

# EFFECTS OF SEVERE SIGNAL DEGRADATION ON EAR DETECTION

*J. Wagner, A. Pflug, C. Rathgeb and C. Busch*

da/sec – Biometrics and Internet Security Research Group  
Hochschule Darmstadt, Darmstadt, Germany

{johannes.wagner, anika.pflug, christian.rathgeb, christoph.busch}@cased.de

## ABSTRACT

Ear recognition has recently gained much attention, as for surveillance scenarios identification remains feasible, in case the facial characteristic is partly or fully covered. However video footage stemming from surveillance cameras is often of low quality. In this work we investigate the impact of signal degradation, i.e. out-of-focus blur and thermal noise, on the segmentation accuracy of automated ear detection. Realistic acquisition scenarios are constructed and various intensities of signal degradation are simulated on a comprehensive dataset. In experiments different ear detection algorithms are employed, pointing out the effects of severe signal degradation on ear segmentation performance.

**Index Terms**— Ear biometrics, ear detection, signal degradation, simulation, surveillance

## 1. INTRODUCTION

In April 2013, NYPD detectives were able to track down an arsonist who was accused of burning mezuzahs in Brooklyn, USA. Police got onto the suspect's trail through a clear picture of his ear from a surveillance camera. The image was run through the police facial recognition database, which contained a profile image of the person where the outer ear was visible, returning the suspect's name. Thus investigators reported, "*We had a good angle on his ear that helped to identify him*" [1]. Success stories like this and a constantly increasing number of surveillance cameras underline the potential of automated ear recognition for forensic identification and confirm that the intricate structure of the outer ear represents a reliable and stable biometric characteristic. However, the ability to determine a person's identity based on automated ear recognition highly depends on the image quality and resolution [2, 3]. In addition, images captured by surveillance cameras may suffer from signal degradations particularly, in case of outdoor installations.

Automated ear biometric recognition systems hold tremendous promise for the future, especially in the forensic area [4]. While the long standing success story of ear recognition goes back to the 19th century [5] nowadays forensic applications have only recently started to pay attention to automated ear

recognition. In past years numerous approaches focusing on ear detection, feature extraction, and feature comparison have been proposed, achieving promising biometric performance (for a detailed survey on detection and recognition techniques for ears see [6]). However, the vast majority of experimental evaluations are performed on datasets acquired under rather favourable conditions, which in most cases does not reflect image data acquired in forensic scenarios. So far, no studies have been conducted on the impact of signal degradation on automated ear detection, which represents a considerable significant point of failure for any automated ear recognition system.

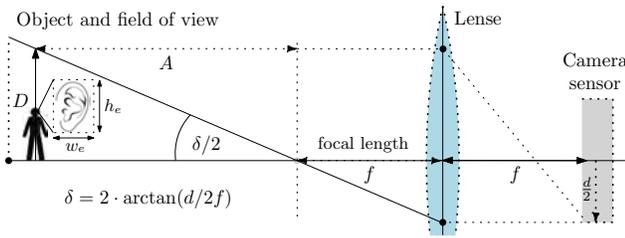
The contribution of this work is the investigation of the effects of severe signal degradation on automated ear detection. Considering different reasonable scenarios of data acquisition (according to surveillance scenarios), profile images of a comprehensive dataset are systematically degraded, simulating frequent distortions, i.e. out-of-focus blur and thermal noise. In our experiments well-established ear detection algorithms are evaluated, which clearly illustrate the impact of signal degradation on ear detection. Furthermore, a detailed discussion of consequential issues is given.

The remainder of this paper is organized as follows: in Sect. 2 considered scenarios and applied signal degradations are described in detail. The effects of signal degradation on ear detection algorithms are investigated in Sect. 3. Finally, conclusions are drawn in Sect. 4.

## 2. ACQUISITION AND SIGNAL DEGRADATION

### 2.1. Acquisition Scenarios

Table 1 summarizes different state-of-the-art surveillance cameras made available by major vendors and relevant characteristics, i.e. focal length, resolution, and sensor type (characteristics refer to currently best products). Based on this comparison we simulate a camera providing (1) a focal length of 8mm, (2) a resolution of  $1920 \times 1080$ , and (3) a sensor diagonal of 1/2.5 inch. We examine two different acquisition scenarios  $\mathcal{S}_1$ ,  $\mathcal{S}_2$  with respect to the distance of the subject to the camera considering distances of 2m and 4m, respectively. Fig. 1 schematically depicts the considered



**Fig. 1.** Simulated data acquisition scenario.

acquisition scenario.

We assume that we are able to detect the presence of a subject in a video by one of the state-of-the-art detection techniques, that are summarized in [7]. After successfully detecting a capture subject, the head region can be roughly segmented (cf. Fig. 2). These pre-segmented images are the basis of further processing, such as ear detection.

**Table 1.** State-of-the-art camera models and characteristics.

Vendor	Product	Focal length	Resolution	Sensor
ACTi <sup>1</sup>	D82	2.8-12mm	1920×1080	1/3.2"
AXIS <sup>2</sup>	P3367V	3-9mm	1920×1080	1/3.2"
GeoVision <sup>3</sup>	GV-FD220G	3-9mm	1920×1080	1/2.5"
Veilux <sup>4</sup>	VVIP-2L2812	2.8-12mm	1920×1080	1/2.5"

<sup>1</sup> <http://www.acti.com/>

<sup>2</sup> <http://www.axis.com/>

<sup>3</sup> <http://www.geovision.com.tw/>

<sup>4</sup> <http://www.veilux.net/>

Let  $C(f, d, w, h)$  be a camera with focal length  $f$ , sensor diagonal  $d$ , and resolution  $w \times h$ . Then the diagonal  $D$  of the field of view at a distinct distance  $A$  is estimated as,

$$D = A \cdot \tan(2 \cdot \arctan((d/2f)/2)) = A \cdot d/2f. \quad (1)$$

In our scenario the aspect ratio is 16:9, i.e. the field of view in object space corresponds to

$$16 \cdot \sqrt{D^2/(16^2 + 9^2)} \text{ m} \times 9 \cdot \sqrt{D^2/(16^2 + 9^2)} \text{ m}. \quad (2)$$

In [8] the average size of the outer ear of males and females across different age group is measured as 61.7 mm × 37.0 mm and 57.8 mm × 34.5 mm, respectively. For an average angle of auricle of 32.5 degrees across age groups and sex we approximate the bounding box of an ear of any subject as 70 mm × 60 mm. For both scenarios  $\mathcal{S}_1, \mathcal{S}_2$  the considered camera  $C(8\text{mm}, 1/2.5'', 1920\text{px}, 1080\text{px})$  would yield images where ear regions comprise approximately  $w_e \times h_e = 110 \times 90$  and  $55 \times 45$  pixels, respectively.

**Table 2.** Blur and noise conditions considered for signal degradation (denotations of  $\sigma$  are defined in 2.2.1 and 2.2.2).

Blur condition		Noise condition		Intensity
Abbrev.	Description	Abbrev.	Description	
B-0	–	N-0	–	none
B-1	$\sigma = 2$	N-1	$\sigma = 20$	low
B-2	$\sigma = 3$	N-2	$\sigma = 25$	medium
B-3	$\sigma = 4$	N-3	$\sigma = 30$	high

## 2.2. Signal Degradation

Signal degradation in this work is simulated by means of blur and noise where blur is applied prior to noise (out-of-focus blur is caused before noise occurs). Four different intensities (including absence) of blur and noise and combinations of these are considered and summarized in Table 2.

### 2.2.1. Blur Conditions

Out-of-focus blur represents a frequent distortion in image acquisition mainly caused by an inappropriate distance of the camera to the eye (another type of blur is motion blur caused by rapid movement which is not considered in this work). We simulate the point spread function of the blur as a Gaussian

$$f(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\pi\sigma^2}} \quad (3)$$

which is then convolved with the specific image, where the image is divided into  $16 \times 16$  pixel blocks.

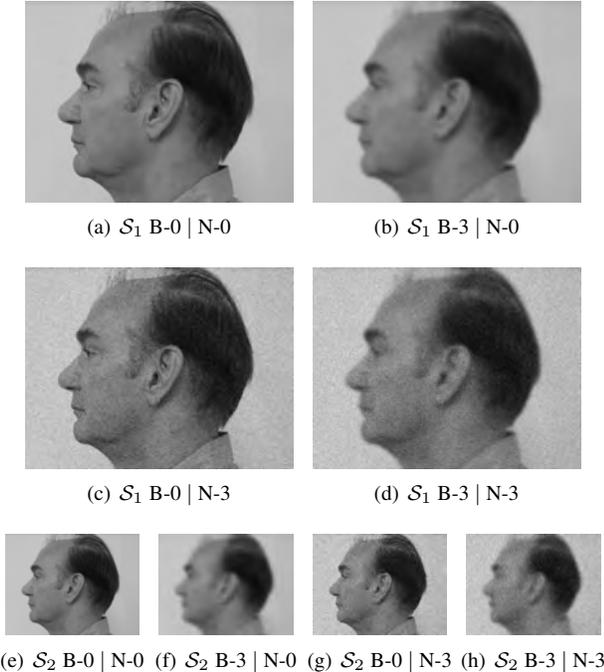
### 2.2.2. Noise Conditions

Amplifier noise is primarily caused by thermal noise. Due to signal amplification in dark (or underexposed) areas of an image, thermal noise has a high impact on these areas. Additional sources contribute to the noise in a digital image such as shot noise, quantization noise and others. These additional noise sources however, only make up a negligible part of the noise and are therefore ignored during this work.

Let  $P$  be the set of all pixels in image  $I \in \mathbb{N}^2$ ,  $w = (w_p)_{p \in P}$ , be a collection of independent identically distributed real-valued random variables following a Gaussian distribution with mean  $m$  and variance  $\sigma^2$ . We simulate thermal noise as additive Gaussian noise with  $m = 0$ , variance  $\sigma^2$  for pixel  $p$  at  $x, y$  as

$$N(x, y) = I(x, y) + w_p, p \in P, \quad (4)$$

with  $N$  being the noisy image, for an original image  $I$ . Examples of results of simulated signal degradation are depicted in Fig. 2 for images considered in both scenarios.



**Fig. 2.** Maximum intensities of blur and/ or noise (b)-(d) and (f)-(h) applied to the image (a) and (e) (id\_02463d677).

### 3. EXPERIMENTAL EVALUATIONS

#### 3.1. Experimental Setup

For our evaluation, we have composed a dataset of mutually different images of the UND-G [9], UND-J2 [10] and UND-NDOff-2007 [11] database. The merged dataset contains 2369 left profile images from 510 subjects with yaw poses between  $-60$  and  $-90$  degrees. Right profile views from UND-G were mirrored horizontally. The manually annotated ground truth in form of ear bounding boxes yields an average size of  $125 \times 95$  pixels for the entire data set, i.e. original images are employed in scenario  $\mathcal{S}_1$ . For the second scenarios  $\mathcal{S}_2$  images are scaled with factor 0.5 prior to applying blur and noise.

We evaluate the performance of cascaded object detectors with a fixed-size sliding window. The object detectors are trained with (1) Haar-like features [12], (2) local binary patterns [13] (LBP) and (3) histograms of oriented gradients [14](HOG). The detectors were trained with images from WPUTEDB [15] and negative samples from the INIRA person detection dataset.

#### 3.2. Performance Evaluation

We calculate the detection error  $E$  from the segmentation result  $S(I)$  for an image  $I$  with corresponding ground truth mask  $G_I$  (both of dimension  $m \times n$  pixels), such that for

**Table 3.** Error rates for different detection algorithms for both scenarios (errors have been multiplied by  $10^2$ ). Results are visualized in Fig 3.

Blur	Noise	Scenario $\mathcal{S}_1$			Scenario $\mathcal{S}_2$				
		PSNR	$E_1$	$E_2$	$E_3$	PSNR	$E_1$	$E_2$	$E_3$
B-0	N-0	$\infty$	2.12	2.86	1.94	$\infty$	2.64	2.75	1.97
B-1	N-0	33.69 db	2.62	2.91	2.10	31.66 db	3.77	3.50	5.56
B-2	N-0	32.18 db	2.89	3.12	2.88	29.58 db	4.18	3.90	6.01
B-3	N-0	31.20 db	3.22	3.24	3.42	28.38 db	4.30	3.87	6.08
B-0	N-1	22.56 db	2.08	2.77	3.45	22.56 db	2.09	3.23	3.18
B-1	N-1	22.21 db	2.09	2.82	3.76	22.01 db	2.34	3.37	5.51
B-2	N-1	22.07 db	2.09	2.59	4.02	21.71 db	2.53	3.41	5.94
B-3	N-1	21.96 db	2.16	2.77	4.00	21.48 db	2.61	3.58	6.12
B-0	N-2	20.74 db	2.19	2.70	3.75	20.74 db	2.18	3.20	3.73
B-1	N-2	20.50 db	2.17	2.80	4.00	20.35 db	2.39	3.40	5.30
B-2	N-2	20.40 db	2.17	2.59	4.14	20.14 db	2.56	3.52	5.60
B-3	N-2	20.32 db	2.22	2.77	4.14	19.97 db	2.66	3.59	5.29
B-0	N-3	19.27 db	2.22	2.82	3.95	19.26 db	2.27	3.40	4.04
B-1	N-3	19.09 db	2.25	2.68	4.13	18.98 db	2.40	3.52	4.86
B-2	N-3	19.02 db	2.27	2.57	4.18	18.82 db	2.59	3.50	4.87
B-3	N-3	18.96 db	2.33	2.84	4.18	18.69 db	2.77	3.74	4.82

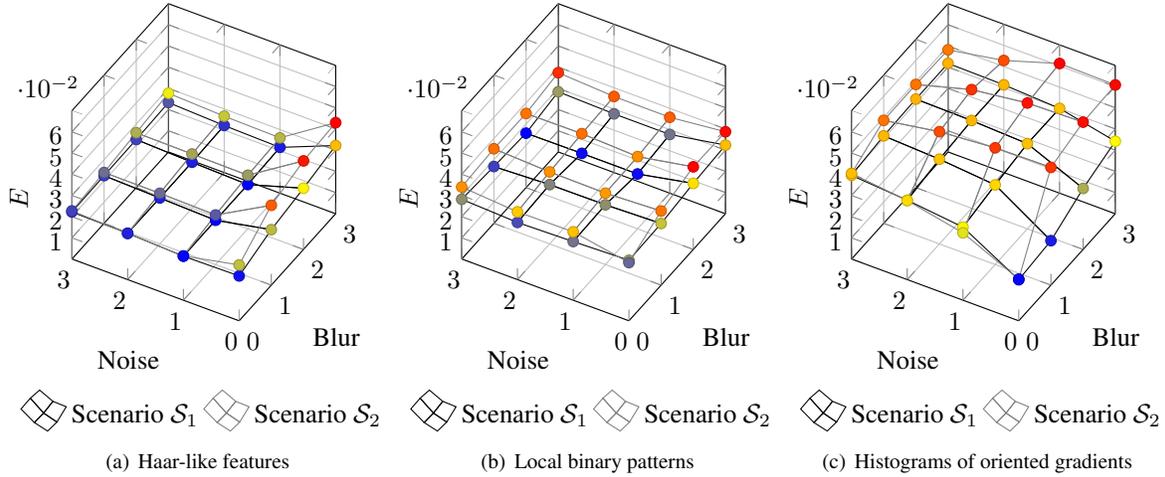
all positions  $x, y$ ,  $G_I[x, y] = 1$  labels “ear” pixels (otherwise  $G_I[x, y] = 0$ ), as

$$E = \frac{1}{m \cdot n} \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} G_I[x, y] \oplus S(I)[x, y]. \quad (5)$$

Table 3 summarizes the detection errors for different detection algorithms for intensities of blur, noise and combination of these for both considered scenarios. The quality of generated images is estimated in terms of average peak signal to noise ratio (PSNR). Errors  $E_1$ ,  $E_2$  and  $E_3$  correspond to the detection results employing Haar-like, LBP, and HOG features, respectively. In Fig. 3 detection errors are plotted for all detection algorithms and scenarios for all combinations of signal degradation. See Table 4 for a collection of examples for ground truth and detection under different conditions.

#### 3.3. Discussion

As can be seen in Fig. 3 and Table 3, Haar-like features turn out to be most robust against noise, followed by LBP where we observe slightly higher error rate. Haar-like features rely on the position of edges and are using the ratio of dark and light pixels in an image patch. This makes these features robust to noise but vulnerable to blur. Combinations between blur and noise perform better, because adding noise after blur results in images with more intense edges.



**Fig. 3.** Errors for different segmentation algorithms for intensities of blur, noise and combination of these for both scenarios.

For LBP, we obtain a mostly stable, although higher, error rate than for Haar-like features. Generally, degradations have the least impact in LBP, because it encodes local texture information as a histogram without the need of particular local features, such as edges. Alterations on a pixel level add noise uniformly for all local features to a mostly skin coloured texture, which can be compensated by the detector.

HOG performs well under ideal conditions, but with increasing noise and blur, the accuracy degenerates quickly. Blur and noise alter the local gradient orientation, length and direction of the image, which makes it difficult for HOG to match the trained pattern with local texture information. Although noise and blur are causing this effect, blur has a significantly larger impact on local gradients than noise.

#### 4. CONCLUSION

We have quantified the impact of signal degradation in particular, noise and blur, on ear detection systems for surveillance purposes. Experiments were carried out for three well-established detection algorithms, Haar-like features, LBPs and HOG. With respect to the simulated conditions, the tested algorithms turn out to be vulnerable to both, to blur and noise. Our future work will comprise research on other forms of signal degradations as well as the impact of signal degradation on feature extraction and recognition performance.

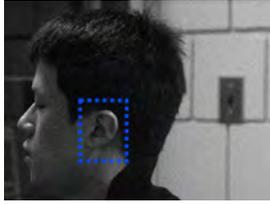
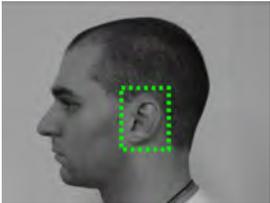
#### 5. ACKNOWLEDGEMENTS

This work is partially funded by the Federal Ministry of Education and Research (BMBF) of Germany, the European FP7 FIDELITY project (SEC-2011-284862) and the Center for Advanced Security Research Darmstadt (CASED).

#### 6. REFERENCES

- [1] NY Daily News, “Mezuzah arsonist snagged by an ear thanks to facial recognition technology,” 2013, 11th April, 2013.
- [2] A. J. Hoogstrate, H. Van Den Heuvel, and E. Huyben, “Ear identification based on Surveillance Samera Images,” *Science & Justice*, vol. 41, pp. 167 – 172, 2001.
- [3] C. Sanderson, A. Bgdeli, T. Shan, S. Chen, E. Berglund, and B. C. Lovell, “Intelligent CCTV for Mass Transport Security: Challenges and Opportunities for Video and Face Processing,” *Electronic Letters on Computer Vision and Image Analysis*, vol. 6, pp. 30–41, 2007.
- [4] A. Abaza, A. Ross, C. Hebert, M. A. F. Harrison, and M. S. Nixon, “A survey on ear biometrics,” *ACM Comput. Surv.*, 2013.
- [5] A. Bertillon, *La Photographie Judiciaire: Avec Un Appendice Sur La Classification Et L’Identification Anthropométriques*, Gauthier-Villars, Paris, 1890.
- [6] A. Pflug and C. Busch, “Ear biometrics: a survey of detection, feature extraction and recognition methods,” *Biometrics, IET*, vol. 1, pp. 114–129, 2012.
- [7] N. A. Ogale, “A survey of techniques for human detection from video,” *Survey, University of Maryland*, 2006.
- [8] C. Sforza, G. Grandi, M. Binelli, D. G. Tommasi, R. Rosati, and V. F. Ferrario, “Age- and sex-related changes in the normal human ear,” *Forensic Science International*, vol. 2009, pp. 110e1–110e7, 2009.
- [9] P. Yan and K. W. Bowyer, “An Automatic 3D Ear Recognition System,” in *3rd Symposium on 3D Data Processing, Visualization, and Transmission*, 2006.

**Table 4.** Detection accuracy of different feature sets with strong noise and blur. The left column shows the ground truth, the middle column shows the detection result in the original image and the right column shows the result after signal degradation.

		
(a) Ground Truth	(b) B-0 N-0 — Haar Detection	(c) B-0 N-3 — Haar Detection
		
(a) Ground Truth	(b) B-0 N-0 — Haar Detection	(c) B-0 N-1 — Haar Detection
		
(a) Ground Truth	(b) B-0 N-0 — HOG Detection	(c) B-0 N-3 — HOG Detection
		
(a) Ground Truth	(b) B-0 N-0 — HOG Detection	(c) B-3 N-0 — HOG Detection
		
(a) Ground Truth	(b) B-0 N-0 — LBP Detection	(c) B-3 N-0 — LBP Detection
		
(a) Ground Truth	(b) B-0 N-0 — LBP Detection	(c) B-3 N-3 — LBP Detection

- [10] P. Yan and K. W. Bowyer, "Biometric Recognition Using 3D Ear Shape," *Pattern Analysis and Machine Intelligence*, vol. 29, pp. 1297–1308, 2007.
- [11] T. C. Faltemier, K. W. Bowyer, and P. J. Flynn, "Rotated Profile Signatures for robust 3D Feature Detection," in *Automatic Face and Gesture Recognition*, 2008.
- [12] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, 2001, vol. 1, pp. 511–518.
- [13] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 7, pp. 971–987, Jul 2002.
- [14] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, June 2005, vol. 1, pp. 886–893.
- [15] D. Frejlichowski and N. Tyszkiewicz, "The West Pomeranian University of Technology Ear Database A Tool for Testing Biometric Algorithms," in *Image Analysis and Recognition*. Springer, 2010.