FACE ANTI-SPOOFING BASED ON DYNAMIC COLOR TEXTURE ANALYSIS USING LOCAL DIRECTIONAL NUMBER PATTERN

Junwei Zhou[1], Ke Shu[1], Peng Liu[2], Jianwen Xiang[1] and Shengwu Xiong[1]
[1]School of Computer Science and Technology, Wuhan University of Technology, Wuhan, China
[2]College of Information Sciences and Technology, The Pennsylvania State University, Pennsylvania 16802, USA
junweizhou@msn.com

Abstract

Face anti-spoofing is becoming increasingly indispensable for face recognition systems, which are vulnerable to various spoofing attacks performed using fake photos and videos. In this paper, a novel "LDN-TOP representation followed by ProCRC classification"pipeline for face anti-spoofing is proposed. We use local directional number pattern (LDN) with the derivative-Gaussian mask to capture detailed appearance information resisting illumination variations and noises, which can influence the texture pattern distribution. To further capture motion information, we extend LDN to a spatial-temporal variant named local directional number pattern from three orthogonal planes (LDN-TOP). The multi-scale LDN-TOP capturing complete information is extracted from color images to generate the feature vector with powerful representation capacity. Finally, the feature vector is fed into the probabilistic collaborative representation based classifier (ProCRC) for face anti-spoofing. Our method is evaluated on three challenging public datasets, namely CASIA FASD, Replay-Attack database, and UVAD database using sequence-based evaluation protocol. The experimental results show that our method can achieve promising performance with 0.37% EER on CASIA and 5.73% HTER on UVAD. The performance on Replay-Attack database is also competitive.

Introduction

In summary, the main contributions of this work are:

(1) We propose a novel pipeline for face anti-spoofing based on LDN and ProCRC. The LDN is more robust to noises and illumination variations and can capture more texture information compared to other texture descriptors such as LBP.

(2) We extend the LDN to spatial-temporal variant LDN-TOP. It can capture both spatial texture and motion information, which are useful for discriminating real faces from fake ones. The multi-scale LDN-TOP is extracted from color images instead of gray-scale ones incorporating useful color information.

(3) Our method achieved the lowest EER of 0.37% on CASIA FASD **?** and the lowest HTER of 5.37% on UVAD **?** outperforming the existing methods and showed competitive results on Replay-Attack database **?**. We opensource our implementation on GitHub[1].

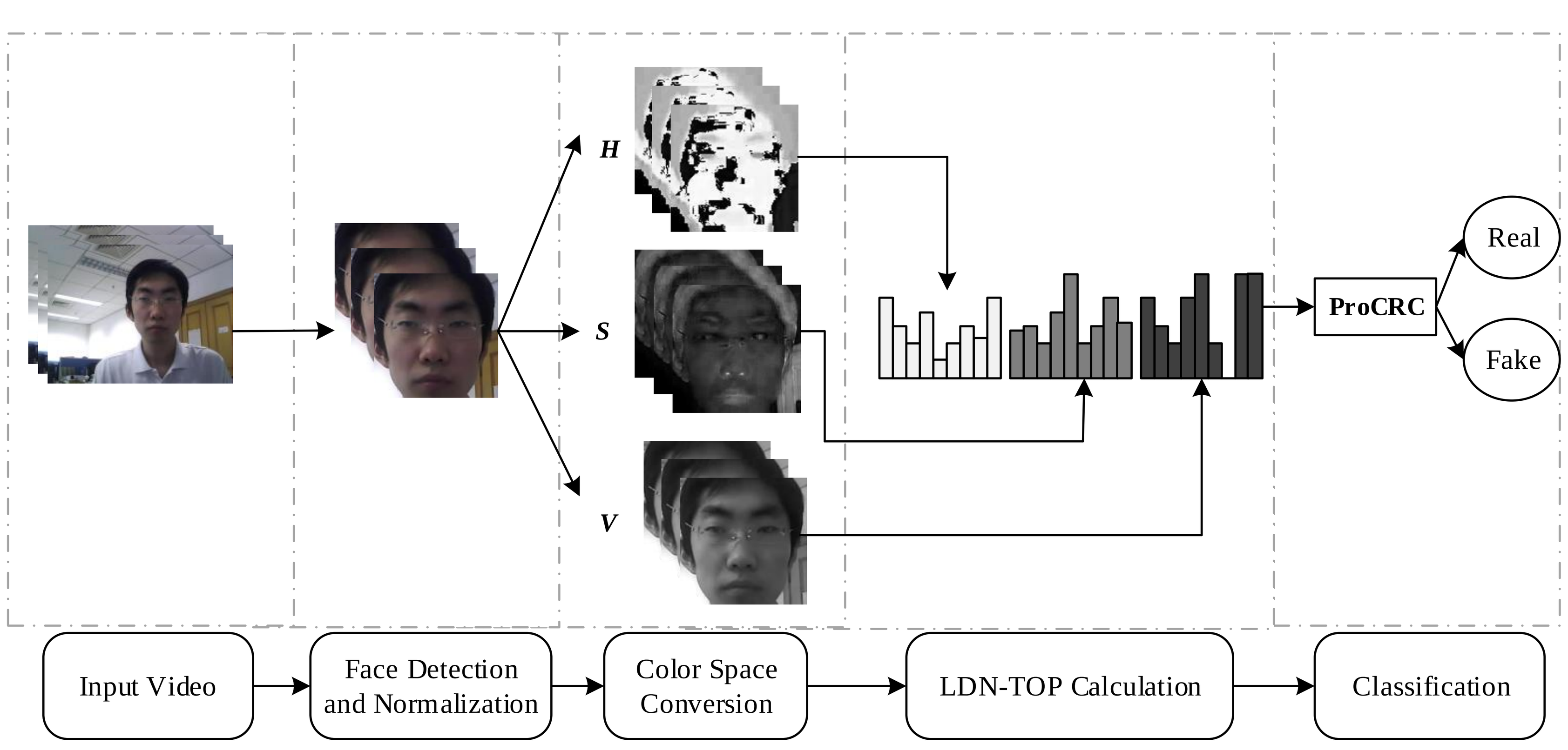

Figura 1: The framework of the proposed method. There are four major steps: 1) face detection and normalization, 2) color space conversion (e.g., HSV), 3) LDN-TOP calculation and 4) classification.

The Proposed Method

In this paper, we propose a face anti-spoofing countermeasure based on LDN and ProCRC. The LDN descriptor encodes each pixel using the complete information among its neighborhood resulting in a more informative representation. Moreover, it is more robust to illumination variations and noises which are essential factors involved in the face anti-spoofing task with the aid of derivative-Gaussian masks. The multi-scale representation which can improve the sensitivity to the variations of input image resolution, is realized by combining LDN histograms calculated using different Gaussian bell widths. To capture motion information which is also useful for detecting spoofing attacks, we extend the LDN to its dynamic variant LDN-TOP. Considering that the texture features calculated from the gray-scale image are sensitive to acquisition conditions such as lighting variations, while the color image contains not only the luminance component but also the chroma component which is useful for discriminating real faces from fake ones. Therefore, we calculate LDN-TOP features on the color images. We further use the ProCRC to detect fake faces, since the ProCRC has advantages over SVM in terms of three aspects: 1) higher interpretability, 2) lower computational complexity, and 3) higher overall classification accuracy.

The whole framework is shown in Fig. 1. Given an input video, we first locate the face part and normalize each facial frame to the fixed size. Then, the resultant facial image sequence is transformed into color spaces (e.g., HSV). The LDN-TOP feature vectors are extracted from each color channel and combined into one single vector. Finally, the ProCRC is used for discriminating real faces from fake ones.

[1]Source code available at: https://github.com/BetaSK/ldn-top-FAS

Experimental Results and Analysis

	Method	C → R[1]	R → C[2]	Average
Hand-Crafted Features	LBP-TOP [5]	49.7	60.6	55.2
	Guided Scale Texture [25]	48.4	40.3	44.4
	Color LBP [15]	47.0	39.6	43.3
	Visual Codebooks [17]	34.4	50.0	42.2
	CLNF [36]	33.7	49.3	41.5
	Videolet [37]	35.4	44.6	40.0
	Color Texture Markov [16]	32.3	45.9	39.1
	IQA [33]	38.1	39.0	38.6
	Colour Texture [3]	30.3	37.7	34.0
Deep Learning	Noise Modeling [38]	28.5	41.1	34.8
	Attention CNN [27]	30.0	33.4	31.7
	CNN-Stacking [26]	**20.6**	40.4	30.5
	Auxiliary [21]	27.6	28.4	28.0
	GFA-CNN [39]	21.4	34.3	27.9
	DDGL [9]	22.8	**27.4**	**25.1**
	Ours	38.8	42.9	40.9

[1] "C → R": Train on CASIA FASD and then test on Replay-Attack.
[2] "R → C": Train on Replay-Attack and then test on CASIA FASD.

	Method	CASIA EER	Replay-Attack HTER
Hand-Crafted Features	CASIA Baseline [11]	17	–
	LBP-TOP [5]	10	7.6
	Color Texture Markov [16]	8	4.4
	MBSIF-TOP + MLPQ-TOP [6]	7.2	1.0
	CVLBC [7]	6.48	0.75
	Colour LBP [15]	6.2	2.9
	Scale Space Texture [14]	4.2	3.1
	Colour Texture [3]	3.2	3.5
	Guided Scale Texture [25]	2.53	3.13
Deep Learning	CNN-Stacking [26]	6.7	0.38
	Attention CNN [27]	3.145	**0**
	Patch & Depth based CNN [20]	2.67	0.72
	LBP Network [28]	2.5	1.3
	Deep LBP [29]	2.3	0.9
	3D Virtual Synthesis [30]	2.22	0.63
	3D CNN [31]	1.4	1.2
	Ours	**0.37**	3.13

Conclusions

In this paper, we proposed a novel and effective face anti-spoofing method based on dynamic color texture and probabilistic collaborative representation based classifier (ProCRC). The experiments showed that the combination of HSV and YCbCr achieved the best results and the ProCRC outperformed the SVM classifiers with two different kernels. We also investigated the effectiveness of time window length and the results indicated that the larger time window length leads to better performance if the video is long enough to provide required frames. Our method achieved the best performance on CASIA with EER and UVAD with HTER. And our method also showed the competitive results on the Replay-Attack database.