Blinking-Based Live Face Detection Using Conditional Random Fields

Lin Sun[1], Gang Pan[1,⋆], Zhaohui Wu[1], and Shihong Lao[2]

[1] Dept. of Computer Science, Zhejiang University, Hangzhou, P.R. China
Tel.:(86)571-8795-1647
{sunlin,gpan,wzh}@zju.edu.cn
[2] Sensing and Control Technology Laboratory, OMRON Corporation, Japan
lao@ari.ncl.omron.co.jp

Abstract. This paper presents a blinking-based liveness detection method for human face using Conditional Random Fields (CRFs). Our method only needs a web camera for capturing video clips. Blinking clue is a passive action and does not need the user to to any hint, such as speaking, face moving. We model blinking activity by CRFs, which accommodates long-range contextual dependencies among the observation sequence. The experimental results demonstrate that the proposed method is promising, and outperforms the cascaded Adaboost method and HMM method.

1 Introduction

Biometrics are emerging technologies that enable the authentication of an individual based on physiological or behavioral characteristics, which including faces, fingerprints, irises, voices, etc [1]. However, spoofing attack (or copy attack) is a serious threat for biometrics. Liveness detection in biometric systems can prevent spoofing attack, based on recognition of physiological activities as the sign of liveness.

Face recognition is one of the most useful biometrics and has many applications. However, the liveness detection in face recognition system has be little addressed. The vein map of the faces using ultra-violet cameras may be a secure method of identifying a live individual, but it needs additional expensive devices. It is a big challenge to detect live face from a web camera in face recognition system. A little work on live face detection has been done. Robert et al[2] used the multi-modal approaches against spoofing, which need voice recorder and user collaboration. Choudhary et al[3] used 3D depth information of a human head to detect live person. However, it is hard to estimate depth information when head is still. Li et al[4] proposed Fourier spectra to classify live faces and the faked images, but it strongly depends on the data quality. Kollreider et al[5] provided a method to determine liveness by applying optical flow to obtain the information of face motion. This kind of methods is vulnerable to photo motion, such as

[⋆] Corresponding author.

S.-W. Lee and S.Z. Li (Eds.): ICB 2007, LNCS 4642, pp. 252–260, 2007.
© Springer-Verlag Berlin Heidelberg 2007

photo bending. Moriyama et al[6] proposed an eyeblink detection method based on the changes of mean intensity of the upper half and the lower half in eye region, usually the images of high quality are required.

Eye blinking is a spontaneous physiological behavior. The normal resting blink rate of a human being is 20 times per minute, with the average blink time lasting one quarter of a second[21]. Thus, eye blinking could be a very useful clue to identify a live face.

This paper focuses on using blinking clue for face liveness detection. We use CRFs to model blinking activities, for its accommodating long-range dependencies on the observation sequence [7]. We compare CRF model with cascaded AdaBoost [17], which is a discriminative model, and HMM, which is a generative model. The experiment results show that CRF model has better performance than the others.

2 Modelling Blinking Activities Using CRFs

Blinking is an action sequence that consists of two continuous sub-actions, from open to close and from close to open. Blinking activity could be sampled by web camera into an image sequence. The typical eye states in the images are *open*, *half-open* and *close*. Every state should not be considered independently for blinking recognition. It means that contextual dependencies in eye blinking sequence need to be considered when modelling blinking activities. It is hard to predict blinking activity at a particular time point using only the previous state and the current observation alone. Hidden Markov model (HMM)[13], which is a generative model, can't accommodate long-range dependencies on the observation sequence. Conditional random fields(CRFs) are probabilistic models for segmenting and labeling sequence data and mainly used in natural language processing for its accommodating long-range dependencies on the observation sequence [7,8,9].

We employ a linear chain structure of CRFs. It has discrete eye state label data $y_t \in \chi = \{1, 2, \ldots, c\}$, $t = 1 \ldots T$, and observation x_t. Half-open state is hard to define commonly over the different individuals, since the eye size of half-open state depends on the person eye appearance, for example, the open state of a small eye may look like the half-open state of a big eye. As for our blinking model, we use two state labels, C for close state and NC for non-close (include half-open and open), to label eye states. Observation data, which are features extracted from the image, will be discussed in next section. Graphical structure of our CRF-based blinking model is shown in Fig. 1. For notational compactness, we consider $Y = (y_1, y_2, \ldots, y_T)$ to be the label sequence, $X = (x_1, x_2, \ldots, x_T)$ to be the observation sequence.

Let $G = (V, E)$ be a graph and Y is indexed by the vertices of G. Then (Y, X) is called a *conditional random field*, when conditioned on X, the random variables X and Y obey the Markov property w.r.t. the graph:

$$p(y_v|X, Y_w, w \neq v) = p(y_v|X, Y_w, w \sim v), \tag{1}$$

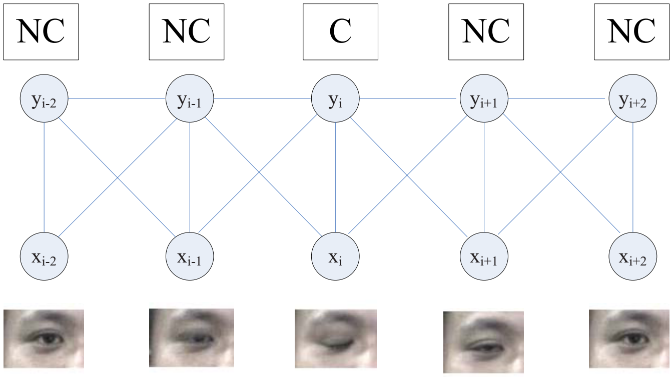

Fig. 1. Graphic structure of CRF-based blinking model. Here we show the model based on contexts of observation of size 3. Labels C and NC are for close state and non-close state respectively.

where $w \sim v$ means that w and v are neighbors in G. Using the Hammersley Clifford theorem [10], the joint distribution over the label sequence Y given observation X has the form:

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

$$Z_\theta(X) = \sum_{Y} exp(\sum_{t=1}^{T} F_\theta(y_t, y_{t-1}, X)), \tag{3}$$

where $Z_\theta(X)$ is a normalized factor summing over all label sequences. Potential functions $F_\theta(y_t, y_{t-1}, X)$ is the sum of CRF features at time t:

$$F_\theta(y_t, y_{t-1}, X) = \sum_{i} \lambda_i f_i(y_t, y_{t-1}, X) + \sum_{j} \mu_j g_j(y_t, X), \tag{4}$$

with parameters $\theta = (\lambda_1, \lambda_2, \ldots; \mu_1, \mu_2, \ldots)$, to be estimated from training data. f_i and g_j are feature functions for inter-label and observation-label 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. Inter-label feature functions f_j are:

$$f_i(y_t, y_{t-1}, X) = [y_t = l \wedge y_{t-1} = l'], \tag{5}$$

$l, l' \in \chi$. $[e]$ is equal to 1 if logical expression e is true, 0 otherwise. Given a temporal context window of size $2W+1$ around the current observation, observation-label feature functions g_j are:

$$g_j(y_t, X) = [y_t = l][D(x_{t-w}) = o], \tag{6}$$

$l \in \chi$,$o \in \tau$, $w \in [-W, W]$. τ is a set of discreted observation value. w is size of a context window around the current observation. $D(x_t) = \lceil x_t/H \rceil$, which is a function to convert float x_t into an integer. To compact the size of the discreted observation set τ, x_t is divided by H first, which is set to 6 in this paper.

CRFs could be trained by searching the set of weights $\theta = \{\lambda_1, \lambda_2, \ldots; \mu_1, \mu_2, \ldots\}$ to maximize the log-likelihood, L_θ, of a given training data set $D = \{Y^d, X^d\}_{d=1\ldots N}$

$$L_\theta = \sum_{d=1}^{N} log(p_\theta(Y^d|X^d)) = \sum_{d=1}^{N}(\sum_{t=1}^{T} F_\theta(y_t^d, y_{t-1}^d, X^d) - logZ_\theta(X^d)). \qquad (7)$$

Learning the model parameters leads to a convex problem with guaranteed global optimality. This optimization can be solved by a gradient ascent (BFGS) method[9], and inference can be performed efficiently using dynamic programming as HMM.

3 Observations in CRF-Based Blinking Model

Observation features can be selected variously according to different problems. For example, silhouette features are commonly used in human motion recognition [14,15,16]. Blinking activity is an action represented by the image sequence which consists of images with close and non-close state. Observation features that convey the eye state would be much helpful for the blinking model. We, integrating Adaboost algorithm [12], define the observation function as:

$$x_t(I) = \sum_{i=1}^{N}(log(1/\beta_i)h_i(I)), \qquad (8)$$

where $h(I, f, p, \theta)$ is a weak classifier which consists of a feature(f), a threshold(θ) and a polarity(p) indicating the direction of the inequality:

$$h(I, f, p, \theta) = \begin{cases} 1 & if \quad pf(I) < p\theta \\ 0 & otherwise \end{cases}. \qquad (9)$$

The parameter β_i and $h_i(I, f, p, \theta)$ can be obtained using Adaboost algorithm. In our implementation, the positive samples and negative samples are for open state and close state respectively. The over-complete haar-like features [11] are chosen as the feature(f). Some observation samples for the whole blinking activity are shown in Fig. 2.

4 Experiments

4.1 Database

In order to test our approach, we build a video database including blinking video clips and imposter video clips, which are captured by Logitech Pro5000, a

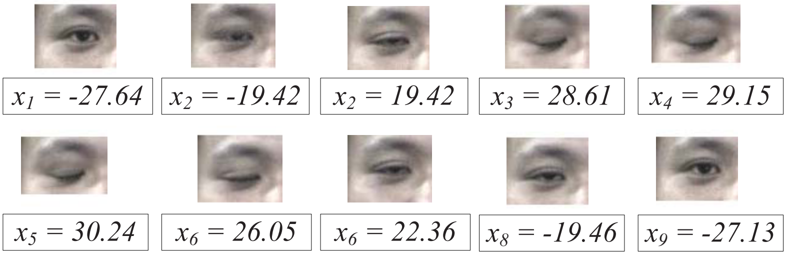

Fig. 2. The samples of observations in CRFs for an entire blinking activity. The observation value x_t of each frame is under the corresponding frame.

common web camera. There are totally 80 clips in blinking video database for 20 individuals, 4 clips for each individual: the first clip without glasses in frontal view, the second clip with thin rim glasses in frontal view, the third clip with black frame glasses in frontal view, and the last clip without glasses in upward view, shown in Fig. 4. Each video clip is about five seconds length with 30 fps and size of 320 × 240. The blinking number varies from 1 to 6 times in each clip. A typical blinking activity in our blinking videos is shown in Fig. 3. To test the ability against photo imposters, we also sampled 180 photo imposter video clips of 20 persons with various motions of photo, including rotating, folding and moving.

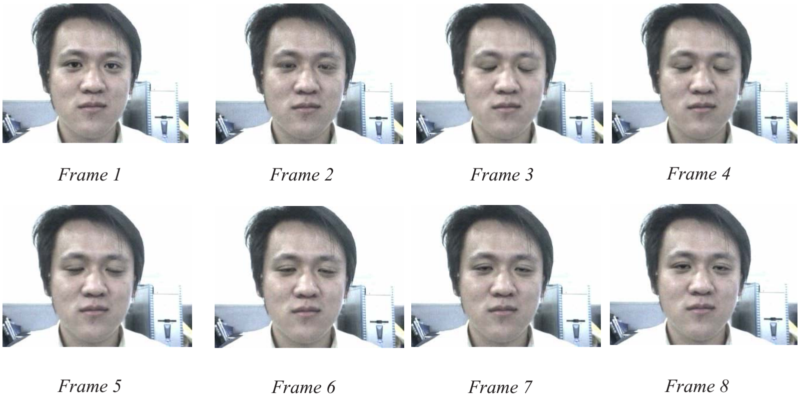

Fig. 3. A blinking example from blinking video database

4.2 CRF Training

State labels for CRF training, C and NC, are labeled manually. Observations calculating is base on the eye images that extracted from face images by OMRON Vision^{TM} software library and resized to 24 × 24 pixels. We have trained CRF that can model long-range dependencies between observations to various degrees.

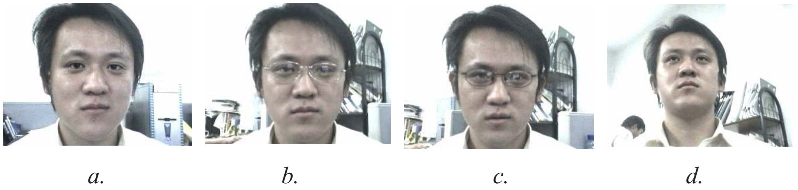

Fig. 4. Four styles in blinking video database. (a) Frontal + without glasses. (b) Frontal + thin rim glasses. (c) Frontal + black frame glasses. (d) Upward+without glasses.

Here window size $w = 2$ is chosen, meaning that we considered contexts of observations of size 5 centered at the current observation. Observation sequences and label sequences of five persons are used to estimate parameters of CRF.

The parameters β_i are obtained by AdaBoost, with 1016 labeled close eyes and 1200 open eyes, which are scaled to a base resolution of 24×24 pixels. Close eye samples are from CAS-PEAL-R1 face database [20] and Asian Face Image Database PF01 [19]. Open eye samples are from FERET face database [18]. The samples of close eye and open eye are shown in Fig. 5.

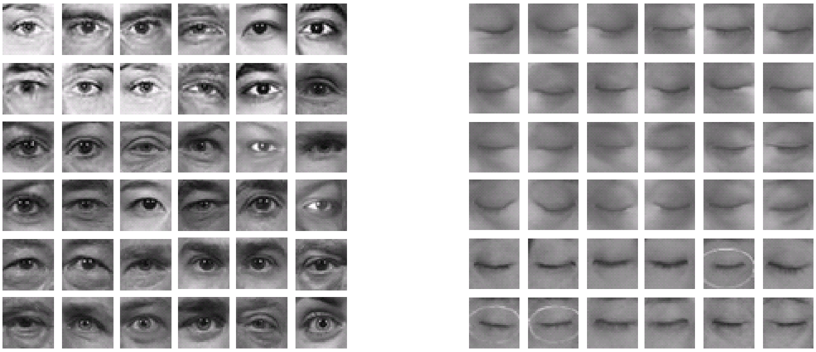

Fig. 5. Training samples for computing observation

4.3 Results

Using the blinking video database, we compare the CRF-based blinking detection with cascaded Adaboost and HMM approaches. Table 1 shows one eye blinking detected rate for the three methods. Table 2 shows two eyes' blinking detection rate, assuming that a blinking is detected if either left eye or right eye blinking is detected. Table 3 shows live face detection rate for the three models, assuming that live face is detected if there is one blinking of either left or right eye is detected in the clip. Table 1, 2 and 3 demonstrate that CRF-based approach outperforms the other two methods in blinking detection from Table 1 and 2 and also has less effect for glasses-wearing and upward view than the other two

Table 1. One-eye blinking detection rate for cascaded AdaBoost, HMM and CRF

Different styles	Cascaded AdaBoost	HMM	CRF($w = 2$)
Without glasses	96.5%	73.7%	98.2%
With thin rim glasses	60.0%	46.9%	68.5%
With black frame glasses	46.9%	39.1%	75.0%
Upward without glasses	52.5%	43.4%	77.0%

Table 2. Two-eye blinking detection rate for cascaded AdaBoost, HMM and CRF

Different styles	Cascaded AdaBoost	HMM	CRF($w = 2$)
Without glasses	98.2%	82.5%	100%
With thin rim glasses	80.0%	63.1%	84.6%
With black frame glasses	71.9%	50.0%	92.2%
Upward without glasses	62.3%	50.9%	100%

Table 3. Live face detection rate for cascaded AdaBoost, HMM and CRF

Different styles	Cascaded AdaBoost	HMM	CRF($w = 2$)
Without glasses	100%	100%	100%
With thin rim glasses	95%	95%	90%
With black frame glasses	80%	85%	100%
Upward without glasses	85%	95%	100%

Table 4. Imposter detection rate for cascaded AdaBoost, HMM and CRF

Cascaded AdaBoost	HMM	CRF($w = 2$)
95%	97.8%	98.3%

methods. Table 4 shows imposter detection rate for these three models tested by imposter video database.

5 Conclusions

This paper presented a CRF-based framework for face liveness detection using blinking clue. CRFs, which can accommodate arbitrary overlapping features of the observation as well as long-range contextual dependencies on observation sequence, are suitable for modeling blinking activities for liveness detection. The experimental results with cascaded AdaBoost, CRF and HMM show that CRF model significantly outperforms the other two models both in liveness detection and imposter detection.

Acknowledgements

This work is partly supported by NSF of China (60503019, 60525202), Key Program of NSF of China (60533040), Program for New Century Excellent Talents in University (NCET-04-0545) and a grant from OMRON corporation. The authors would be grateful to Mr. Yoshihisa Ijiri and Mr. Motoo Yamamoto for valuable discussions.

References

1. Jain, A., Bolle, R., Pankanti, S.: Personal Identification in Networked Society. Springer, Heidelberg (1999)
2. Robert, W.F., Ulrich, D.: BioID: A Multimodal Biometric Identification System. IEEE Computer 33(2), 64–68 (2000)
3. Choudhury, T., Clarkson, B., Jebara, T., Pentland, A.: Multimodal person recognition using unconstrained audio and video. In: Proc. 2nd Int. Conf. Audio-Video Based Person Authentication, pp. 176–181 (1999)
4. Li, J.W., Wang, Y.H., Tan, T.N., Jain, A.K.: Live Face Detection Based on the Analysis of Fourier Spectra. In: Proc. SPIE. Biometric Technology for Human Identification, vol. 5404, pp. 296–303 (2004)
5. Kollreider, K., Fronthaller, H., Bigun, J.: Evaluating liveness by face images and the structure tensor. AutoID (2005)
6. Moriyama, T., Kanade, T., Cohn, J.F., Xiao, J., Ambadar, Z., Gao, J., Imamura, H.: Automatic Recognition of Eye Blinking in Spontaneously Occurring Behavior. In: Proc. Int. Conf. on Pattern Recognition, vol. 4, pp. 78–81 (2002)
7. Lafferty, J., McCallum, A., Pereira, F.: Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In: Proc. 18th Int. Conf. Machine Learning, pp. 282–289 (2001)
8. Sha, F., Pereira, F.: Shallow Parsing with Conditional Random Fields. Proc. Human Language Technology, NAACL, 213–220 (2003)
9. McCallum, A.: Efficiently Inducing Features of Conditional Random Fields. Proc. 19th Uncertainty in Artificial Intelligence, 403–410 (2003)
10. Hammersley, J., Clifford, P.: Markov fields on finite graphs and lattices (Unpublished manuscript 1971)
11. Lienhart, R., Kuranov, A., Pisarevsky, V.: Empirical Analysis of Detection Cascades of Boosted Classifiers for Rapid Object Detection. In: Proc. 25th German Pattern Recognition Symposium, pp. 297–304 (2003)
12. Freund, Y., Schapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences 55(1), 119–139 (1997)
13. Rabiner, L.: A tutorial on hidden markov models and selected applications in speech recognition. In: Proc. IEEE (1989)
14. Cristian, S., Kanaujia, A., Li, Z.G., Metaxas, D.: Conditional Models for Contextual Human Motion Recognition. In: Proc. Int. Conf. Computer Vision, pp. 1808–1815 (2005)
15. Bobick, A., Davis, J.: The recognition of human movement using temporal templates. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 928–934 (2001)

16. Gavrila, D.: The Visual Analysis of Human Movement: A Survey. Computer Vision and Image Understanding 73(1), 82–98 (1999)
17. Viola, P., Jones, M.J.: Rapid Object Detection using a Boosted Cascade of Simple Features. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 511–518 (2001)
18. Phillips, P.J., Moon, H., Rauss, P.J., Rizvi, S.: The FERET Evaluation Methodology for Face-Recognition Algorithms. IEEE Trans. Pattern Analysis and Machine Intelligence 22(10), 1090–1104 (2000)
19. Hyoja-Dong, Nam-Gu: Asian Face Image Database PF01. Technical Report, Pohang University of Science and Technology `http://nova.postech.ac.kr`
20. Zhang, X.H., Shan, S.G., Cao, B., Gao, W., Zhou, D.L., Zhao, D.B.: CAS-PEAL: A Large-Scale Chinese Face Database and Some Primary Evaluations. Journal of Computer-Aided Design and Computer Graphics 17(1), 9–17 (2005) (in Chinese)
21. Karson, C.N.: Spontaneous eye-blink rates and dopaminergic systems. Brain 106, 643–653 (1983)