Scattering features for multimodal gait recognition

Srđan Kitić

Technicolor R&I
Home Experience Lab - Data Analytics

GlobalSIP 2017

Table of contents

1. Introduction
2. Normalized scattering for gait signals
3. Performance and wrap-up

Introduction

Identification is a core component in many applications:

- Recommender systems,
- Online banking and commerce,
- Surveillance,
- Gaming,
- Administration etc.

Different biometrics: fingerprint, face, speech, retinal scan, *gait (this work)*...

Each comes with advantages and drawbacks, *e.g.* accuracy or intrusiveness.

Gait-based identification

Prior art - various modalities exploited:

- Video (silhouette) (1, 2): high accuracy, privacy issues.
- Mechanical force sensors (3, 4): high setup cost.
- Wearables (5, 6): instrusive.
- WiFi (7): limited accuracy and range.
- Sound (8, 9, 10, 11): (assuming VAD) privacy-preserving, wideband, widespread availability.
- Seismic (12): privacy-preserving, robust, secure, narrowband.

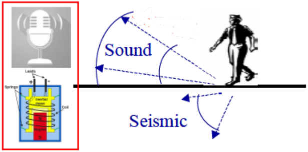
Complementary properties of sound and seismic cues indicate that a *bimodal* approach may be effective.

Gait-based identification

Prior art - various modalities exploited:

- Video (silhouette) (1, 2): high accuracy, privacy issues.
- Mechanical force sensors (3, 4): high setup cost.
- Wearables (5, 6): instrusive.
- WiFi (7): limited accuracy and range.
- Sound (8, 9, 10, 11): (assuming VAD) privacy-preserving, wideband, widespread availability.
- Seismic (12): privacy-preserving, robust, secure, narrowband.

Complementary properties of sound and seismic cues indicate that a *bimodal* approach may be effective.



technicolor

Gait-based idetification

Open set identification:

1. Identify a person, if coming from a known set.
2. Otherwise, decide that the person is unknown.

Addressed through *GMM-UBM framework* (13).

Remaining challenges:

- No publicly available bimodal data.
- No generally acclaimed feature type.
- Seamless feature fusion?

Gait-based idetification

Open set identification:

1. Identify a person, if coming from a known set.
2. Otherwise, decide that the person is unknown.

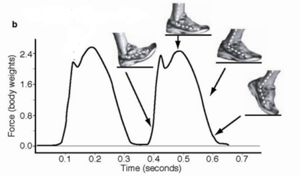
Addressed through *GMM-UBM framework* (13).

Remaining challenges:

- No publicly available bimodal data.
  - We recorded a small scale dataset (size precludes deep learning).
- No generally acclaimed feature type.
  - Tailored *scattering transform* (14) based features.
- Seamless feature fusion?
  - Surprisingly simple - stay tuned.

technicolor

Gait signals



Particle velocity:

$$\hat{v}(\omega) = \mathcal{F}\left(v(\mathrm{t})\right) \propto \mathcal{F}\left(\int \vec{F}_{\mathsf{GRF}} dt\right)$$

Footfall ≈ 0.15s.
Period $\approx 2 \times 0.61$s. (15)

Acquired signals are band-passed and convoluted:

- Sound, *for* $200Hz \lesssim \omega \lesssim 20kHz$:

$$\hat{x}_{\mathsf{a}}(\omega, \vec{r}(\mathrm{t})) = \hat{h}_{\mathsf{a}}(\omega, \vec{r}(\mathrm{t}))\hat{v}(\omega) + \hat{e}_{\mathsf{a}}(\omega) = \hat{g}_{\mathsf{a}}(\omega, \vec{r}(\mathrm{t}))\frac{\hat{v}(\omega)}{\hat{z}(\omega)} + \hat{e}_{\mathsf{a}}(\omega)$$

- Seismic, *for* $20Hz \lesssim \omega \lesssim 300Hz$:

$$\hat{x}_{\mathsf{g}}(\omega, \vec{r}(\mathrm{t})) = \hat{h}_{\mathsf{g}}(\omega, \vec{r}(\mathrm{t}))\hat{v}(\omega) + \hat{e}_{\mathsf{g}}(\omega) = S_{\mathsf{g}}\hat{g}_{\mathsf{g}}(\omega, \vec{r}(\mathrm{t}))\hat{v}(\omega) + \hat{e}_{\mathsf{g}}(\omega).$$

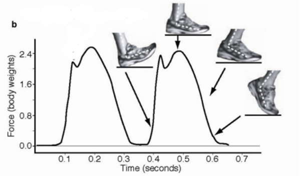
Local stationarity assumption (LSA)

Within (short) temporal segment of duration τ:

$$\hat{g}_{\cdot}(\omega, \vec{r}(\mathrm{t}+\mathrm{t}')) \approx \hat{g}_{\cdot}(\omega, \vec{r}(\mathrm{t})), \text{ analogously } \hat{h}_{\cdot}(\omega, \vec{r}(\mathrm{t}+\mathrm{t}')) \approx \hat{h}_{\cdot}(\omega, \vec{r}(\mathrm{t})).$$

technicolor

Gait signals



Particle velocity:

$$\hat{v}(\omega) = \mathcal{F}\left(v(\mathrm{t})\right) \propto \mathcal{F}\left(\int \vec{F}_{\mathsf{GRF}} dt\right)$$

Footfall ≈ 0.15s.
Period $\approx 2 \times 0.61$s. (15)

Acquired signals are band-passed and convoluted:

- Sound, *for* $200Hz \lesssim \omega \lesssim 20kHz$:

$$\hat{x}_{\mathsf{a}}(\omega, \vec{r}(\mathrm{t})) = \hat{h}_{\mathsf{a}}(\omega, \vec{r}(\mathrm{t}))\hat{v}(\omega) + \hat{e}_{\mathsf{a}}(\omega) = \hat{g}_{\mathsf{a}}(\omega, \vec{r}(\mathrm{t}))\frac{\hat{v}(\omega)}{\hat{z}(\omega)} + \hat{e}_{\mathsf{a}}(\omega)$$

- Seismic, *for* $20Hz \lesssim \omega \lesssim 300Hz$:

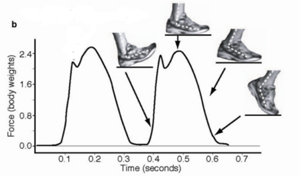
$$\hat{x}_{\mathsf{g}}(\omega, \vec{r}(\mathrm{t})) = \hat{h}_{\mathsf{g}}(\omega, \vec{r}(\mathrm{t}))\hat{v}(\omega) + \hat{e}_{\mathsf{g}}(\omega) = S_{\mathsf{g}}\hat{g}_{\mathsf{g}}(\omega, \vec{r}(\mathrm{t}))\hat{v}(\omega) + \hat{e}_{\mathsf{g}}(\omega).$$

Local stationarity assumption (LSA)

Within (short) temporal segment of duration τ:

$$\hat{g}.(\omega, \vec{r}(\mathrm{t}+\mathrm{t}')) \approx \hat{g}.(\omega, \vec{r}(\mathrm{t})), \text{ analogously } \hat{h}.(\omega, \vec{r}(\mathrm{t}+\mathrm{t}')) \approx \hat{h}.(\omega, \vec{r}(\mathrm{t})).$$

Gait signals



Particle velocity:

$$\hat{v}(\omega) = \mathcal{F}\left(v(\mathrm{t})\right) \propto \mathcal{F}\left(\int \vec{F}_{\mathrm{GRF}} d\mathrm{t}\right)$$

Footfall ≈ 0.15s.
Period $\approx 2 \times 0.61$s. (15)

Acquired signals are band-passed and convoluted:

- Sound, *for* $200Hz \lesssim \omega \lesssim 20kHz$:

$$\hat{x}_{\mathrm{a}}(\omega, \vec{r}(\mathrm{t})) = \hat{h}_{\mathrm{a}}(\omega, \vec{r}(\mathrm{t}))\hat{v}(\omega) + \hat{e}_{\mathrm{a}}(\omega) = \hat{g}_{\mathrm{a}}(\omega, \vec{r}(\mathrm{t}))\frac{\hat{v}(\omega)}{\hat{z}(\omega)} + \hat{e}_{\mathrm{a}}(\omega)$$

- Seismic, *for* $20Hz \lesssim \omega \lesssim 300Hz$:

$$\hat{x}_{\mathrm{g}}(\omega, \vec{r}(\mathrm{t})) = \hat{h}_{\mathrm{g}}(\omega, \vec{r}(\mathrm{t}))\hat{v}(\omega) + \hat{e}_{\mathrm{g}}(\omega) = S_{\mathrm{g}}\hat{g}_{\mathrm{g}}(\omega, \vec{r}(\mathrm{t}))\hat{v}(\omega) + \hat{e}_{\mathrm{g}}(\omega).$$

Local stationarity assumption (LSA)

Within (short) temporal segment of duration τ:

$\hat{g}_.(\omega, \vec{r}(\mathrm{t}+\mathrm{t}')) \approx \hat{g}_.(\omega, \vec{r}(\mathrm{t}))$, analogously $\hat{h}_.(\omega, \vec{r}(\mathrm{t}+\mathrm{t}')) \approx \hat{h}_.(\omega, \vec{r}(\mathrm{t}))$.

Feature extraction

- Signals depend on impact velocity ☺ *and* relative position ☹
- Sound and seismic signals represent different physical quantities.
- To cope, we rely on a "CNN-like" scattering trnsform (16).

Feature extraction up to the *order* p:

0: $S_0(x) = \phi_T * x$,
1: $S_1^{\lambda_1}(x) = \phi_T * |\psi_{\lambda_1} * x|$,
2: $S_2^{\lambda_1,\lambda_2}(x) = \phi_T * |\psi_{\lambda_2} * |\psi_{\lambda_1} * x||$,
...
p: $S_p^{\lambda_1,\dots\lambda_p}(x) =$
$\phi_T * |\psi_p * \dots |\psi_{\lambda_2} * |\psi_{\lambda_1} * x|| \dots |$.

$\phi_T := \phi_T(\mathrm{t})$ - a lowpass $(2\pi/T)$ filter, $\psi_\lambda := \psi_\lambda(\mathrm{t})$ - a complex wavelet at scale λ

Rule of thumb

1. Computational cost increases with T ("time-invariance").
2. $T \propto$ duration of a classified event (crucial for performance!).

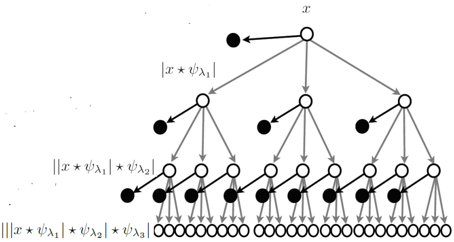
technicolor

Feature extraction

- Signals depend on impact velocity ☺ *and* relative position ☹
- Sound and seismic signals represent different physical quantities.
- To cope, we rely on a “CNN-like” scattering trnsform (16).

Feature extraction up to the *order* p:

0: $S_0(x) = \phi_T * x,$
1: $S_1^{\lambda_1}(x) = \phi_T * |\psi_{\lambda_1} * x|,$
2: $S_2^{\lambda_1,\lambda_2}(x) = \phi_T * |\psi_{\lambda_2} * |\psi_{\lambda_1} * x||,$
...
p: $S_{\mathrm{p}}^{\lambda_1,\ldots\lambda_{\mathrm{p}}}(x) =$
$\phi_T * |\psi_{\mathrm{p}} * \ldots |\psi_{\lambda_2} * |\psi_{\lambda_1} * x|| \ldots |.$



$\phi_T := \phi_T(\mathrm{t})$ - a lowpass $(2\pi/T)$ filter, $\psi_\lambda := \psi_\lambda(\mathrm{t})$ - a complex wavelet at scale λ

Rule of thumb

1. Computational cost increases with T (“time-invariance”).
2. $T \propto$ duration of a classified event (crucial for performance!).

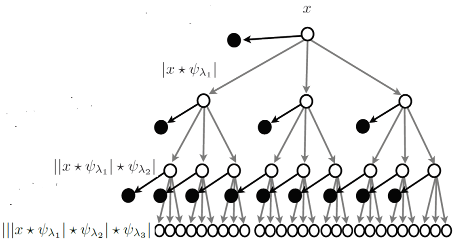
technicolor

Feature extraction

- Signals depend on impact velocity ☺ *and* relative position ☹
- Sound and seismic signals represent different physical quantities.
- To cope, we rely on a "CNN-like" scattering trnsform (16).

Feature extraction up to the *order* p:

0: $S_0(x) = \phi_T * x$,
1: $S_1^{\lambda_1}(x) = \phi_T * |\psi_{\lambda_1} * x|$,
2: $S_2^{\lambda_1,\lambda_2}(x) = \phi_T * |\psi_{\lambda_2} * |\psi_{\lambda_1} * x||$,
...
p: $S_p^{\lambda_1,\ldots\lambda_p}(x) =$
$\phi_T * |\psi_p * \ldots |\psi_{\lambda_2} * |\psi_{\lambda_1} * x|| \ldots |$.



$\phi_T := \phi_T(t)$ - a lowpass ($2\pi/T$) filter, $\psi_\lambda := \psi_\lambda(t)$ - a complex wavelet at scale λ

Rule of thumb

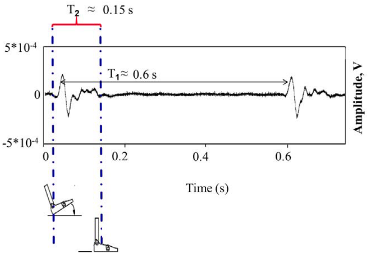
1. Computational cost increases with T ("time-invariance").
2. $T \propto$ duration of a classified event (crucial for performance!).

Feature extraction

Competing requirements for T:

1. Short ($T \sim 0.15$s): characterizes only the footfall event, requires $\text{p} = 1$.
2. Large ($T \sim 1.22$s): captures also the temporal dynamics, but violates LSA and increases cost.

Can we avoid this tradeoff?



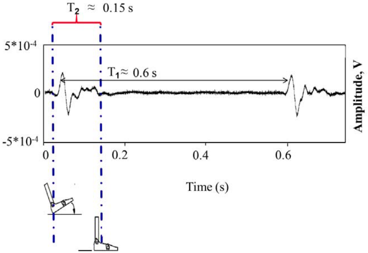
Visual comparison - two $\text{p} = 1$ scattering matrices (audio):

Feature extraction

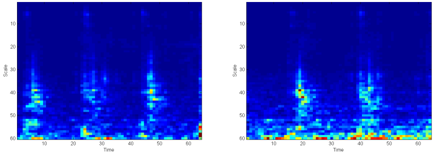
Competing requirements for T:

1. Short ($T \sim 0.15$s): characterizes only the footfall event, requires $\mathsf{p} = 1$.
2. Large ($T \sim 1.22$s): captures also the temporal dynamics, but violates LSA and increases cost.

Can we avoid this tradeoff?



Visual comparison - two $\mathsf{p} = 1$ scattering matrices (audio):



Invariances mostly due to a global temporal offset!

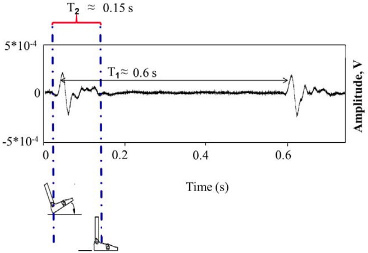
technicolor

Feature extraction

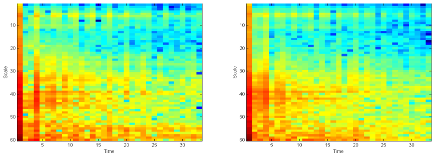
Competing requirements for T:

1. Short ($T \sim 0.15$s): characterizes only the footfall event, requires $\mathbf{p} = 1$.
2. Large ($T \sim 1.22$s): captures also the temporal dynamics, but violates LSA and increases cost.

Can we avoid this tradeoff?



Visual comparison - two $\mathbf{p} = 1$ scattering matrices (audio):



Remedy - compute Fourier modulus across rows (time).

Robust scattering features: normalized scattering

What about feature dependency on $\vec{r}$?

Normalized scattering

Under certain assumptions on $h := h(\mathrm{t})$, it can be shown:

$$S_{\mathrm{p}}^{\lambda_1,\dots\lambda_{\mathrm{p}}}(h * x) \approx |\hat{h}(\lambda_1)| S_{\mathrm{p}}^{\lambda_1,\dots\lambda_{\mathrm{p}}}(x),$$

then:

$$\tilde{S}_{\mathrm{p}}^{\lambda_1,\dots\lambda_{\mathrm{p}}}(h * x) := \frac{S_{\mathrm{p}}^{\lambda_1,\dots\lambda_{\mathrm{p}}}(h * x)}{S_{\mathrm{p}}^{\lambda_1,\dots\lambda_{\mathrm{p}-1}}(h * x)} \approx \tilde{S}_{\mathrm{p}}^{\lambda_1,\dots\lambda_{\mathrm{p}}}(x).$$

Consequence: if LSA holds, normalized scattering features depend *only* on $v(\mathrm{t})$!

A cheap channel normalization technique - “scattering CMS”.

technicolor

Robust scattering features: normalized scattering

What about feature dependency on $\vec{r}$?

Normalized scattering

Under certain assumptions on $h := h(\mathrm{t})$, it can be shown:

$$S_{\mathrm{p}}^{\lambda_1,\ldots\lambda_{\mathrm{p}}}(h * x) \approx |\hat{h}(\lambda_1)| S_{\mathrm{p}}^{\lambda_1,\ldots\lambda_{\mathrm{p}}}(x),$$

then:

$$\tilde{S}_{\mathrm{p}}^{\lambda_1,\ldots\lambda_{\mathrm{p}}}(h * x) := \frac{S_{\mathrm{p}}^{\lambda_1,\ldots\lambda_{\mathrm{p}}}(h * x)}{S_{\mathrm{p}}^{\lambda_1,\ldots\lambda_{\mathrm{p}-1}}(h * x)} \approx \tilde{S}_{\mathrm{p}}^{\lambda_1,\ldots\lambda_{\mathrm{p}}}(x).$$

Consequence: if LSA holds, normalized scattering features depend *only* on $v(\mathrm{t})$!

A cheap channel normalization technique - "scattering CMS".

technicolor

Robust scattering features: normalized scattering

What about feature dependency on $\vec{r}$?

Normalized scattering

Under certain assumptions on $h := h(\mathrm{t})$, it can be shown:

$$S_{\mathrm{p}}^{\lambda_1,\ldots\lambda_{\mathrm{p}}}(h * x) \approx |\hat{h}(\lambda_1)| S_{\mathrm{p}}^{\lambda_1,\ldots\lambda_{\mathrm{p}}}(x),$$

then:

$$\tilde{S}_{\mathrm{p}}^{\lambda_1,\ldots\lambda_{\mathrm{p}}}(h * x) := \frac{S_{\mathrm{p}}^{\lambda_1,\ldots\lambda_{\mathrm{p}}}(h * x)}{S_{\mathrm{p}}^{\lambda_1,\ldots\lambda_{\mathrm{p}-1}}(h * x)} \approx \tilde{S}_{\mathrm{p}}^{\lambda_1,\ldots\lambda_{\mathrm{p}}}(x).$$

Consequence: if LSA holds, normalized scattering features depend *only* on $v(\mathrm{t})$!

A cheap channel normalization technique - “scattering CMS”.

technicolor

Feature fusion

What about fusion?

- Recall that $\hat{x}_{\text{a}}$ and $\hat{x}_{\text{g}}$ have (approx) complementary frequency range.
- Hence, $\tilde{S}_1^{\lambda_1}(x_{\text{a}}) > 0$ and $\tilde{S}_1^{\lambda_1}(x_{\text{g}}) > 0$ should be complementary as well.
- Due to channel normalization, $\tilde{S}_1^{\lambda_1}(x_{\text{a}})$ and $\tilde{S}_1^{\lambda_1}(x_{\text{g}})$ "live" in the same feature space, we can simply sum them up[1]:

$$\tilde{S}_{\text{fused}}^{\lambda_1} = \alpha_{\text{a}} \tilde{S}_1^{\lambda_1}(x_{\text{a}}) + \alpha_{\text{g}} \tilde{S}_1^{\lambda_1}(x_{\text{g}})$$

$\tilde{S}_0(x_{\text{a}})$ and $\tilde{S}_0(x_{\text{g}})$ are concatenated to $\tilde{S}_{\text{fused}}^{\lambda_1}$.

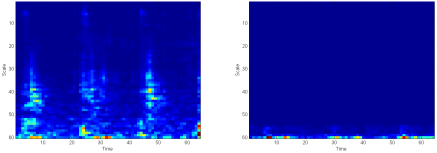
[1] $\alpha_{.}$ is a normalization constant

technicolor

Feature fusion

What about fusion?

- Recall that $\hat{x}_{\text{a}}$ and $\hat{x}_{\text{g}}$ have (approx) complementary frequency range.
- Hence, $\tilde{S}_1^{\lambda_1}(x_{\text{a}}) > 0$ and $\tilde{S}_1^{\lambda_1}(x_{\text{g}}) > 0$ should be complementary as well.



- Due to channel normalization, $\tilde{S}_1^{\lambda_1}(x_{\text{a}})$ and $\tilde{S}_1^{\lambda_1}(x_{\text{g}})$ "live" in the same feature space, we can simply sum them up[1]:

$$\tilde{S}_{\text{fused}}^{\lambda_1} = \alpha_{\text{a}}\tilde{S}_1^{\lambda_1}(x_{\text{a}}) + \alpha_{\text{g}}\tilde{S}_1^{\lambda_1}(x_{\text{g}})$$

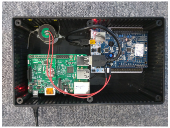
$\tilde{S}_0(x_{\text{a}})$ and $\tilde{S}_0(x_{\text{g}})$ are concatenated to $\tilde{S}_{\text{fused}}^{\lambda_1}$.

[1] $\alpha_{.}$ is a normalization constant

Experiments

Experimental setup (17):

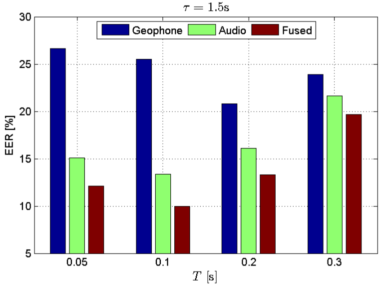
- Data collected internally, on a prototype dual sensor setup.
- 12 participants (8m and 4f), up to two types of shoes per person.
- (Low noise) recordings in a carpet-covered room, on 3 different days[2].
- 6 persons randomly chosen for training the UBM.
- From the remaining, randomly chosen 3 targets and 3 unknowns.
- Hyperparameters: τ, T, N (the number of retained coefficients after PCA).



[2] To avoid environmental effects: 2 days for training, 3rd day for evaluation.

Results

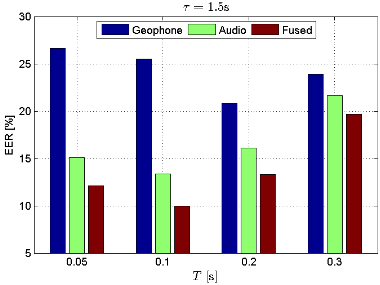
- Performance metric: *Equal Error Rate (EER)*, lower is better.
- Median results for the best-performing N, after 100 random partitions.

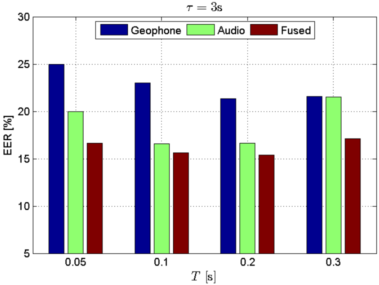


- "Optimal" hyperparameters agree with predictions:
  1. T on the order of the footfall impact duration.
  2. Larger τ degrades performance (violates LSA).
  3. "Richer" representations (*i.e.* audio and fused) favor larger N.

Results

- Performance metric: *Equal Error Rate (EER)*, lower is better.
- Median results for the best-performing N, after 100 random partitions.

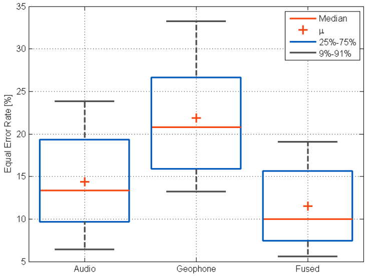




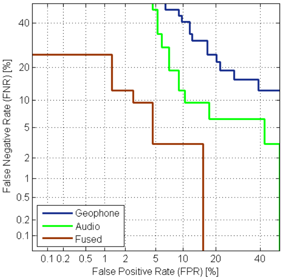
- "Optimal" hyperparameters agree with predictions:
  1. T on the order of the footfall impact duration.
  2. Larger τ degrades performance (violates LSA).
  3. "Richer" representations (*i.e.* audio and fused) favor larger N.

technicolor

Results



Best setting for each modality



Typical DET curves

Classification with fused features:

- exhibits the smallest variance,
- is the most robust wrt parameterization.

technicolor

Summary

Bimodal gait-based identification wrap-up:

- Confirmed identification by both sound and seismic observations.
- Performance gradation: fused $>$ sound $>$ seismic.
- Further research directions:
  - Recognition in noisy conditions and using cheap MEMS sensors.
  - "Walker diarization"?
  - Relevance of the shoe type, gender and/or environment.
  - A better way to fuse / extract features (new datasets), etc.

technicolor

technicolor

References I

(1) L. Wang, T. Tan, H. Ning, and W. Hu, "Silhouette analysis-based gait recognition for human identification," *IEEE transactions on pattern analysis and machine intelligence*, vol. 25, no. 12, pp. 1505–1518, 2003.

(2) C. Chen, J. Liang, H. Zhao, H. Hu, and J. Tian, "Frame difference energy image for gait recognition with incomplete silhouettes," *Pattern Recognition Letters*, vol. 30, no. 11, pp. 977–984, 2009.

(3) D. Bales, P. A. Tarazaga, M. Kasarda, D. Batra, A. Woolard, J. D. Poston, and V. S. Malladi, "Gender classification of walkers via underfloor accelerometer measurements," *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 1259–1266, 2016.

(4) M. Köhle and D. Merkl, "Identification of gait patterns with self-organizing maps based on ground reaction force." in *ESANN*, vol. 96, 1996, pp. 24–26.

(5) J. Mantyjarvi, M. Lindholm, E. Vildjiounaite, S.-M. Makela, and H. Ailisto, "Identifying users of portable devices from gait pattern with accelerometers," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'05), 2005.*, vol. 2. IEEE, 2005, pp. ii–973.

(6) D. Gafurov, K. Helkala, and T. Søndrol, "Biometric gait authentication using accelerometer sensor." *JCP*, vol. 1, no. 7, pp. 51–59, 2006.

(7) W. Wang, A. X. Liu, and M. Shahzad, "Gait recognition using wifi signals," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing.* ACM, 2016, pp. 363–373.

(8) M. Hofmann, J. Geiger, S. Bachmann, B. Schuller, and G. Rigoll, "The tum gait from audio, image and depth (gaid) database: Multimodal recognition of subjects and traits," *Journal of Visual Communication and Image Representation*, vol. 25, no. 1, pp. 195–206, 2014.

(9) A. Itai and H. Yasukawa, "Footstep classification using simple speech recognition technique," in *Circuits and Systems, 2008. ISCAS 2008. IEEE International Symposium on.* IEEE, 2008, pp. 3234–3237.

(10) J. T. Geiger, M. Hofmann, B. Schuller, and G. Rigoll, "Gait-based person identification by spectral, cepstral and energy-related audio features," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2013.* IEEE, 2013, pp. 458–462.

(11) J. T. Geiger, M. Kneißl, B. W. Schuller, and G. Rigoll, "Acoustic gait-based person identification using hidden markov models," in *Proceedings of the 2014 Workshop on Mapping Personality Traits Challenge and Workshop.* ACM, 2014, pp. 25–30.

(12) S. Pan, N. Wang, Y. Qian, I. Velibeyoglu, H. Y. Noh, and P. Zhang, "Indoor person identification through footstep induced structural vibration," in *Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications.* ACM, 2015, pp. 81–86.

technicolor

References II

(13) D. A. Reynolds and W. M. Campbell, "Text-independent speaker recognition," in *Springer Handbook of Speech Processing.* Springer, 2008, pp. 763–782.

(14) J. Andén and S. Mallat, "Deep scattering spectrum," *IEEE Transactions on Signal Processing*, vol. 62, no. 16, pp. 4114–4128, 2014.

(15) A. Ekimov and J. Sabatier, "Rhythm analysis of orthogonal signals from human walking," *The Journal of the Acoustical Society of America*, vol. 129, no. 3, pp. 1306–1314, 2011.

(16) J. Bruna and S. Mallat, "Invariant scattering convolution networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 8, pp. 1872–1886, 2013.

(17) S. Kitić, G. Puy, P. Pérez, and P. Gilberton, "Scattering features for multimodal gait recognition," in *IEEE Global Conference on Signal and Information Processing (GlobalSIP), 2017.*, 2017.