Aggression Detection in Urban Environments based on Audio Analysis

By Edward Fleri Soler | Faculty of ICT | edward.fleri.14@um.edu.mt

Supervised by Dr. George Azzopardi

UNIVERSITY OF MALTA
L-Universita ta' Malta

Contribution

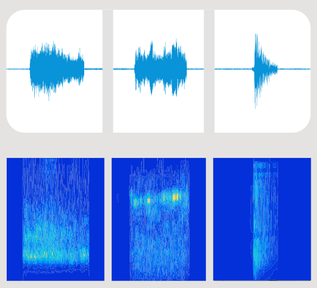
- Aids in the fight against crime
- Reduces reliance on personnel
- Dedicated to urban environments
- Processing and logging of court evidence

Motivation

- Cheaper alternative to visual surveillance
- Certain events only produce audio signatures
- Omnidirectionality and greater area of cover
- Less intrusive

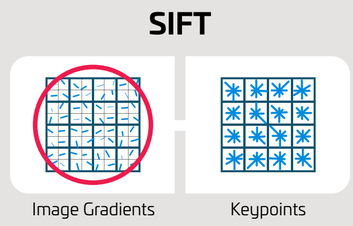
Image Representation

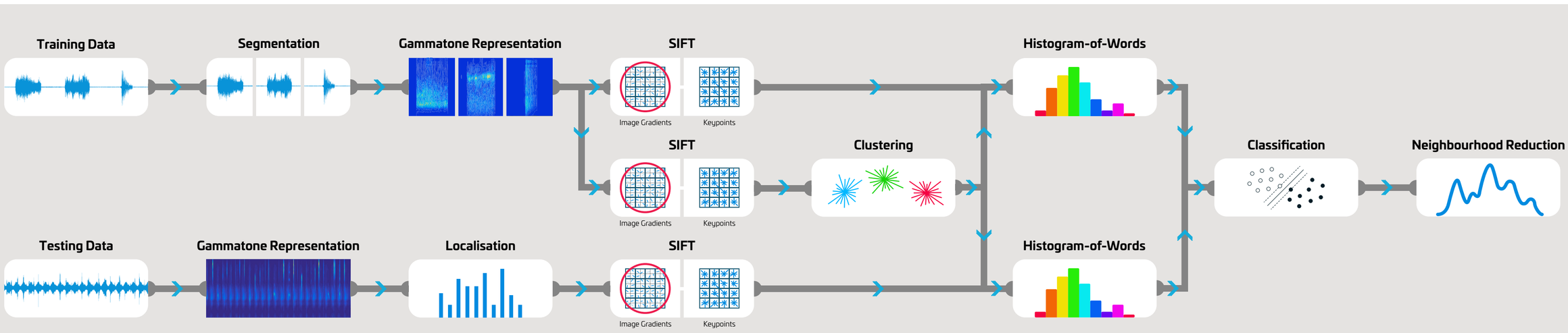
- Option of different representations
- Better illustrates distinct audio events
- Suitable for the detection and classification of events



Feature Extraction

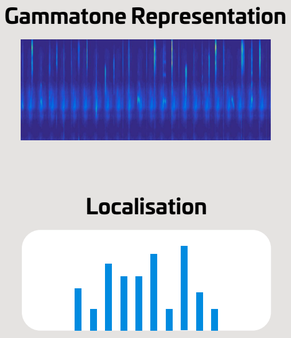
- Employs the Dense SIFT algorithm [1] for feature extraction
- Employs the Bag-of-Words model [2] for high level representation
- Employs spatial pyramids for better description





Localisation

- Application of B-COSFIRE line detection algorithm [3]
- Suspected events located through summed response
- Reduces classification search space



Classification

- Support Vector Machine (SVM) classification
- Neighbourhood reduction to deal with false alarms
- Adapted to deal with varying noise conditions

Results

- Tested on MIVIA Audio Events Dataset
- Capable of handling impulsive and sustained events
- Retains performance in noisy environments

		Predicted Events			
		Background Noise	Breaking Glass	Gunshot	Scream
Actual Events	Background Noise	1814	34	154	46
	Breaking Glass	36	1717	45	2
	Gunshot	55	149	1587	9
	Scream	51	27	104	1595

SNR Level	Breaking Glass	Gunshot	Scream	Total	Foggia et al. Total
5dB	93.00%	83.00%	73.65%	83.26%	81.1%
10dB	96.67%	87.33%	89.19%	91.07%	85%
15dB	95.67%	89.33%	94.26%	93.08%	87%
20dB	96.00%	89.00%	94.26%	93.08%	88.4%
25dB	95.33%	90.33%	93.58%	93.08%	88.7%
30dB	95.67%	90.00%	93.92%	93.19%	90%
Total	95.39%	88.17%	89.81%	91.13%	86.7%

References

[1] David G Lowe. Object recognition from local scale-invariant features. In Computer vision, 1999. The proceedings of the seventh IEEE international conference on, volume 2, pages 1150{1157. Ieee, 1999.

[2] Jun Yang, Yu-Gang Jiang, Alexander G Hauptmann, and Chong-Wah Ngo. Evaluating bag-of-visual-words representations in scene classication. In Proceedings of the international workshop on Workshop on multimedia information retrieval, pages 197{206. ACM, 2007.

[3] George Azzopardi, Nicola Strisciuglio, Mario Vento, and Nicolai Petkov. Trainable cosfire filters for vessel delineation with application to retinal images. Medical image analysis, 19(1):46{57, 2015.