascaded Adaboost

Data	cas-Adaboost	HMM	W=0	W=1	W=2	W=3	W=4
One-eye detection rate							
Frontal w/o glasses	96.5%	69.6%	93.8%	93.8%	93.8%	93.8%	94.6%
Frontal w/ thin rim glasses	60.0%	43.9%	83.3%	84.1%	85.6%	85.6%	85.6%
Frontal w/ black frame glasses	46.9%	42.5%	80.6%	79.9%	82.1%	84.3%	84.3%
Upward w/o glasses	52.5%	45.5%	78.8%	79.6%	82.6%	84.9%	84.1%
Average	64.0%	49.6%	83.7%	84.1%	86.9%	**88.8%**	88.8%
Two-eye detection rate							
Frontal w/o glasses	98.2%	80.4%	98.2%	98.2%	98.2%	98.2%	98.2%
Frontal w/ thin rim glasses	80.0%	60.6%	93.9%	93.9%	93.9%	93.9%	93.9%
Frontal w/ black frame glasses	71.9%	55.2%	94.0%	92.5%	89.6%	91.0%	91.4%
Upward w/o glasses	62.3%	59.1%	87.9%	89.4%	92.4%	95.5%	95.5%
Average	78.1%	63.4%	93.3%	93.3%	93.7%	**95.7%**	95.7%

Table 3. Comparison with the cascaded Adaboost and HMM. (false alarm rate $< 0.1\%$)

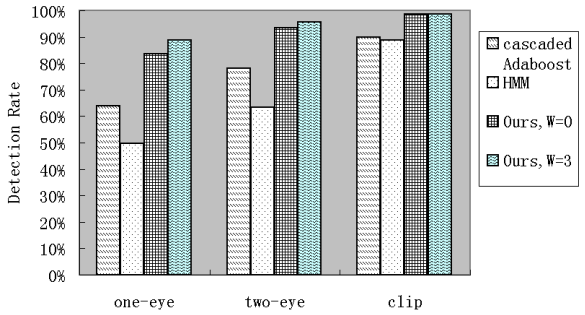


Figure 7. Comparison with cascaded Adaboost and HMM using three performance measures.

are similar to the training data for the *eye closity* computation, include 1,016 close eye samples with size of 24×24 and 1,200 background samples with the open eye (larger than 24×24). Finally, an optimal classifier consisting of eight stages and 73 features is obtained. For HMM, the eye closity of each frame is used as the observation data, same as our approach. The false alarm rates of all the three methods are controlled below 0.1% on the test data.

Figure 7 shows the performance of cascaded Adaboost, HMM and our approach using three measures, one-eye detection rate, two-eye detection rate and clip detection rate. From the figure, it is obvious that our method (with W=3) always significantly outperforms cascaded Adaboost and HMM when different performance measures are used. Note that our approach exploits only 50 features while the cascaded Adaboost uses 73 features.

The detailed detection rates of the three methods are shown in Tab. 3, where the results of four conditions are listed respectively. Although the glasses-wearing and upward view have distinct effect on performance of all the three approaches, our approach still achieves good performance of the average one-eye rate of 88.8% and the average two-eye rate of 95.7% (W=3).

3.6. Photo Imposter Tests

The three methods trained above, cascaded Adaboost, HMM and our method, are also tested for their capability against photo spoofing using the photo-imposter video database. A total of five photo attacks are simulated in the database. Table 4 depicts the results. The number in the table shows how many clips failed during the attack test. It can be seen that the three methods have very similar performance, only 1-2 clips failed out of 100 clips.

4. Conclusions

This paper investigates eyeblinks as liveness detection against photo spoofing in face recognition. The advantages of eyeblink-based method are non-intrusion, no extra hardware requirement, and prominence of activity. To recognize the eyeblink behavior, we model the dependencies among the observations and states in an undirected conditional graphical framework, embedded a new-defined discriminative measure of eye state in order to hasten inference as well as convey the most effective discriminative information. We demonstrated that the proposed approach achieves high performance by just using one generic webcamera under uncontrolled indoor lighting conditions, even was worn glasses-wearing allowed. The comparison experiments showed our approach outperforms cascaded Adaboost and HMM. The proposed eyeblink detection approach, in nature, can be applied to a wide range of applications such as fatigue monitoring, psychological experiments, medical testing, and interactive gaming.

Acknowledgements

This work was in part supported by Program for New Century Excellent Talents in University (NCET-04-0545),

Category of attacks	cas-Adaboost	HMM	W=0	W=1	W=2	W=3	W=4
Keep photo still	0	0	0	0	1	0	0
Move vert., hor., back and front	0	0	0	0	1	0	0
Rotate in depth	1	1	0	0	0	0	0
Rotate in plane	0	0	1	1	0	0	0
Bend inward and outward	0	1	0	0	0	1	0
Total	1	2	1	1	2	1	0

Table 4. Comparison of photo attack test using *photo-imposter video database*, which includes 20 subjects, five categories of photo attacks for each, thus totally 100 video clips. The number shown in the table is the failed clip number.

NSFC grants (60503019, 60525202, 60533040) and a grant from OMRON corporation. The authors would thank Mr. Yoshihisa Ijiri and Mr. Motoo Yamamoto for valuable discussions.

References

[1] A.Ross, K.Nandakumar, A.K.Jain, *Handbook of Multibiometrics*. Springer Verlag, 2006.

[2] S.Schuckers, Spoofing and Anti-Spoofing Measures. *Information Security Technical Report*, 7(4):56-62, Elsevier, 2002.

[3] J.Bigun, H.Fronthaler, K.Kollreider, Assuring liveness in biometric identity authentication by real-time face tracking, *IEEE Conf. on Computational Intelligence for Homeland Security and Personal Safety (CIHSPS'04)*, pp.104-111, 21-22 July 2004.

[4] S.Parthasaradhi, R.Derakhshani, L.Hornak, S.Schuckers, Time-series detection of perspiration as a liveness test in fingerprint devices. *IEEE Trans. Systems, Man and Cybernetics, Part C*, 35(3):335-343, Aug. 2005.

[5] A.Antonelli, R.Cappelli, D.Maio, D.Maltoni, Fake finger detection by skin distortion analysis. *IEEE Trans. Information Forensics and Security*, 1(3):360-373, 2006.

[6] W. Zhao, R. Chellappa, J. Phillips, and A. Rosenfeld, Face Recognition: A Literature Survey. A*CM Computing Surveys*, pp.399-458, 2003.

[7] T.Choudhury, B.Clarkson, T.Jebara, A.Pentland, Multimodal person recognition using unconstrained audio and video, *AVBPA'99*, pp.176-181, Washington DC, 1999.

[8] K.Kollreider, H.Fronthaler, J.Bigun, Evaluating liveness by face images and the structure tensor, *Fourth IEEE Workshop on Automatic Identification Advanced Technologies*, pp.75-80, 17-18 Oct. 2005.

[9] Robert W. Frischholz, Ulrich Dieckmann, BioID: A Multimodal Biometric Identification System, *IEEE Computer*, vol. 33, no. 2, pp.64-68, February 2000.

[10] Girija Chetty, Michael Wagner, Multi-level Liveness Verification for Face-Voice Biometric Authentication, *Biometrics Symposium 2006*, Baltimore, Maryland, Sep. 19-21, 2006.

[11] R.W.Frischholz, A.Werner, Avoiding Replay-Attacks in a Face Recognition System using Head-Pose Estimation, *IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG'03)*, pp.234- 235, 2003.

[12] J.Li, Y.Wang, T.Tan, A.Jain, Live Face Detection Based on the Analysis of Fourier Spectra, Biometric Technology for Human Identification, *Proc. SPIE*, vol. 5404, pp. 296-303, 2004.

[13] D.A. Socolinsky, A. Selinger, J. D. Neuheisel, Face Recognition with Visible and Thermal Infrared Imagery, *CVIU*, vol.91, no. 1-2, pp. 72-114, 2003

[14] Craig Karson, Spontaneous eye-blink rates and dopaminergic systems. *Brain*, vol.106, pp.643-653, 1983.

[15] Kazuo Tsubota, Tear Dynamics and Dry Eye. *Progress in Retinal and Eye Research*, vol.17, no.4, pp565-596, 1998.

[16] Ying-li Tian, Kakeo Kanade, J.F..Cohn, Recognizing Action Units for Facial Expression Analysis. *IEEE PAMI*, vol.23, no.2, pp.97-115, 2001.

[17] T.Moriyama, T.Kanade, J.F.Cohn, J.Xiao, Z.Ambadar, J.Gao, H.Imamura, Automatic Recognition of Eye Blinking in Spontaneously Occurring Behavior. *ICPR'02*, 2002.

[18] Qiang Ji, Zhiwei Zhu, Peilin Lan, Real Time Nonintrusive Monitoring and Prediction of Driver Fatigue, *IEEE Trans. Vehicular Technology*, vol.53, no.4, pp.1052-1068, 2004.

[19] J.Lafferty, A.McCallum, and F.Pereira, Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. *ICML'01*, pp.282-289, 2001.

[20] S.Z.Li, *Markov Random Field Modeling in Image Analysis*. Springer-Verlag, 2001.

[21] P.Viola, M.J.Jones, Rapid Object Detection using a Boosted Cascade of Simple Features. *CVPR'01*, pp.511-518, 2001.

[22] L.R.Rabiner, A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257-286, 1989.

[23] F.Sha, F.Pereira, Shallow Parsing with Conditional Random Fields. *Proc. Human Language Technology, NAACL*, pp. 213-220, 2003.

[24] S.Cristian, A.Kanaujia, Z.Li, D.Metaxas, Conditional Models for Contextual Human Motion Recognition, *ICCV'05*, pp.1808-1815, 2005.

[25] D.Gavrila, The Visual Analysis of Human Movement: A Survey, *CVIU*, vol.73, no.1, pp.82-98, 1999.

[26] ZJU Eyeblink Database, http://www.cs.zju.edu.cn/~gpan or http://www.stat.ucla.edu/~gpan.

[27] Y.Freund, R.Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, *Second European Conference on Computational Learning Theory*, pp.23-37, 1995.