We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS BOOK CITATION INDEX INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

9

Liveness Detection for Face Recognition

Gang Pan, Zhaohui Wu and Lin Sun
Department of Computer Science, Zhejiang University
China

1. Introduction

Biometrics is an emerging technology that enables uniquely recognizing humans based upon one or more intrinsic physiological or behavioral characteristics, such as faces, fingerprints, irises, voices (Ross et al., 2006). However, spoofing attack (or copy attack) is still a fatal threat for biometric authentication systems (Schukers, 2002). Liveness detection, which aims at recognition of human physiological activities as the liveness indicator to prevent spoofing attack, is becoming a very active topic in field of fingerprint recognition and iris recognition (Schuckers, 2002; Bigun et al., 2004; Parthasaradhi et al., 2005; Antonelli et al., 2006).

In face recognition community, although numerous recognition approaches have been presented, the effort on anti-spoofing is still very limited (Zhao et al., 2003). The most common faking way is to use a facial photograph of a valid user to spoof face recognition systems. Nowadays, video of a valid user can also be easily captured by needle camera for spoofing. Therefore anti-spoof problem should be well solved before face recognition could be widely applied in our life.

Most of the current face recognition works with excellent performance, are based on intensity images and equipped with a generic camera. Thus, an anti-spoofing method without additional device will be preferable, since it could be easily integrated into the existing face recognition systems.

In Section 2, we give a brief review of spoofing ways in face recognition and some related work. The potential clues will be also presented and commented. In Section 3, a real-time liveness detection approach is presented against photograph spoofing in a non-intrusive manner for face recognition, which does not require any additional hardware except for a generic webcamera. In Section 4, databases are introduced for eyeblink-based anti-spoofing. Section 5 presents an extensive set of experiments to show effectiveness of our approach. Discussions are in Section 6.

2. Spoofing in face recognition

Generally speaking, there are three ways to spoof face recognition:

a. Photograph of a valid user
b. Video of a valid user
c. 3D model of a valid user

Source: Recent Advances in Face Recognition, Book edited by: Kresimir Delac, Mislav Grgic and Marian Stewart Bartlett,
ISBN 978-953-7619-34-3, pp. 236, December 2008, I-Tech, Vienna, Austria

Photo attack is the cheapest and easiest spoofing approach, since one's facial image is usually very easily available for the public, for example, downloaded from the web, captured unknowingly by a camera. The imposter can rotate, shift and bend the photo before the camera like a live person to fool the authentication system. It is still a challenging task to detect whether an input face image is from a live person or from a photograph.

Video spoofing is another big threat to face recognition systems, because it is very similar to live face and can be shot in front of legal user's face by a needle camera. It has many physiological clues that photo does not have, such as head movement, facial expression, blinking et al.

3D model has 3D information of face, however, it is rigid and lack of physiological information. It is also not very easy to be realistic with live person who the 3D model imitates. So photo and video are most common spoofing ways to attack face recognition system.

In general, human is able to distinguish a live face and a photograph without any effort, since human can very easily recognize many physiological clues of liveness, for example, facial expression variation, mouth movement, head rotation, eye change. However, the tasks of computing these clues are often complicated for computer, even impossible for some clues under the unconstrained environment.

From the static view, an essential difference between a live face and a 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 employed the structure from motion yielding the depth information of the face to detect live person or still photo (Choudhary et al., 1999). The disadvantages of depth information are that, firstly it is hard to estimate depth information when head is still. Secondly, the estimate is very sensitive to noise and lighting condition, becoming unreliable.

Compared with photographs, another prominent characteristic of live faces is the occurrence of the non-rigid deformation and appearance change, such as mouth motion, expression variation. The accurate and reliable detection of these changes usually needs either the input data of high-quality or user collaboration. Kollreider et al applies the optical flow to the input video to obtain the information of face motion for liveness judgement (Kollreider et al., 2005), but it is vulnerable to photo motion in depth and photo bending. Some researchers use the multi-modal approaches of face-voice against spoofing (Frischholz, & Dieckmann, 2000; Chetty & Wagner, 2006), exploiting the lip movement during speaking. This kind of method needs voice recorder and user collaboration. An interactive approach is tried by Frischholz et al, requiring user to act an obvious response of head movement (Frischholz & Dieckmann, 2000).

Besides, Li et al presented Fourier spectra to classify live faces or faked images, based on the assumption that the high frequency components of the photo is less than those of live face images (Li et al., 2004). With thermal infrared imaging camera, face thermogram also could be applied in to liveness detection (Socolinsky et al., 2003).

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

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

3. Blinking-based liveness detection

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 will be preferable, since it could be easily integrated into the existing face recognition approach and system.
In this section, a blinking-based liveness detection approach is introduced for prevention of photograph-spoofing. It 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.

3.1 Why blinking used?

We hope to find some easily computational, also hardly disguising clue for the photo-spoofing protection. Eyeblink is a physiological activity of rapid closing and opening of eyelids, which is an essential function of eyes that helps 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 (Karson,1983; Tsubota, 1998), 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 a blink lasts averagely 250 milliseconds. Currently a generic camera can easily capture a 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 a generic camera to capture two or more frames for each blink when a face looks into the camera. It is feasible to adopt eyeblink as the 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. Eyeblink behaviour is the prominently distinguishing character of a live face from a facial photo, which would be much helpful for liveness detection only from the generic camera.

There is little work addressing vision-based detection of eyeblink in the literature. Most of the previous efforts need highly controlled conditions and high-quality input data, for instance, the automatic recognition system of human facial action units (Tian et al., 2001). Moriyama's blinking detection method (Moriyama et al., 2002) is based on variation of average intensity in the eye region, sensitive to lighting conditions and noise. Ji et al have attempted to use an active IR camera to detect eyeblinks for prediction of driver fatigue (Ji et al., 2004).

3.2 Overview of the approach

An eyeblink behaviour 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 close eye or open eye, for example, using the Viola's cascaded Adaboost approach like face detection (Viola & Jones, 2001). 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 eventually from opening to closure, then to opening. 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) (Rabiner, 1989) models a sequence of observations by assuming that there is an underlying sequence of states drawn from a finite state set. The features of image could be regarded as the observations, and the eye state label is for the underlying states. A 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 classification anyway. On the other hand, for our task of eyeblink recognition, the two independence assumptions are too restrictive, since, in fact, there exist dependencies among observations and states, which will benefit blink detection, in particular 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 sates for simplifying the complex of inference and simultaneously improving the performance. One of advantage of the proposed method is that allows us to relax the assumption of conditional independence of the observed data.

3.3 Conditional modeling of blinking behaviours

An eyeblink activity can be represented by an image sequence **S** consisting of T images, where $S=\{I_i,\ i=1,\ldots,T\}$. The typical eye states in the images are opening and closing, in

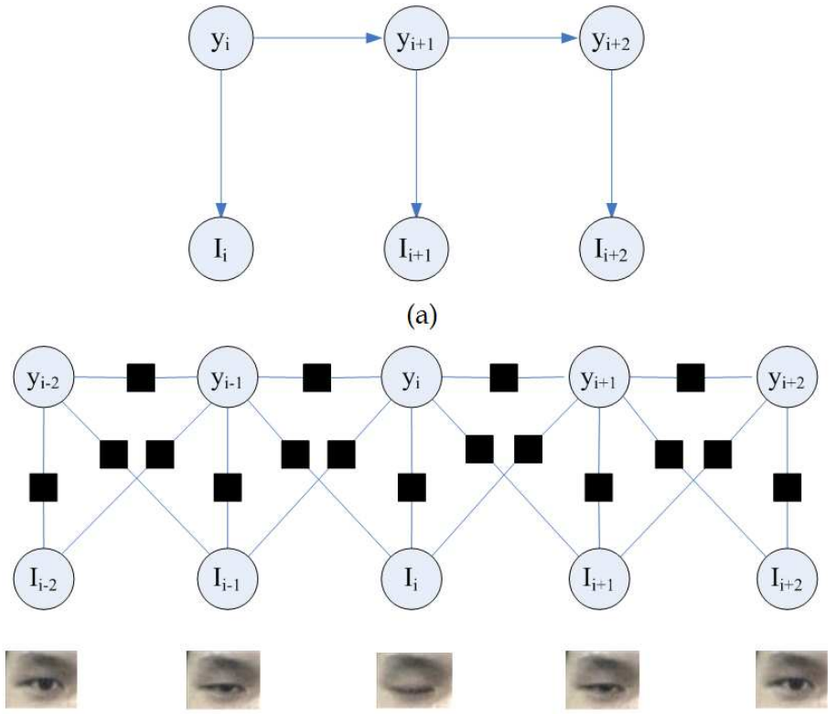


Fig. 1. Illustration of graphical structures. (a) Hidden Markov Model, (b) graphical model of a linear-chain CRF, where the circles are variable nodes and the black boxes are factor nodes, in this example the state depends on contexts of 3 neighbouring observations, that is, w=1.

addition, there is an ambiguous state when the eyeblinks from open state to close or from close state to open. We define a three-state set for eyes,

$$Q=\{a : open, \gamma : close, \beta: ambiguous\}.$$

Thus, a typical blink activity can be described as a state change pattern of $a \rightarrow \beta \rightarrow \gamma \rightarrow \beta \rightarrow a$. Suppose that 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 Q. Let $G=(V, E)$ be a graph and Y is indexed by the vertices of G. Then (Y, S) is called a *conditional random field* (*CRF*) (Lafferty et al., 2001), when conditioned on S, the random variables Y and S obey the Markov property w.r.t. the graph:

$$p(y_v|S, y_u, u \neq v)=p(y_v|S, y_u, u \sim v) \quad (1)$$

Where $u \sim v$ means that u and v are neighbours 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 the other word, it introduces the long-range dependencies in the model. Using the fundamental theorem by Hammersley & Clifford (Li, 2001), the joint distribution over the label sequence Y given the observation S can be written as the following form:

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

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

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

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

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

With parameter $\theta=\{\lambda_1,...,\lambda_A;\mu_1,...,\mu_B\}$, 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. Feature functions f_i and g_j are based on conjunctions of simple rules. The *within-label* feature functions f_i are:

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

Where $l,l'\in Q$, and $1_{x=x'}$ 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:

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

Where $l\in Q$, $w\in[-W,W]$, $U(\cdot)$ is the eye closity, described in the next section. W is for a context window size around the current observation.
Parameter estimation of $\theta=\{\lambda_1,...,\lambda_A;\mu_1,...,\mu_B\}$ is typically performed by penalized maximum likelihood. Given a labeled training set $\{Y^{(i)}, S^{(i)}\}_{i=1,...,N}$, the conditional log likelihood is appropriate:

$$L_\theta=\sum_{i=1}^{N}log(p_\theta(Y^{(i)}|S^{(i)}))=\sum_{i=1}^{N}(\sum_{t=1}^{T}\psi_\theta(y_t^{(i)},y_{t-1}^{(i)},S^{(i)})-log(Z_\theta(S^{(i)}))) \quad (7)$$

In order to avoid over-fitting of a large number of parameters, the regularization technique is used, 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 convex. Regularization will ensure that L_θ is strictly convex. Finally, the optimization is solved by a limited-memory version of BFGS (Sha & Pereira, 2003), of quasi-Newton methods. The normalization factor $Z_\theta(S)$ can be computed by the idea forward-backward.
The inference tasks, for instance, to label an unknown instance $\boldsymbol{Y}^*=\text{argmax}_Y\ p(Y|S)$, can performed efficiently and exactly by variants of the standard dynamic programming methods for HMM.

3.4 Eye closity: definition and computation

From the theoretical view, the original image data could be directly incorporated into the conditional model framework described above. However, obviously, 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 (Cristian et al., 2005; Gavrila, 1999}. 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 (Freund & Schapire, 1995), we define a real-value discriminative feature for the eye image, called *eye closity*, *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.

$$U_M(I) = \sum_{i=1}^{M} (\log \frac{1}{\beta_i}) h_i(I) - \frac{1}{2} \sum_{i=1}^{M} \log \frac{1}{\beta_i} \tag{8}$$

Where,

$$\beta_i = \varepsilon_i / (1 - \varepsilon_i) \tag{9}$$

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 labelled 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 (Freund & Schapire, 1995).

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 Adaboosted classifier will classify the input as the close eye, and the negative value as the open eye. Bigger the value of *closity is*, higher 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 Adaboosted classifier.

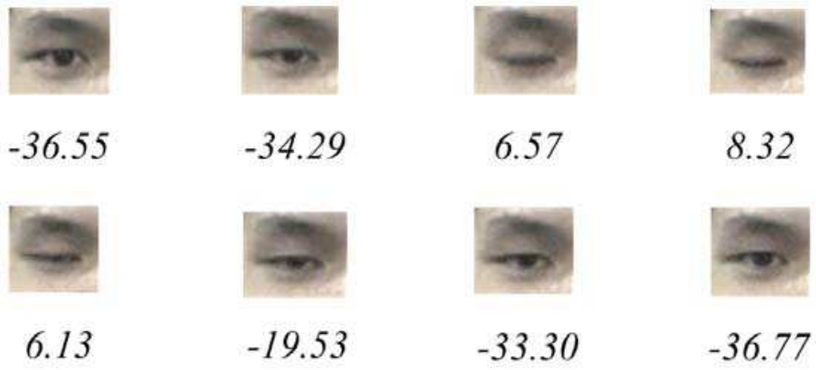


Fig. 2. Illustration of the closity for a blinking activity sequence. The closity value of each frame is below the corresponding frame. Bigger the value is, higher the degree of closeness. The closity value of zero is exactly the threshold of the Adaboost classifier.

4. Databases

To evaluate the proposed approach, we collected and built two databases: ZJU Eyeblink Database and ZJU Photo-Imposter Database.

4.1 ZJU eyeblink database

The ZJU Eyeblink Database is publicly available (http://www.cs.zju.edu.cn/~gpan or http://www.stat.ucla.edu/~gpan). It contains 80 video clips in AVI format of 20 individuals, collected by Logitech Pro 5000. There are 4 clips per subject: a clip for frontal view without glasses, a clip with frontal view and wearing thin rim glasses, a clip for frontal view and black frame glasses, and a clip with upward view without glasses. Each individual is required to perform blinking spontaneously in normal speed with the above four configurations. Each video clip is captured with 30 fps and size of 320x240 for each configuration, lasting about 5 seconds. The blink number in a video clip varies from 1 to 6 times. There are totally 255 blinks in the database. All the data are collected indoor without lighting control. Table 2 is demography 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.

Fig. 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.

4.2 ZJU photo-imposter database

To test the ability against photo imposters, we also collect a *photo-imposter 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 horizontally, vertically, 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 with size of 320 × 240. Five categories of the photo-attacks are shown in Fig. 4.

(1) (2) (3) (4) (5)

Fig. 4. Five categories of photo-attacks: (1) keep the photo still, (2) move the photo horizontally, vertically, 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.

5. Experiments

5.1 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 and negative samples of open eyes are shown in Fig. 5. 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 face key-point 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 $a \rightarrow \beta \rightarrow \gamma \rightarrow \beta \rightarrow a$ is accounted as one blink for this eye.

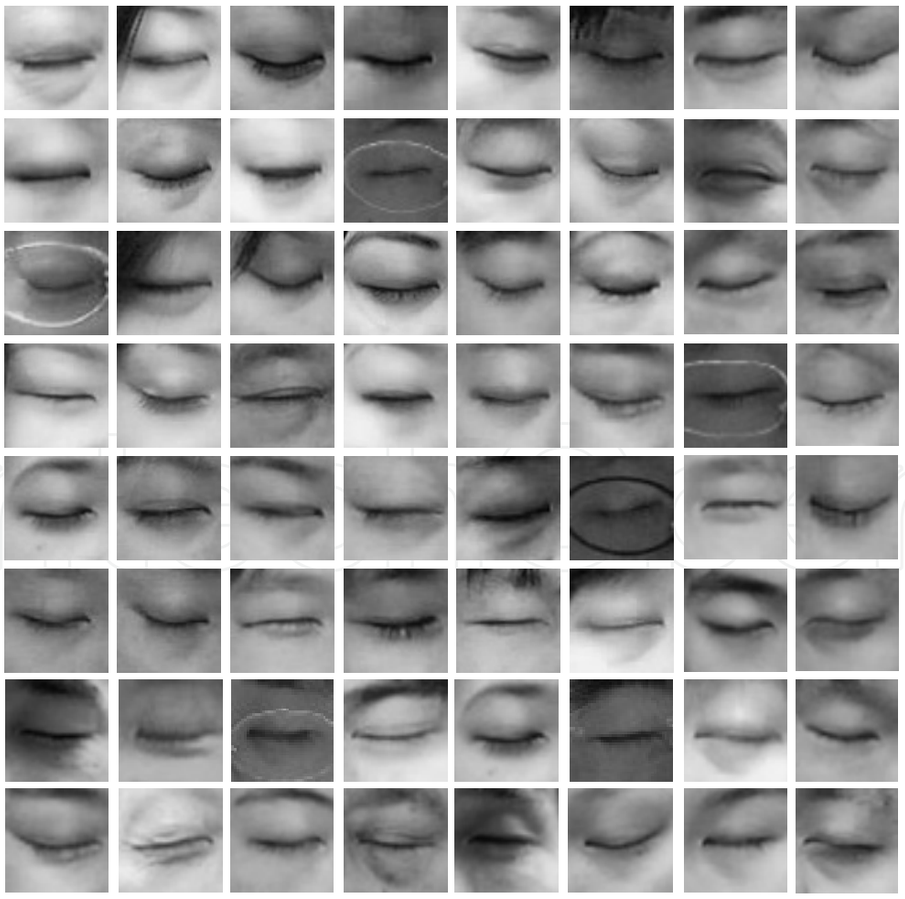
5.2 Performance measures

Three types of detection rates are for measuring the approach performance of liveness detection.

1. *One-eye detection rate*: it 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.
2. *two-eye detection rate*: 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.
3. *clip detection rate*: 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.

5.3 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. 6., 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



(a)

(b)

Fig. 5. Samples for computation of *eye closity*. (a) positive samples, (b) negative samples. Note that it includes glasses-wearing samples.

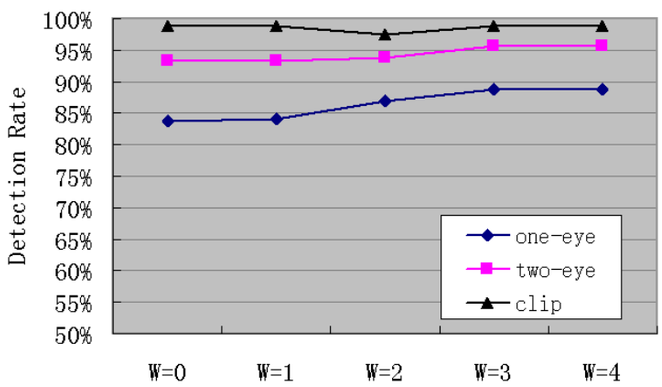


Fig. 6. Results of various window size: W = {0, 1, 2, 3, 4}.

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.

Fig.7 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. 7(a) is a typical blinking. The closity values of both eyes are greater than zero during blinking. The left eye in Fig. 7(a) and the right eye of Fig. 7(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. 7(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.

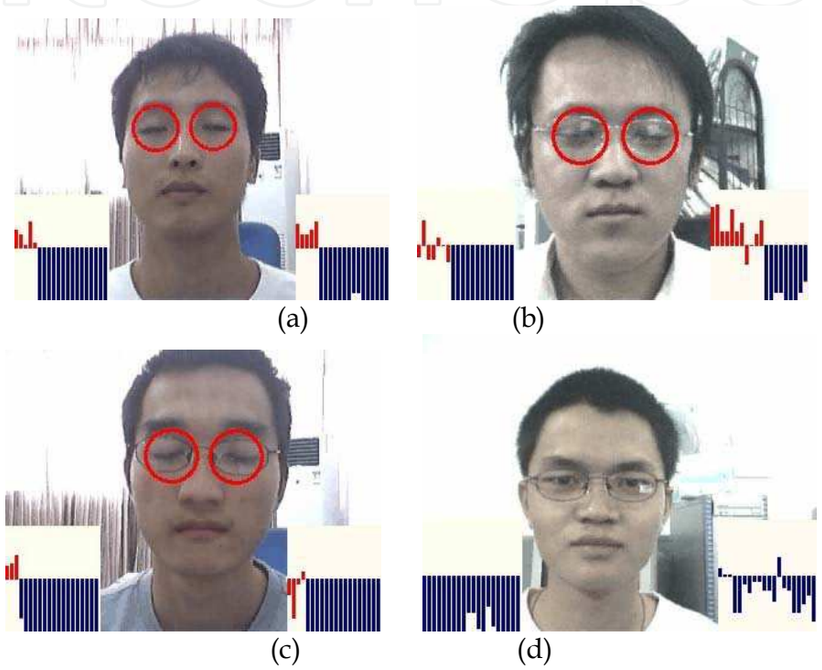


Fig. 7. Illustration of temporal variation of closity and blinking detection results. A bar graph shows the temporal variation of closity for each eye. In the bar graph, the vertical axis means eye closity, and the horizontal axis is for time steps. The closity of horizontal axis is equal to zero. The 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.

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.

5.4 Comparison with cascaded Adaboost, HMM

The comparison experiments with cascaded Adaboost and HMM are also conducted. The labeled training samples for the cascaded Adaboost 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.

Fig. 8. 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.

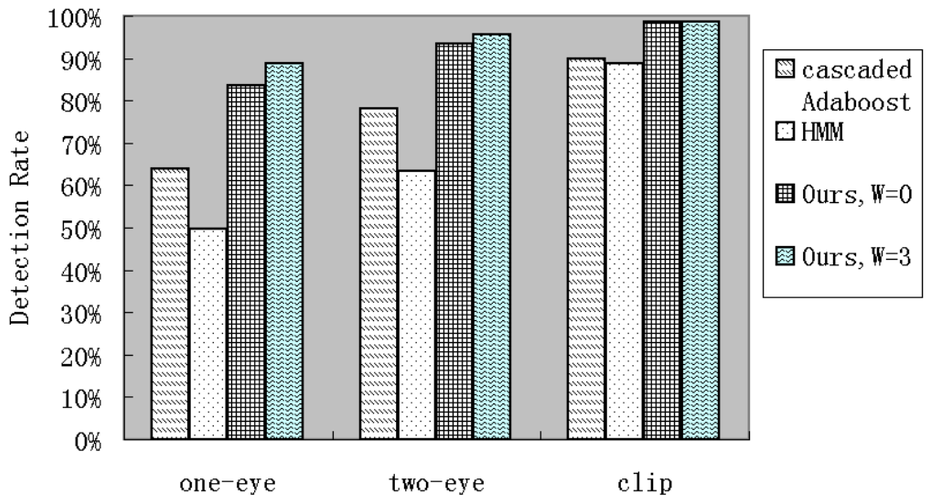


Fig. 8. Comparison with cascaded Adaboost and HMM using three performance measures.

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	8.1%	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%)

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).

5.5 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.

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 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.

6. Discussions

We investigate eyeblinks as a liveness detection clue against photo spoofing in face recognition. The advantages of eyeblink-based method are non-intrusion, no requirement of extra hardware. Undirected conditional graphical framework, which assumes dependencies among the observations and states, is employed to model eyeblink. A new-defined discriminative measure of eye states, called eye closity, can hasten inference as well as convey most effective discriminative information. Experiments demonstrate that the proposed approach achieves high performance by just using one generic webcamera under uncontrolled indoor lighting conditions, even glasses are worn. The comparison experiments show 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.

However, blinking-based liveness detection has some limitations. It would be affected by strong glasses reflection, which may cover eyes partially or totally. Blink clue also does not work for video spoofing. Anti-video spoofing is still a challenge to researchers.

7. Acknowledgements

This work was partly supported by NSFC grants (60503019, 60525202, 60533040), PCSIRT Program (IRT0652), 863 Program (2008AA01Z149), and a grant from OMRON corporation.

8. References

Antonelli, A.; Cappelli, R. & Maio, D. & Maltoni, D. (2006). Fake finger detection by skin distortion analysis. IEEE Trans. Information Forensics and Security, Vol.1, No.3, pp. 360-373, 2006

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

Chetty, G. & Wagner, M. (2006). Multi-level Liveness Verification for Face-Voice Biometric Authentication, Biometric Symposium 2006, Baltimore, Maryland, Sep 2006

Choudhury, T.; Clarkson, B. & Jebara, T. & Pentland, A. (1999). Multimodal person recognition using unconstrained audio and video, International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA'99), pp.176-181, Washington DC, 1999

Cristian, S.; Kanaujia, A. & Li, Z. & Metaxas, D. (2005). Conditional Models for Contextual Human Motion Recognition, IEEE International Conference on Computer Vision (ICCV'05), pp.1808-1815, 2005

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

Frischholz, R.W. & Dieckmann, U. (2000). BioID: A Multimodal Biometric Identification System, IEEE Computer, Vol. 33, No. 2, pp.64-68, February 2000

Frischholz, R.W. & Werner, A. (2003). 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

Gavrila,D. (1999). The Visual Analysis of Human Movement: A Survey, Computer Vision and Image Understanding, Vol.73, No.1, pp.82-98, 1999

Ji, Q.; Zhu, Z. & Lan, P. (2004). Real Time Nonintrusive Monitoring and Prediction of Driver Fatigue, IEEE Trans. Vehicular Technology, Vol.53, No.4, pp.1052-1068, 2004

Karson, C. (1983). Spontaneous eye-blink rates and dopaminergic systems. Brain, Vol.106, pp.643-653, 1983

Kollreider, K.; Fronthaler, H. & Bigun, J. (2005). Evaluating liveness by face images and the structure tensor, Fourth IEEE Workshop on Automatic Identification Advanced Technologies, pp.75-80, Oct. 2005

Lafferty, J.; McCallum, A. & Pereira, F. (2001) Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. International Conference on Machine Learning (ICML'01), pp.282-289, 2001

Li, J.; Wang, Y. & Tan, T. & Jain, A. (2004). Live Face Detection Based on the Analysis of Fourier Spectra, Biometric Technology for Human Identification, Proceedings of SPIE, Vol. 5404, pp. 296-303, 2004

Li, S.Z. (2001) Markov Random Field Modeling in Image Analysis. Springer-Verlag, 2001

Moriyama, T.; Kanade, T. & Cohn,J.F. & Xiao, J. & Ambadar, Z. & Gao, J. & Imamura, H. (2002). Automatic Recognition of Eye Blinking in Spontaneously Occurring Behavior. IEEE International Conference on Pattern Recognition (ICPR'02), 2002

Parthasaradhi, S.; Derakhshani R. & Hornak, L. & Schuckers, S. (2005). Time-series detection of perspiration as a liveness test in fingerprint devices. IEEE Trans. Systems, Man and Cybernetics, Part C, Vol.35, No.3, pp. 335-343, Aug. 2005

Rabiner, L.R. (1989). A tutorial on hidden markov models and selected applications in speech recognition. Proceedings of the IEEE, Vol.77, No.2, pp.257-286, 1989

Ross, A.; Nandakumar, K. & Jain, A.K. (2006). Handbook of Multibiometrics, Springer Verlag.

Schuckers, S. (2002). Spoofing and Anti-Spoofing Measures. Information Security Technical Report, Vol.7, No.4, 56-62, Elsevier

Sha, F. & Pereira, F. (2003). Shallow Parsing with Conditional Random Fields. Proc. Human Language Technology, NAACL, pp. 213-220, 2003

Socolinsky, D.A.; Selinger, A. & Neuheisel, J. D. (2003). Face Recognition with Visible and Thermal Infrared Imagery, Computer Vision and Image Understanding, vol.91, no. 1-2, pp. 72-114, 2003

Tian, Y.; Kanade, K. & Cohn, J.F. (2001). Recognizing Action Units for Facial Expression Analysis. IEEE Trans. Pattern Analysis and Machine Intelligence, Vol.23, No.2, pp.97-115, 2001

Tsubota, K. (1998). Tear Dynamics and Dry Eye. Progress in Retinal and Eye Research, Vol.17, No.4, pp565-596, 1998

Viola, P. & Jones, M.J. (2001). Rapid Object Detection using a Boosted Cascade of Simple Features. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'01), pp.511-518, 2001.

Zhao, W.; Chellappa, R. & Phillips, J. & Rosenfeld, A. (2003). Face Recognition: A Literature Survey. ACM Computing Surveys, pp.399-458, 2003

Recent Advances in Face Recognition

Edited by Kresimir Delac, Mislav Grgic and Marian Stewart Bartlett

ISBN 978-953-7619-34-3
Hard cover, 236 pages
Publisher InTech
Published online 01, June, 2008
Published in print edition June, 2008

The main idea and the driver of further research in the area of face recognition are security applications and human-computer interaction. Face recognition represents an intuitive and non-intrusive method of recognizing people and this is why it became one of three identification methods used in e-passports and a biometric of choice for many other security applications. This goal of this book is to provide the reader with the most up to date research performed in automatic face recognition. The chapters presented use innovative approaches to deal with a wide variety of unsolved issues.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Gang Pan, Zhaohui Wu and Lin Sun (2008). Liveness Detection for Face Recognition, Recent Advances in Face Recognition, Kresimir Delac, Mislav Grgic and Marian Stewart Bartlett (Ed.), ISBN: 978-953-7619-34-3, InTech, Available from: http://www.intechopen.com/books/recent_advances_in_face_recognition/liveness_detection_for_face_recognition

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen