Desarrollo de una plataforma de procesamiento de audio en tiempo real con aplicación a seguridad urbana

Pasantía Laboral en SONDA Uruguay

Maestría en Ingeniería Eléctrica del *Instituto de Ingeniería Eléctrica, Facultad de Ingeniería, Universidad de la República, Uruguay.*

Ing. Juan Braga

1 de julio de 2016

Período:	Diciembre 2015 a Marzo 2016
Docente Responsable:	*Msc. Ing. Guillermo Carbajal*
Responsable SONDA Uruguay:	*Nestor Rossi*, Gerente de Proyectos
Equipo de SONDA Uruguay:	*Ing. Florencia Lanzaro*
	Msc. Ing. Guillermo Carbajal

Resumen

El presente informe tiene como objetivo la documentación del trabajo realizado en SONDA Uruguay en el período de Diciembre 2015 a Marzo 2016 para su aprobación en créditos de la Maestría en Ingeniería Eléctrica de la UdelaR. Se trató de un estudio del estado del arte sobre la aplicación de analíticas a seguridad urbana y el desarrollo de una plataforma de procesamiento de audio en tiempo real para prueba de concepto en condiciones de laboratorio de dos analíticas de audio en particular,

1. Introducción

El monitoreo en tiempo real de las actividades humanas para seguridad urbana se ha vuelto masivo en los últimos años a nivel mundial. La cantidad de sensores distribuidos por las ciudades ha crecido enormemente. Es entonces en los centros de monitoreo, donde confluyen las señales desde variados puntos de la ciudad, que surge la necesidad de automatizar los procesos de visualización y control para hacer la operativa más eficiente (Crocco et al., 2016). El desarrollo de analíticas en tiempo real para detección automática de patrones de interés, ha captado la atención de la comunidad científica así como a empresas con fines

comerciales. En este escenario es donde la empresa SONDA Uruguay motiva un proyecto de desarrollo de software de analíticas de audio y video, para el monitoreo automático en tiempo real de los espacios urbanos.

En el marco del proyecto de la empresa anteriormente mencionado, se realizó una prueba de concepto sobre la implementación de analíticas de audio en tiempo real para seguridad urbana. Con este fin se realizó un estudio del estado del arte para luego seguir con el desarrollo de una plataforma de procesamiento de audio a nivel de laboratorio.

En lo que sigue el documento está dividido en tres secciones. En la Sección 2 se detalla sobre la plataforma de procesamiento desarrollada. En la Sección 3 sobre los algoritmos utilizados para la implementación de dos analíticas de audio particulares: Detección de Niveles Predefinidos de Sonido y Detección de Sirenas. Por último en la Sección 4 se encuentra una breve reflexión sobre la sinergia de la pasantía laboral y los estudios de académicos, en el marco de la Maestría de Ingeniería Eléctrica actualmente en curso, a modo de conclusión.

1.1. Sobre SONDA Uruguay

SONDA Uruguay es fundamentalmente una empresa de servicios, proyectos de integración de sistemas y provisión de plataformas en el campo de las tecnologías de la información. Actualmente la empresa está incorporando proyectos de I+D en el área de tratamiento de señales. Tal es el caso del proyecto de desarrollo de analíticas de audio y video en tiempo real con aplicación a seguridad urbana. En el mismo, se encuentran involucradas ocho personas, profesionales y estudiantes de Ingeniería de Sistemas y Eléctrica, que complementan la rama de computación científica con el manejo de la ingeniería de software para aplicaciones críticas de gran volumen de datos.

2. Plataforma de procesamiento

Los sensores para Seguridad Urbana (en particular las cámaras de video y micrófonos) son diseñados para trabajar en tiempo real, transmitiendo a través de una red IP los datos adquiridos, a una central donde se realiza el monitoreo y la toma de decisiones. Es en este escenario donde los algoritmos de procesamiento de audio se encuentran embebidos y los tiempos de procesamiento no deben exceder la cadencia de disponibilidad de datos.

La plataforma de procesamiento, se puede dividir en dos grandes bloques: por un lado el de genereración de datos de audio y comunicaciones, y por otro la adquisición desde la red y procesamiento para la toma de decisiones, como se observa en la Figura 1.

Del lado de la generación de los datos de audio el tipo de micrófono (patrón polar, sensibilidad, etc.) y el codec (frecuencia de muestreo, bitdepth, etc.) definen la información sonora que se extrae del escenario monitoreado. El tipo de

análisis que se pueda hacer va a depender de las anteriores características del sistema.

Por otro lado el módulo de adquisición hace disponible las muestras de audio, en ventanas de largo configurable según los requerimientos del algoritmo. El análisis se hace ventana a ventana (frame by frame en la literatura) y en el caso particular de detección de sirenas (sección: 3.2), se utiliza una estrategia de integración temporal para aumentar la robustez a falsas alarmas, debidas al análisis local ventana a ventana.

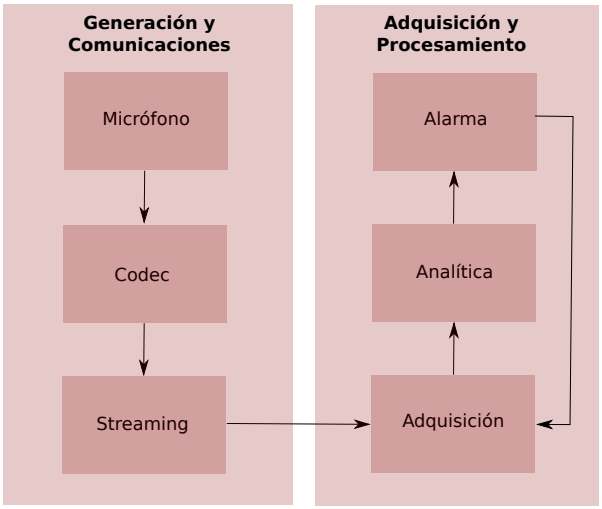

Figura 1: Esquema de la plataforma de procesamiento. A la izquierda el bloque de generación y comunicaciones, a la derecha el de adquisición y procesamiento.

3. Analíticas

Las analíticas son el modulo de procesamiento de audio que transforman los datos de entrada en una decisión para generar una alarma. Tienen el objetivo de alertar ante hechos predefinidos que puedan estar asociados a actividades vandálicas. En esta sección se describe sobre dos analíticas desarrolladas en la primer etapa del proyecto de desarrollo del software de analíticas de audio.

3.1. Analítica de Detección de Niveles Predefinidos de Sonido

Tiene como objetivo alertar la presencia de momentos de alto nivel sonoro o ausencia del mismo. Se hace una estimación de la energía, de la señal de entrada, frame a frame y se umbraliza para generar alertas en niveles energéticos predefinidos por el usuario.

3.1.1. Medidas de energía utilizadas

Se utilizan dos estimaciones de energía para detección de cambios energéticos de distinta naturaleza, por un lado cambios sostenidos en el tiempo y por otro lado los impulsivos.

RMS

Se utiliza para estimar la energía promedio de la ventana temporal. La ecuación para una ventana de largo N es de la siguiente forma (Lerch, 2012, Chapter 4):

$$RMS = \sqrt{\frac{1}{N}\sum_{n=1}^{N} x[n]^2} \tag{1}$$

Envolvente de pico

Se utiliza para detectar eventos energéticos impulsivos que pueden ser disimulados por el promediado del valor RMS. El cálculo para una ventana de largo N se realiza de la siguiente forma (Lerch, 2012, Chapter 4):

$$PICO = max(x[n]) \forall n \in 1..N \tag{2}$$

3.1.2. Detección de altos niveles sonoros

Para esta analítica se utiliza una estrategia de doble umbral sobre las estimaciones de energía detalladas anteriormente. En primer lugar se aplica al valor RMS de la ventana temporal, siendo útil frente a cambios energéticos sostenidos en el tiempo. Por otro lado se umbraliza el valor de pico, para detectar eventos impulsivos que no sean detectables en el promediado del RMS.

3.1.3. Detección de bajos niveles sonoros

Para la umbralización inferior sólo se trabaja con el valor RMS.

3.2. Analítica de Detección de Sirenas

Tiene como objetivo alertar frente a al presencia de una sirena en el audio de análisis. Para esta analítica se utilizó una estrategia de reconocimiento de patrones con el entrenamiento de clasificadores. Se resolvió con el enfoque clásico de reconocimiento de sonidos particulares para Seguridad Urbana de dos etapas, Lecomte et al. (2011):

a. Detección de eventos anómalos

b. Reconocimiento de los eventos anómalos detectados

Se trata de dos clasificadores anidados con una etapa de integración temporal intermedia. En la primer etapa, con un modelado no supervisado del ruido ambiente se detectan anormalidades en el audio, si existieron los suficientes eventos anómalos según la integración temporal, se pasa a la tercer y última etapa, donde se clasifica en sirena o no. Se observa un esquema de lo dicho anteriormente en la Figura 2.

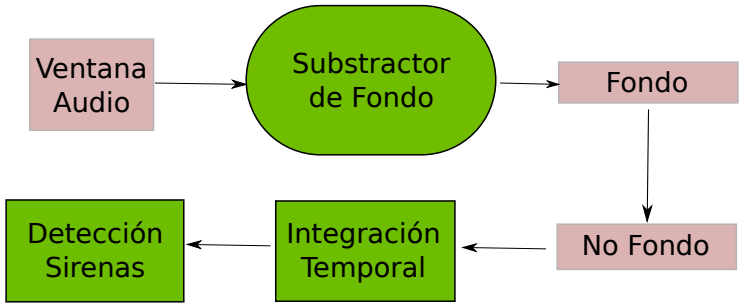

Figura 2: Esquema de la analítica de detección de sirenas. Se observa una primer etapa de substracción de fondo y posterior clasificación en sirena o no.

3.2.1. Substractor de Fondo

En aplicaciones de Seguridad Urbana, se utilizan algoritmos de substracción de fondo (en la literatura *Background Substraction*, Crocco et al. (2016)) como una primer etapa de análisis y conceptualización de la señal de entrada. En otras palabras estos algoritmos modelan los patrones que dominan el ambiente y detectan anormalidades.

El modelado del ruido ambiente con un enfoque supervisado tiene la dificultad de la generación de una base de datos que resuma y caracterice la infinidad de posibilidades que existen en un entorno urbano. Por esta razón se optó por un enfoque de clase unitaria utilizando One-Class SVM (Rabaoui et al. (2008), (Lecomte et al., 2011)).

Extracción de Características

Teniendo en cuenta que no se conoce nada a priori sobre el ruido ambiente y además el sistema está basado en un análisis por ventana (en la literatura frame by frame), para la extracción de características se utiliza el enfoque genérico de un banco de filtros lineal, para extracción de la energía en bandas de frequencia de la transformada de Fourier (Fourier-based linear filterbank (Lecomte et al., 2011)).

Entrenamiento y Clasificación

El entrenamiento se realiza en tiempo real, con fragmentos del propio ruido ambiente en donde el sistema se encuentra funcionando. Permite independencia de una base de datos para entrenamiento y robustez frente a cambios naturales en el escenario de análisis, como son la noche y el día por ejemplo.

El largo del fragmento es variable permitiendo adaptarse a escenarios complejos, sacrificando costo computacional. Se utiliza un clasificador One-Class SVM que detecta en cada ventana de entrada si pertenece o no al modelo de fondo generado en el One-Class SVM.

3.2.2. Integración temporal

Previo a la etapa de clasificación de sirena o no, se utiliza un esquema de integración temporal y una medida de distancia para decidir si continuar o descartar las muestras de audio.

Se integran ventanas adyacentes y mediante la distancia binaria de Huffman, con un umbral parametrizable, se define si ese fragmento mayor de audio pasa a la siguiente etapa de clasificación, observar Figura 3.

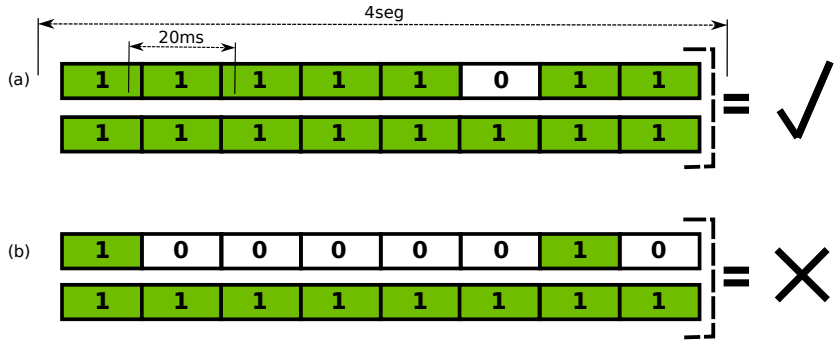

Figura 3: Esquema de la integración temporal de ventanas adyacentes en un fragmento de audio mayor, se ejemplifica con dos valores particulares en segundos. En verde los audios detectados como no fondo, en blanco los que pertenecen al fondo según el substractor de fondo. En (a) se observa un audio que pasa a la segunda etapa de clasificación, en (b) caso contrario.

3.2.3. Clasificación Sirena

La última etapa es un clasificador One-Class SVM que decide si el evento anómalo sostenido en el tiempo, según las etapas anteriores, clasifica como sirena o no.

Extracción de Características

Se utilizan al igual que en la publicación Salamon et al. (2014) los primeros 25 coeficientes MFCC (Mel-Frequency Cepstral Coeficients), calculados sobre un banco de filtros Mel de 40 bandas, generado con ventanas temporales de 20ms, 50 % de solapamiento y enventanado hamming. Lo anterior se computa sobre el audio completo que sale de la integración temporal (en la Figura 3 se ejemplifica con 4 segundos de audio) y se aplican estadísticas para resumir a un vector de 275 características: máximo, mínimo, media, mediana, varianza, skewness, kurtosis y media y varianza de la primer y segunda derivada de los coeficientes Mel a lo largo del tiempo.

Entrenamiento y Clasificación

Para el entrenamiento se utilizó un enfoque de clase unitaria utilizando One-Class SVM. Como base de datos para entrenamiento y test se utilizó la Urban-Sound dataset publicada en Salamon et al. (2014).

4. Conclusiones

La pasantía laboral generó sinergia de mis estudios académicos y mi trabajo en el sector privado. El desarrollo de analíticas de audio requiere de habilidades en procesamiento de audio que están vinculadas directamente con mi tema de tesis. Además se procedió con metodologías y herramientas aprendidas en la Maestría en Ingeniería Eléctrica de la Facultad de Ingeniería, como la realización de una revisión bibliográfica de publicaciones científicas para dominar el estado del arte, y la utilización de herramientas vistas en cursos de la maestría.

Por otro lado la posibilidad de aplicar directamente los conocimientos obtenidos en la Maestría de Ingeniería Eléctrica a la resolución de un problema particular con aplicación práctica, generó motivación tanto para los estudios académicos como para mi rol como ingeniero en la sociedad.

Referencias

Crocco, M., Cristani, M., Trucco, A., and Murino, V. (2016). Audio surveillance: A systematic review. *ACM Comput. Surv.*, 48(4):52:1–52:46.

Lecomte, S., Lengellé, R., Richard, C., Capman, F., and Ravera, B. (2011). Abnormal events detection using unsupervised one-class svm-application to audio surveillance and evaluation. In *Advanced Video and Signal-Based Surveillance (AVSS), 2011 8th IEEE International Conference on*, pages 124–129. IEEE.

Lerch, A. (2012). *An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics*. Wiley-IEEE Press, 1st edition.

Rabaoui, A., Kadri, H., Lachiri, Z., and Ellouze, N. (2008). One-class svms challenges in audio detection and classification applications. *EURASIP Journal on Advances in Signal Processing*, 2008(1):1–14.

Salamon, J., Jacoby, C., and Bello, J. P. (2014). A dataset and taxonomy for urban sound research. In *22st ACM International Conference on Multimedia (ACM-MM'14)*, Orlando, FL, USA.