UNIVERSITÀ
POLITECNICA
DELLE MARCHE

Facoltà di Ingegneria

Diego Droghini

Ambient Intelligence: Computational Audio Processing For Human Fall Detection

Dottorato in Ingegneria dell'Informazione
Progetto Eureka

13-14 Marzo 2019, www.univpm.it

Outline

UNIVERSITÀ
POLITECNICA
DELLE MARCHE

- Introduction
  -Fall Detection System and Challange
  -Motivations and Contributions
- Dataset
  -A3FALL Dataset
  -FAS
- Supervised Approaches
  -SVM based apprach
- Unsupervised Approach
  -End-To-End CNN-AE
- Weakly-supervised Approach
  -OCSVM + User-Aided Template Matching
  -Few-shot Siamese Neural Networks
  -Robust Metric Learning with Siamese Autoencoders
- Conclusions
- References
- Publications list

UNIVERSITÀ POLITECNICA DELLE MARCHE

Where we are going: Ambient Intelligence (AmI)

Capabilities

- integrated into the environment
- context-aware
- tailored to the user need
- able to adapt actions in new scenarios
- minimal user intervention

A digital environment that proactively, but sensibly, supports people in their daily lives.

Requirement

- Sensor interaction and interoperability
- Pervasive Ubiquitous Computing
- Artificial Intelligence

Index

Human fall: a real problem for society

UNIVERSITÀ POLITECNICA DELLE MARCHE

"And why do we fall, Bruce? So we can learn to pick ourselves up" cit. Alfred - Batman Begins

- 62% of injury-related hospitalizations for the people over 65 years are the result of a fall [1]
- main cause of death due to accidents for people over 65 [2]
- can lead to psychophysical repercussions on people [3]

We can introduce an FCS system in an AmI to:

- Monitoring of the elderly or people who live alone
- Assistance time reduction

Fall Classification System Challenge [4]

UNIVERSITÀ POLITECNICA DELLE MARCHE

- Try to collect sufficient human fall data: supervised or analytical methods
  - difficulty in retrieving examples that represent human falls
- Deal with no human fall data: Novelty detection approach
  - good description of "normality"

Related Work

- Wearable
  - **Accelerometers**
  - Gyroscopes
- Environmental
  - **Vision System**
  - Audio microphone
  - Radar doppler

Issues

- no standard datasets;
- no standard methodology
- difficult comparison with other methods;
- no audio dataset available
- scarcity of real human falls

Index

UNIVERSITÀ
POLITECNICA
DELLE MARCHE

Motivations and Contributions

UNIVERSITÀ POLITECNICA DELLE MARCHE

Analytical methods:

- exploiting some a priori knowledge;
- manual tuning of the hyperparameters;
- applying a threshold on aquired signals or features;

They can hardly perform when the operating conditions and the subjects are variable.

Machine learning methods:

- learns from the data;
- can be eforced by exploiting some a priori knowledge;
- no need of manual tuning of the hyperparameters;

They need more data!

Contributions

- audio based fall detection is a reliable alternative;
- proposal and evaluation of the FD task dedicated FAS sensor;
- proposal of the acoustic A3FALL dataset to the community.

Index

The fall events dataset: A3Fall

UNIVERSITÀ POLITECNICA DELLE MARCHE

Recording Rooms

- R0 - rectangular room particularly suitable for the propagation of acoustic waves through the floor;
- R1 - university auditorium room in which the flooring is composed of fitted carpet;
- R2 - a recording studio
  - sensors placed in the live room;
  - audio events performed in the control room.

Recording set-up

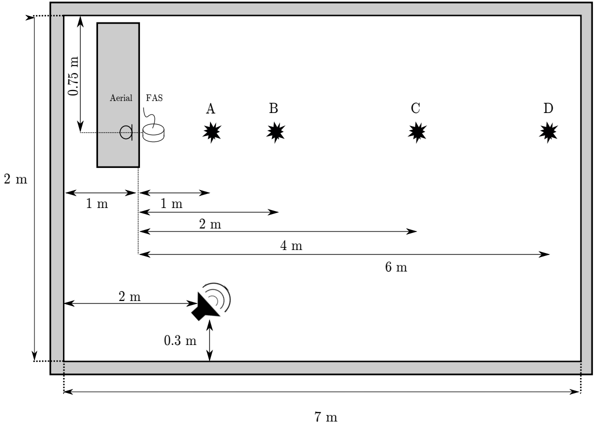
Sampling

- Sample Rate: 44.1 kHz
- Bit-depth: 24 bit

Used sensors:

- Mic. array (3)[1]
- **FAS v1**[1]

The recording room



[1] AKG C400 BLT.

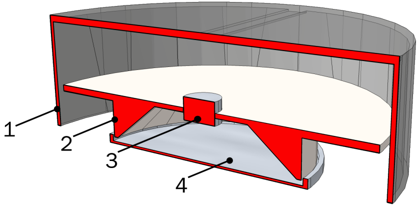
Index

UNIVERSITÀ
POLITECNICA
DELLE MARCHE

Floor Acoustic Sensor[1]

UNIVERSITÀ POLITECNICA DELLE MARCHE

1. The outer container
2. The inner container
3. The microphone slot
4. The membrane touching the floor



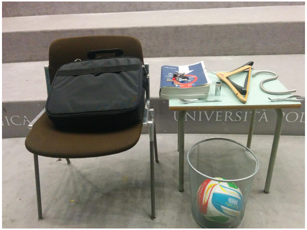
Advantage

- Easily integrated into the environment
- More sensitive to signals related to falls

3D printed prototype

Description

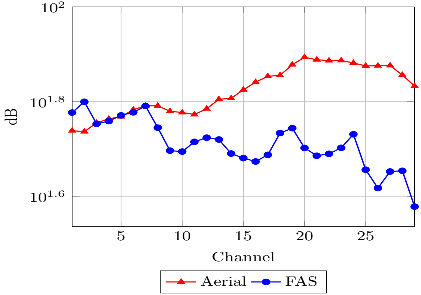
Class	R0	R1	R2
	Nr. of occurrences		
Basket	64	40	40
Fork	64	40	40
Ball	64	40	40
Book	64	40	40
Bag	64	30	40
Chair	96	40	40
Table	0	40	40
Guitar Slide	0	40	40
Nipper	0	40	40
Keys	0	40	40
Hook	0	40	40
Coat Hook	0	40	40
Manikin Doll	44	0	0
Human Fall	0	40	40
	Total length (s)		
Background	2530	9055	5550



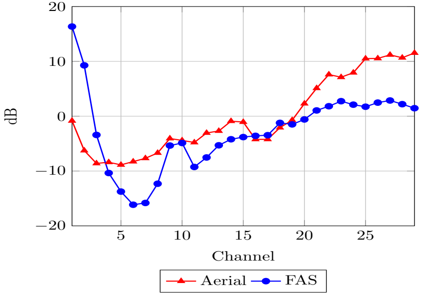
Signals analysis

UNIVERSITÀ
POLITECNICA
DELLE MARCHE

Average value of the mel channels



Average value of the SNR for each mel channel



Remark

- The "valley" in the curves is due to the pitch of the music signal
- The FAS is better at low frequencies

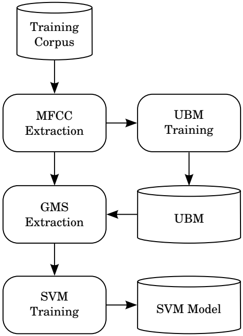
Index

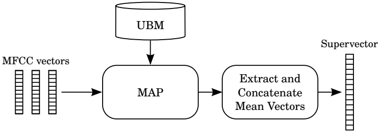
UNIVERSITÀ
POLITECNICA
DELLE MARCHE

Supervised Approaches: SVM based

UNIVERSITÀ POLITECNICA DELLE MARCHE

- Completely supervised scenario
- GMM to model a UBM with EM algorithm
- MFCC features to Supervectors (MAP)
- SVM classificator



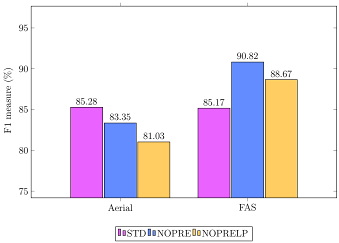


Scenarios

UNIVERSITÀ POLITECNICA DELLE MARCHE

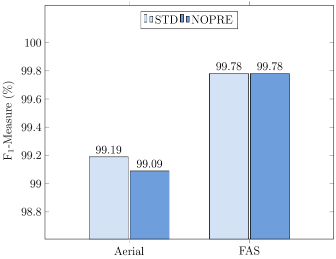
Multi-class

Trained to discriminate the type of the fallen object



Bi-class

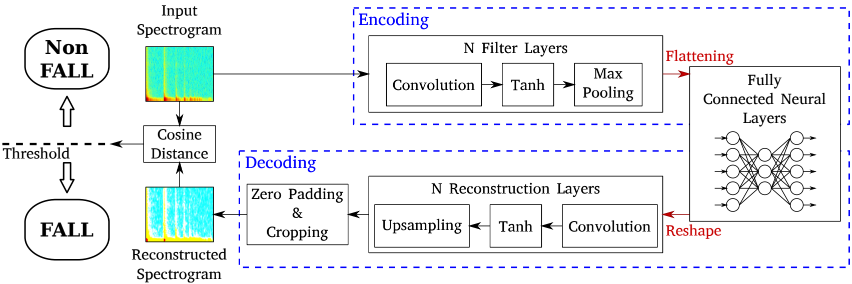
Trained to discriminate fall from non-fall



Index

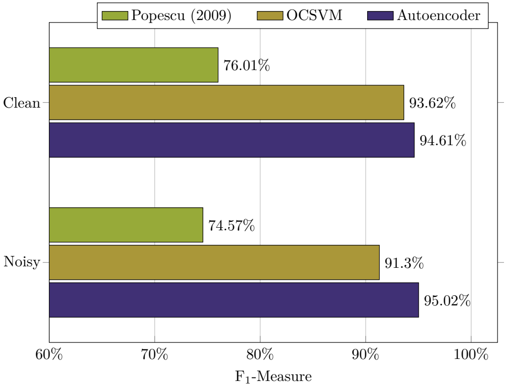
Unsupervised Approach

End-To-End Convolutional Autoencoder



Results

- Trained with background sounds only
- Tested with:
  - background
  - human fall

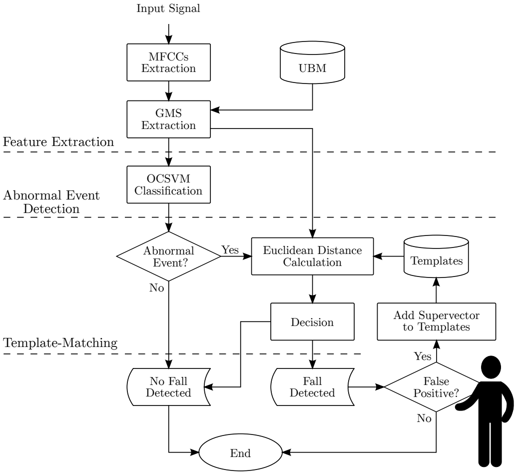


Index

UNIVERSITÀ POLITECNICA DELLE MARCHE

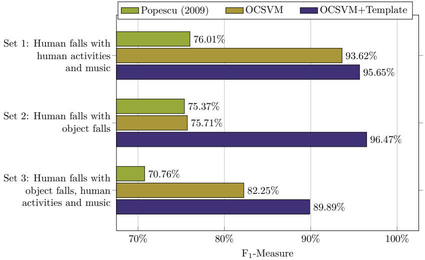
OCSVM user-aided

- 1° Stage: OCSVM
  - trained with non-fall events (background noises)
- 2° Stage: template matching
  - based on Euclidean Distance
- Final decision
  - user-aided: false positive report

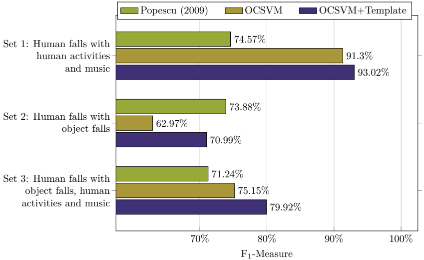


Results

Clean condition

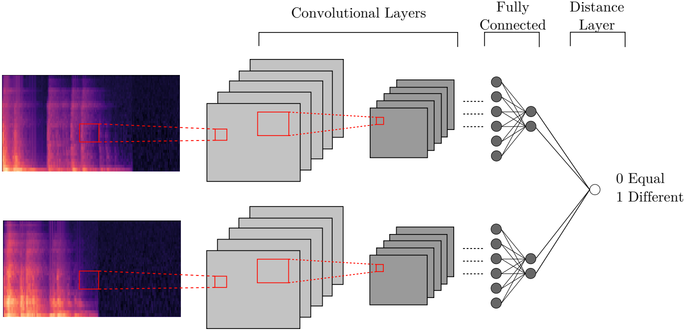


Noisy condition



Index

Few-shot Siamese Neural Networks[1]



$$Loss = (1 - Y)\frac{1}{2}(E_w)^2 + (Y)\frac{1}{2}\{(max(0, m - E_w)\}^2$$

[1] Best Presentation @ International Conference on Pattern Recognition and Artificial Intelligence (PRAI2018) Kean University, Union, NJ, USA

Training Phase

Consider X_1, X_2 as a pair of two input samples.

Euclidean Distance

$$E_w = \|S_e(X_1) - S_e(X_2)\| \,. \tag{1}$$

with $S_e(X1)$ and $S_e(X2)$ the mappings performed by the network

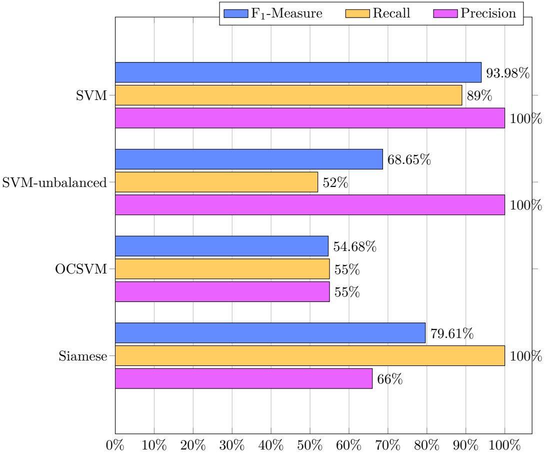
The label for training a SNN are built as follows:

- Positive examples labelled as 0: pair of input samples belonging to the same fall event class
- Negative examples labelled as 1: pair of input samples not belonging to the same fall event class

Results

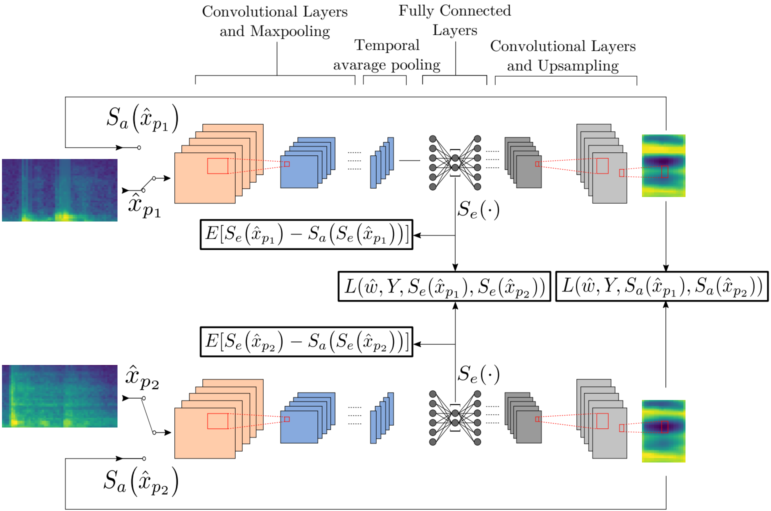
UNIVERSITÀ POLITECNICA DELLE MARCHE

- Trained:
  - objects fall
  - 1 human fall
- Validation:
  - objects fall
  - few human fall
- Tested with:
  - objects fall
  - human fall



Index

Robust Metric Learning with SCAE

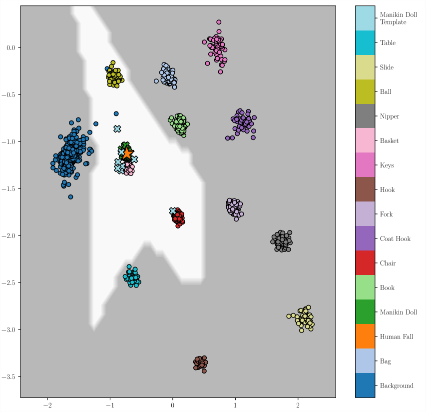


Training procedure

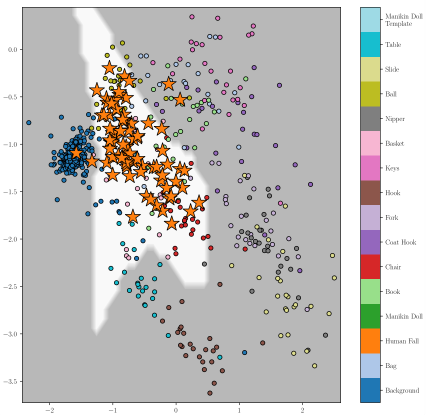
- contrastive loss of latent spaces
- contrastive loss of reconstructions
- recostruction loss (mse) of latent spaces
- to exploit the few human fall samples available a smart selection of the train pairs has been done:
  - neither the real falls nor the simulated falls were coupled with the falls of the objects
  - simulated and real falls were coupled together
- induces the network to make a transformation in the latent layer
- simulated human fall can be used as real human fall template in the final classifier
- KNN classifier trained with extracted feat. from the **Encoder**

Latent space with 2 neurons

Train set

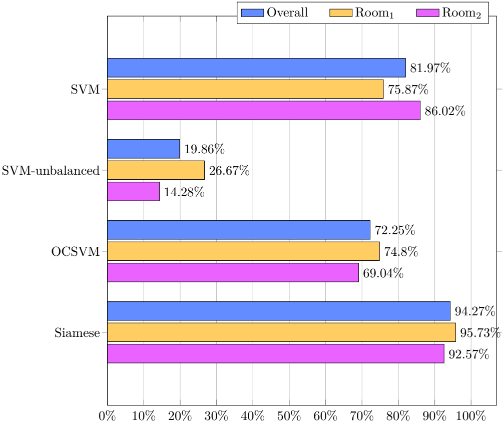


Test set



Results

- Trained (3 room):
  - objects falls
  - backgrounds
  - manikin doll falls
  - 1 human fall per room
- Testset composed of:
  - objects fall
  - human fall
  - backgrounds



Conclusions

- FAS sensor proposed and evaluated
- A3FALL dataset with real human falls
  - real human falls
  - different floor types
  - different rooms
- different scenarios explored from a data availability prospective
  - supervised
  - unsupervised
  - weakly-supervised

Future works

- data fusion (different type of sensors)
- approach fusion
- exploit the AmI ecosystem (different algorithms)

Thanks for your attention

References (1)

R. J. Gurley, N. Lum, M. Sande, B. Lo, and M. H. Katz,
"Persons found in their homes helpless or dead,"
New England Journal of Medicine, vol. 334, no. 26, pp. 1710–1716, 1996.

M. Mubashir, L. Shao, and L. Seed,
"A survey on fall detection: Principles and approaches,"
Neurocomputing, vol. 100, pp. 144–152, 2013.

S. Abbate, M. Avvenuti, P. Corsini, J. Light, and A. Vecchio,
"Monitoring of human movements for fall detection and activities recognition in elderly care using wireless sensor network: a survey,"
in *Wireless Sensor Networks: Application-Centric Design*. InTech, 2010.

S. S. Khan and J. Hoey,
"Review of fall detection techniques: A data availability perspective,"
Medical engineering and physics, vol. 39, pp. 12–22, 2017.

E. Principi, D. Droghini, S. Squartini, P. Olivetti, and F. Piazza,
"Acoustic cues from the floor: a new approach for fall classification,"
Expert Systems with Applications, vol. 60, pp. 51–61, 2016.

References (2)

UNIVERSITÀ POLITECNICA DELLE MARCHE

S. Chopra, R. Hadsell, and Y. LeCun,
"Learning a similarity metric discriminatively, with application to face verification,"
in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. IEEE, 2005, vol. 1, pp. 539–546.

D. Droghini, E. Principi, S. Squartini, P. Olivetti, and P. F.,
"Human fall detection by using an innovative floor acoustic sensor,"
in *Proc. of WIRN*. Vietri sul Mare, Italy, May, 18-20 2016,
(8 pages), to appear.

J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, and R. Shah,
"Signature verification using a" siamese" time delay neural network,"
in *Advances in Neural Information Processing Systems*, 1994, pp. 737–744.

G. Koch, R. Zemel, and R. Salakhutdinov,
"Siamese neural networks for one-shot image recognition,"
in *ICML Deep Learning Workshop*, 2015, vol. 2.

References (3)

UNIVERSITÀ POLITECNICA DELLE MARCHE

D. Droghini, D. Ferretti, E. Principi, S. Squartini, and F. Piazza,
"A combined one-class svm and template matching approach for user-aided human fall detection by means of floor acoustic features,"
Computational Intelligence and Neuroscience, vol. 2017, 2017,
Article ID 1512670.

Publications List (1)

UNIVERSITÀ POLITECNICA DELLE MARCHE

International Journal: *[3 articles, 2 first author]*

- [1] Principi, E., Droghini, D., Squartini, S., Olivetti, P., Piazza, F.:, "'Acoustic cues from the floor: a new approach for fall classification," *Expert Systems with Applications*, 2016.
- [2] Droghini, D., Ferretti, D., Principi, E., Squartini, S., Piazza, F.:, "A combined one-class svm and template matching approach for user-aided human fall detection by means of floor acoustic features," *Computational Intelligence and Neuroscience*, 2017.

International Journal (submitted): *[1 article, 1 first author]*

- [1] Droghini, D., Principi, E., Squartini, S., Gabrielli, L., P., Piazza., "Audio Metric Learning by using Siamese Autoencoders for One-Shot Human Fall Detection," *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2018, submitted.

International Conference: *[14 articles, 5 first author]*

Publications List (2)

UNIVERSITÀ POLITECNICA DELLE MARCHE

[1] Droghini, D., Principi, E., Squartini, S., Olivetti, P., Piazza., F., "Human fall detection by using an innovative floor acoustic sensor," in *Proc. of WIRN*, Vietri sul Mare, Italy, May, 2016.

[2] Gabrielli, L., Cella, E., Vesperini, F., Droghini, D., Principi, E., Squartini, S., "Deep Learning for Timbre Modification and Transfer: An Evaluation Study," in *Proc. of AES*, Milan, Italy, May, 2018.

[3] Droghini, F., Vesperini, F., Principi, E., Squartini, S., Piazza., F., "Few-shot Siamese Neural Networks employing Audio features for Human-Fall Detection", in *Proc. of PRAI*,Kean University, Union, NJ, USA, Aug 2018.

[4] Vesperini, F., Droghini, D., Principi, E., Gabrielli, L., Squartini, S., "Hierarchic ConvNets Framework for Rare Sound Event Detection", in *Proc. of EUSIPCO*, Rome, Italy, Sept. 2018.

Others:

Publications List (3)

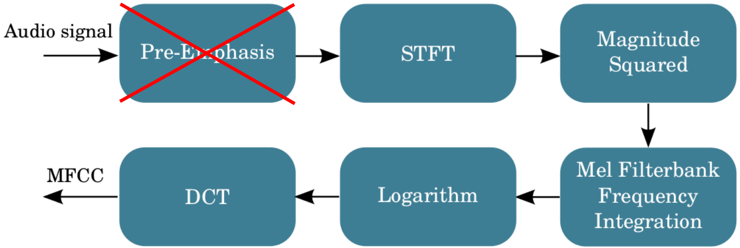
F. Vesperini, D. Droghini, D. Ferretti, E. Principi, L. Gabrielli, S. Squartini, and F. Piazza,
"A hierarchic multi-scaled approach for rare sound event detection",
Ancona, Italy, 2017, DCASE Tech. Report. Copyright-free.

L. Gabrielli, F. Vesperini, D. Droghini, and S. Squartini,
"Rima Glottidis: Experimenting generative raw audio synthesis for a sound installation,"
in *XXII Colloquium of Musical Informatics*, Udine, Italy, 20-23 Nov. 2018.

MFCC pipelines used

UNIVERSITÀ POLITECNICA DELLE MARCHE

- STD: MFCC calculated as described above
- NOPRE: no pre-emphasis filter
  - to reduce the high frequencies emphasis



SVM based methods optimization

UNIVERSITÀ POLITECNICA DELLE MARCHE

SVM and USVM Gridsearch

GMM

- number of gaussians: $[2^0, 2^1, ..., 2^6]$

SVM

- C : $[2^{-5}, 2^{-3}, ..., 2^{15}]$
- γ : $[2^{-15}, 2^{-13}, ..., 2^{3}]$

OCSVM Gridsearch

GMM

- number of gaussians: $[2^0, 2^1, ..., 2^6]$

SVM

- ν : $2^{-5}, 2^{-3}, ..., 2^{15}]$
- γ : $[2^{-15}, 2^{-13}, ..., 2^{3}]$

For all the methods the $F_1 Measure$ ware used for the optimization.

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

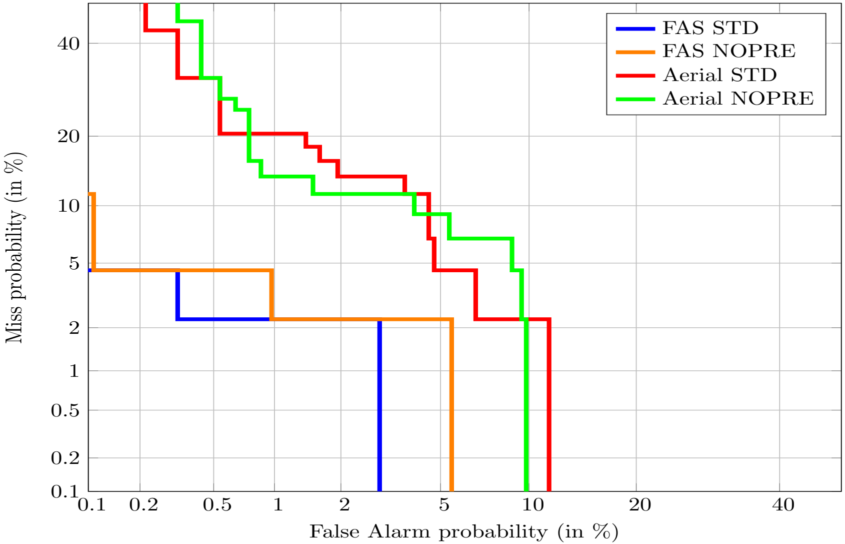
Acoustic Scenarios for SVM based approaches

UNIVERSITÀ POLITECNICA DELLE MARCHE

The experiments have been conducted with a 4-fold cross-validation strategy and a three-way data split in three different operating conditions:

- matched, where the training, validation and test sets share the same acoustic condition, i.e., clean or noisy;
- mismatched, where the training set is composed of clean signals while the validation and test sets are composed of noisy signals;
- multicondition, where the training, validation and test sets contain both clean and noisy data. In this case the sets have been divided so that they contain 1/3 of clean data and 2/3 of noisy data.

Binary-SVM Multicondition DET plot



Optimization of the End-to-End Autoencoder

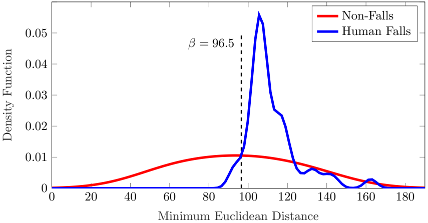
Table: Hyper-parameters optimized in the random-search phase, and their range.

Parameter	Range	Distribution	Parameter	Range	Distribution
Cnn layer Nr.	[1-3]	uniform	Batch size	[10%-25%]	log-uniform
Kernel shape	[3x3-8x8]	uniform	Max pool shape	[1x1-5x5]	uniform
Kernel Nr.	[4-64]	log-uniform	Max Pool	All-Only end	uniform
MLP layers Nr.	[1-3]	uniform	Dropout	[Yes-No]	uniform
MLP layers dim.	[128-4096]	log-unifom	Drop rate	[0.5-0.6]	normal
Stride	[1x1-3x3]	uniform	Learning rate	[10^{-4}-10^{-2}]	log-unifom

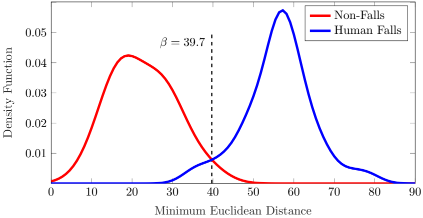
Template-matching decision threshold in clean condition

UNIVERSITÀ POLITECNICA DELLE MARCHE

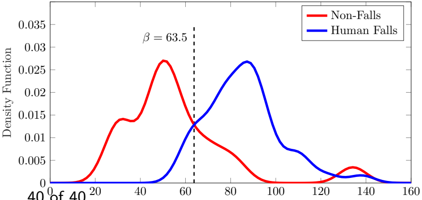
Set 1



Set 2

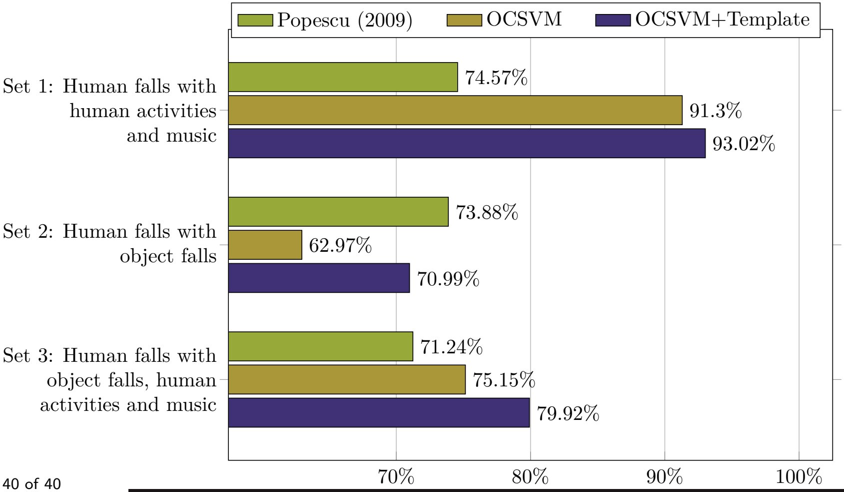


Set 3



The choice of β has been performed by calculating the minimum Euclidean distance between each fall and non-fall event in the validation set and the set of templates

OCSVM Template Matching
Results in Noisy condition



Preliminary results for SCAE approach

Table: Preliminary result for different pairs generation strategies.

Technique	**Result in R1**	**Result in R2**	**Overall**
$\mathcal{P}$-$\mathcal{N}$-PAIRS	55.17%	64.74%	60.13%
$\mathcal{N}$-PAIRS	76.20%	67.53%	72.05%
NO-PAIRS	91.71%	89.88%	90.97%
$\mathcal{P}$-PAIRS	92.54%	92.54%	**92.54%**

Hyper-parameters optimized for SCAE approach

UNIVERSITÀ
POLITECNICA
DELLE MARCHE

Table: Hyper-parameters optimized in the random-search phase and their range.

Parameter	Range	Distribution
Cnn layer Nr.	[1-3]	Uniform
Kernel shape	[1x1-8x8]	Uniform
Kernel Nr.	[1-32]	Uniform
MLP layers Nr.	[1-2]	Uniform
MLP layers dim.	[1-4096]%	Log-uniform
Max pool shape	[0x0-3x3]	Uniform
Drop rate	[0-0.2]%	Uniform