COMPRESIÓN DE LA BASE DE DATOS PARA SÍNTESIS DE VOZ VISUAL

ROSANA IBARRA BECERRA
Ingeniera Electrónica

Trabajo de Investigación para optar por el título de
Magíster en Ingeniería Electrónica

Director
PEDRO RAÚL VIZCAYA GUARÍN
Ingeniero Electrónico Ph.D.

PONTIFICIA UNIVERSIDAD JAVERIANA
FACULTAD DE INGENIERÍA
DEPARTAMENTO DE INGENIERÍA ELECTRÓNICA
BOGOTÁ
2011

PONTIFICIA UNIVERSIDAD JAVERIANA

FACULTAD DE INGENIERÍA

MAESTRÍA EN INGENIERÍA ELECTRÓNICA

RECTOR MAGNÍFICO:	R.P. JOAQUÍN SÁNCHEZ GARCIA S.J.
DECANO ACADÉMICO:	Ing. LUIS DAVID PRIETO MARTINEZ PhD
DECANO DEL MEDIO UNIVERSITARIO:	R.P. ANTONIO JOSÉ SARMIENTO S.J.
DIRECTOR DE LA MAESTRÍA:	Ing. CARLOS ALBERTO PARRA PhD
DIRECTOR DEL TRABAJO DE INVESTIGACIÓN:	Ing. PEDRO RAÚL VIZCAYA GUARÍN PhD.

ARTICULO 23 DE LA RESOLUCIÓN No. 13 DE JUNIO DE 1946

"La universidad no se hace responsable de los conceptos emitidos por sus alumnos en sus proyectos de grado.

Sólo velará porque no se publique nada contrario al dogma y la moral católica y porque los trabajos no contengan ataques o polémicas puramente personales. Antes bien, que se vea en ellos el anhelo de buscar la verdad y la justicia".

AGRADECIMIENTOS

La autora expresa sus agradecimientos:

A Dios por darme la valiosa oportunidad de terminar con éxito esta etapa de mi carrera.

A mi familia por el apoyo y respaldo que me brindaron día a día.

Al Ingeniero Pedro Raúl Vizcaya, PhD y director de esta investigación por tener la paciencia y compartir sus valiosos conocimientos desde el inicio de esta investigación, los cuales fueron fundamentales para llevar a buen término este proyecto.

A mí amado Fernando por su apoyo y confianza.

Y a todas aquellas personas que de una u otra manera me animaron y me dieron fuerzas para finalizar con éxito esta etapa de mi carrera.

TABLA DE CONTENIDO

INTRODUCCIÓN 7
1. MARCO TEÓRICO 9
1.1 Sistema de Telefonía Visual 9
1.2 Base de Datos 10
1.3 Mapas de Autoorganización (SOFM) 10
1.4 Compresión de Video 12
1.5 Métodos de Compresión para Video 13
2. DESCRIPCIÓN GENERAL Y DIAGRAMA DE BLOQUES 14
2.1 Etapa de Entrenamiento 15
2.2 Etapa de Evaluación 16
3. DESARROLLOS 18
3.1 Creación de Secuencias de Video 18
3.1.1 Base de Datos 18
3.1.2 Selección de la Región de Interés (RI) 19
3.1.3 Segmentación y Parametrización 19
3.1.4 Clasificación de las Imágenes 21
3.1.5 Generación de Secuencias 23
3.1.6 Mapas Resultantes y Creación de Secuencias 24
3.2 Compresión de las Secuencias 26
3.2.1 Métodos de Compresión 27
3.2.2 Comparación de Secuencias comprimidas 27
4. ANÁLISIS DE RESULTADOS 32
5. CONCLUSIONES 33
BIBLIOGRAFIA 34
ANEXOS 36

LISTA DE FIGURAS

Figura 1. Transmisión del Servicio de Telefonía Visual 9
Figura 2. Recepción del Servicio de Telefonía Visual 9
Figura 3. Relación entre el mapa de características y el vector de pesos sinápticos [13] 11
Figura 4. Diagrama de Bloques General 14
Figura 5. Diagrama de Bloques de la Etapa de Entrenamiento 15
Figura 6. Diagrama de Bloques de la Etapa de Evaluación 16
Figura 7. Imágenes de la Base de Datos 19
Figura 8. Imagen Segmentada en Escala de Grises 20
Figura 9. Graficas de la DCT bidimensional a) Matriz de 128x64 coeficientes b) Coeficientes de las primeras 8x16 posiciones 20
Figura 10. Ejemplo de topología hexagonal 22
Figura 11. Mapa del Libro de Código generado por SOFM 23
Figura 12. Barridos realizados sobre el Libro de Código a) Zig-Zag Diagonal. b) Serpenteo c) Curva de Peano 24
Figura 13. Mapas Resultantes a) Zig-Zag Diagonal b) Serpenteo c) Curva de Peano 25
Figura 14. Mapas generados con SOFM: a) Aleatorio b) Natural 26
Figura 15. SNR vs Métodos de Compresión 28
Figura 16. Factor de Compresión vs Métodos de Compresión 28
Figura 17. SNR y Métodos de Compresión 29
Figura 18. Factor de Compresión vs Métodos de Compresión en la RI 29
Figura 19. Promedio de la Desviación Estándar de las Imágenes Diferencia y las Secuencias Generadas 30
Figura 20. Promedio de la Desviación Estándar de la Región de la Interés de las imágenes diferencia 30

INTRODUCCIÓN

En los últimos años la demanda de servicios de telecomunicaciones como la transmisión de imágenes, voz y video por una misma red se ha incrementado de forma exponencial lo cual ha llevado a la comunidad científica a un desarrollo acelerado de algoritmos de compresión que reduzcan de forma considerable para cada uno de estos servicios. Teniendo como base esta premisa, se han desarrollado diversos trabajos en la Pontificia Universidad Javeriana ([1], [2], [3], [4], [5]), que buscan reconstruir secuencias de video naturales para el servicio de Telefonía Visual de tal forma que puedan ser transmitidas por canales de muy baja capacidad.

Específicamente, en este trabajo se investigó, desarrolló y evaluó la creación de secuencias de video a partir de la generación de un libro de código reducido y la compresión de estas secuencias para el servicio de Telefonía Visual para luego ser trasmitidas por canales de muy baja capacidad. Actualmente los sistemas de Telefonía Visual se basan en estándares [6], [7] que no aprovechan de la mejor forma las cualidades de este tipo de secuencias, la información que presentan las secuencias de video comparada con las secuencias de audio en sistemas de telefonía visual es bastante menor [8], sin embargo estudios realizados por diversos investigadores como Green y Kuhl, Sumby y Pollack, O'Neill citados por Mark y Allen [9] y el trabajo desarrollado por Bárcenas et al. [10], concluyen que el video es útil para aumentar o mejorar la inteligibilidad de señales de voz en ambientes ruidosos, en valores que van desde uno hasta 15 decibeles, pero con la carencia del audio el mensaje es prácticamente incomprensible mientras que en el caso contrario no sucede lo mismo.

Entre las primeras aproximaciones al problema se encuentra la solución planteada por Bárcenas et al.[10]. En este trabajo se realizó mórfosis entre diferentes imágenes almacenadas para obtener secuencias de video creíbles; sin embargo el tiempo de procesamiento que esto implica no permite tener un sistema en tiempo real. Posteriormente se planteó la idea de realizar el procesamiento en el dominio de los parámetros (Machado y Santa [2]), surgiendo así una alternativa que permitió implementar un conversor de texto a voz visual en tiempo real (AVSS). A partir de las necesidades de segmentación de la región de interés (boca), se creó el sistema SPARV (Sistema de segmentación automática de rostros en video) (Baptiste y Sotomayor [3]) en el cual se analizaron diferentes métodos de parametrización, incluyendo la transformada discreta de Fourier, y un método para la segmentación

automática de la boca utilizando las características de movimiento y la detección de la piel por medio de color. Hacia finales del 2003 se presentó un trabajo que analizó a fondo diferentes transformaciones como métodos de parametrización ([4]), incluyendo la DCT y análisis de componentes principales. Igualmente se desarrolló un algoritmo de interpolación entre imágenes que genera secuencias naturales. Se generó un libro de códigos de tamaño reducido. Ésta investigación permitió la implementación completa del sistema en tiempo real. El proyecto “Telefonía Visual por canales de muy baja capacidad” ([6],[7]) facilito unir todos los logros desarrollados en trabajos anteriores e implementar un sistema prototipo de videofonía.

La investigación se desarrollo en dos etapas: la primera de Entrenamiento que consta de los dos módulos y la segunda de Evaluación que tiene el tercer módulo; en el primer módulo se diseño a partir del libro de código, en el segundo se armaron las secuencias de tal manera que entre cada imagen existieran transiciones muy suaves y finalmente en el tercer modulo se emplearon diferentes métodos de compresión para comprimir el libro de código. El diseño del nuevo libro de código se realizó inicialmente reduciendo la base de datos original utilizada en la investigación de “Telefonía Visual”, por medio de la transformada discreta de coseno (DCT) con el fin de obtener las imágenes que tienen la mayor cantidad datos correlacionados (mayor cantidad de energía) las cuales se clasificaron utilizando los Mapas de Autoorganización de Características (SOFM) de Kohonen ([11])

La creación de las secuencias a partir del libro de códigos obtenido por los Mapas de Autoorganización (SOFM) se desarrollo teniendo en cuenta el mínimo cambio entre imágenes, para esto se realizaron diferentes tipos de barridos sobre una matriz 8x8 la cual contiene los índices de cada imagen. En la etapa de Evaluación se utilizaron diferentes métodos de compresión para video, utilizando el software Adobe Premiere Pro Vr 7.0, el cual trabaja con diez compresores para video de los cuales se emplearon seis de ellos dado que los otros compresores son para otros tipos de video con características diferentes al utilizado en la investigación.

Este documento plantea una nueva propuesta de compresión de la Base de Datos a un tamaño considerablemente reducido, con una relación señal a ruido por encima de lo óptimo y con un excelente Factor de calidad, de esta manera facilita la comunicación de Telefonía Visual.

1. MARCO TEÓRICO

1.1 Sistema de Telefonía Visual

El sistema de Telefonía Visual se compone de un transmisor y receptor como se ilustra en los diagramas en bloques de las Figuras 1 y 2, en donde se indica que la Base de Datos[1] se transmite fuera de línea (off-line), antes de establecer la comunicación con el receptor. Luego de establecer la comunicación lo que se envía durante la llamada al receptor son los índices de las imágenes ya enviadas fuera de línea con el fin de recrear el video.

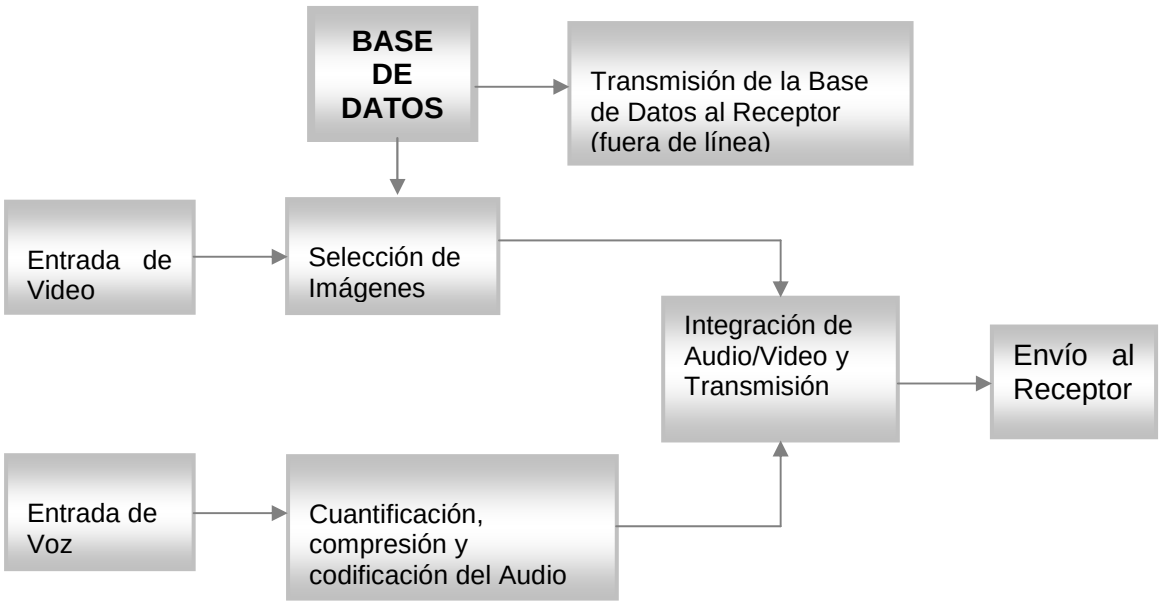


Figura 1. Transmisión del Servicio de Telefonía Visual

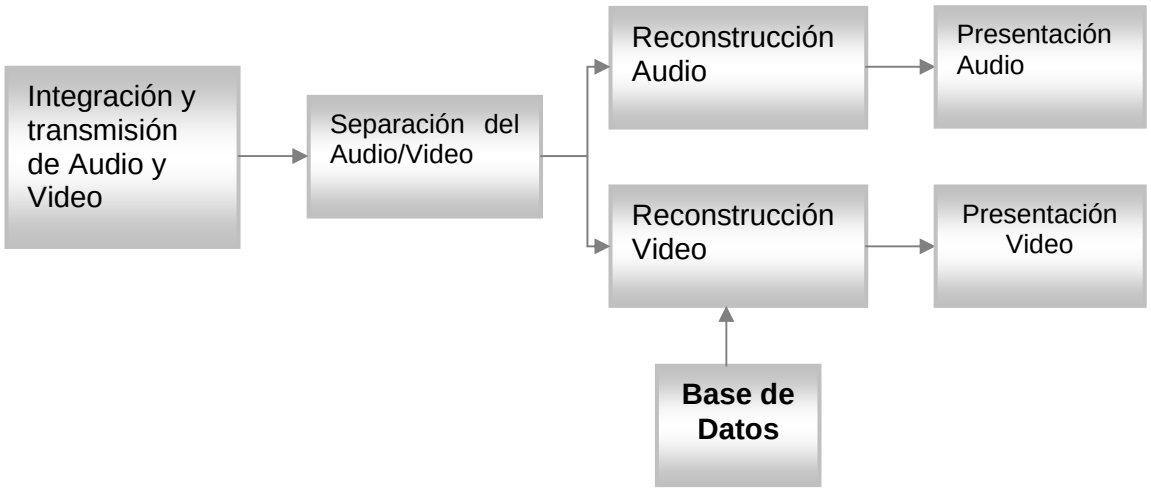


Figura 2. Recepción del Servicio de Telefonía Visual

[1] Conjunto de imágenes

1.2 Base de Datos

La Base de Datos es un video con un tamaño de 495 MB, está conformado por 856 imágenes, el formato del video es .avi, el cual fue diseñado para el servicio de Transmisión de Telefonía Visual ([2]), sus características son: texto natural (es un texto rico en vicemas[2]), duración del video de 30 segundos a 30 cuadros por segundo y resolución de 496x408 a color.

1.3 Mapas de Autoorganización (SOFM)

Una de las principales aplicaciones de los mapas de autoorganización de características, desarrollados por Kohonen ([12], [13]), consiste en transformar una señal patrón de dimensiones arbitrarias en un mapa discreto de una o dos dimensiones. Esta transformación se realizar de forma que se adopte un ordenamiento topológico. Kohonen basó su desarrollo en el funcionamiento del cerebro humano y la forma como las capas de la corteza cerebral se organizan de acuerdo a la función fisiológica que desarrollan. Una aproximación inicial se plantea desde el punto de vista de las redes neuronales. El algoritmo de puede resumir de la siguiente forma:

1. **Iniciación**. Escoger valores aleatorios para el vector inicial de pesos sinápticos $Wj(0)$. La única restricción que existe es que $Wj(0)$ debe ser diferente para $j = 1, 2, ...N$, donde N es el número de neuronas.

2. **Muestreo**. Tomar una muestra x de la entrada con cierta distribución de probabilidad; el vector x representa la señal sensada.

3. **Criterio de similaridad**. Encontrar la neurona que "mejor se acomoda" (neurona ganadora) $i(x)$ en tiempo n, usando el criterio de mínima distancia Euclidiana.

$$i(x) = \arg\min \|x(n) - Wj\| \; j = 1, 2, ...N \qquad [2.1]$$

[2] Un visema es el equivalente visual de un fonema. Es la representación facial que se asocia con uno o más fonemas determinados. Esta asociación entre fonemas y visemas depende de las características articulatorias de cada persona y del idioma o dialecto en particular.

4. **Actualizar.** Ajustar los vectores de pesos sinápticos de todas las neuronas, usando la fórmula de actualización que involucra a los vecinos (regla de Kohonen).

$$W_j(n+1) = \begin{cases} W_j(n) + \eta(n)\left[x(n) - W_j(n)\right] & j \in Ai(x)(\eta) \\ W_j(n), & \text{de otra forma} \end{cases} \qquad [2.2]$$

Donde $\eta(n)$ es un parámetro conocido como la tasa de aprendizaje, y $\mathrm{A}i_{(x)}(\eta)$ es la función de aprendizaje alrededor de la neurona ganadora i; tanto la tasa como la función de aprendizaje cambian dinámicamente para obtener los mejores resultados.

5. Volver al paso 2 hasta no observar cambios en el mapa resultante.

El método de SOFM posee tres propiedades:

La aproximación del espacio de entrada. Un mapa de autoorganización de características Φ, representado por el conjunto de vectores de peso sináptico $\{ Wj \,|\, j = 1, 2, ..., N \}$, en un espacio de salida A, provee una buena aproximación del espacio de entrada X.

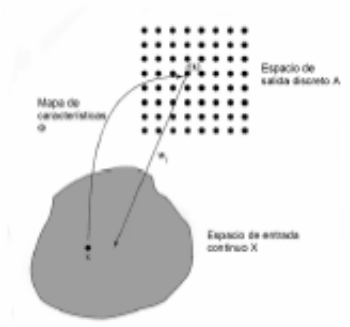


Figura 3. Relación entre el mapa de características y el vector de pesos sinápticos [13]

Ordenamiento Topológico. El mapa de características Φ calculado con el algoritmo SOFM esta topológicamente ordenado en el sentido de la localización espacial de los nodos en la estructura, la cual corresponde a un dominio particular de las características de los patrones de entrada.

Correspondencia de la función densidad de la entrada. El mapa de características Φ refleja las variaciones estadísticas de la distribución de la entrada.

1.4 Compresión de Video

La compresión de video es el proceso en el cual se realiza la eliminación de datos redundantes con el fin de disminuir considerablemente el tamaño de un archivo. Los archivos de video digital son muy grandes, requiriendo gran velocidad de transferencia de datos en la lectura y reproducción, esto ha llevado al desarrollo de nuevos Codec's los cuales son algoritmos que tiene la capacidad de comprimir y descomprimir, ellos utilizan básicamente dos sistemas para realizar este proceso.

Compresión sin pérdida, Es aquella que conserva de forma intacta los datos originales, este tipo de compresión consiste en descartar regiones de similares colores entre imágenes, además su factor de compresión es muy pequeño de 1:2.

Compresión con pérdidas, Es aquella que elimina la información que no percibe el ojo humano, este tipo de compresión no recupera la información perdida. La cantidad de información eliminada depende del factor de compresión y es proporcional al factor de calidad.

Dentro de los algoritmos de compresión con pérdida hay dos tipos:

Algoritmos de compresión espacial, Estos algoritmos comprimen cada imagen del video de forma independiente, sin tener en cuenta el resto de imágenes.

Algoritmos de compresión temporal, Estos algoritmos comprimen basándose en la variación entre frames[3] la cual no se almacena en su totalidad sino que se comprime. Este tipo de algoritmo utiliza los llamados *Keyframes*, que son frames completos del video que se almacenan con poca o ninguna compresión, las cuales se toman como referencia para generar los siguientes frames que son llamados Deltaframes.

Otra característica de lo los algoritmos de compresión es simetría es decir un algoritmo es simétrico cuando el tiempo de compresión y descompresión es el mismo, en caso contrario se dice que el algoritmo es asimétrico.

[3] Imágenes

1.5 Métodos de Compresión para Video

En la actualidad existen diferentes algoritmos de compresión estándares para video los cuales se encuentran en su mayoría reglamentados por la ITU-T[4]. La mayoría de Codec's se descargan de forma gratuita por Internet, además existe programas para edición de video, los cuales incluyen una amplia variedad de Codec's como es el caso de software Adobe Premiere Pro vr 7.0 ([14]) que incluye entre otros los siguientes algoritmos de compresión:

Cinepak Codec by Radius, Es un codec de audio y video temporal de alta calidad, realiza compresión de video Blanco&Negro y a color, la profundidad del color es determinada por "millones de colores", se diseño para tener mínimas pérdidas su compresión es asimétrica, además permite variación en el factor de calidad.

DivX Mpeg4, Este codec que resulta es una variación del formato de compresión de video Mpeg-4 que se caracteriza por una alta tasa de compresión, es muy usado en DVD y para hacer descargas en Internet, no permite variación en el factor de calidad.

Indeo® Video 5.10, Es un codec con una alta calidad, la profundidad de color es determinada por "millones de colores", la compresión puede variar según el factor de calidad, es ampliamente utilizado en Internet, emplea un sistema progresivo de descarga que se adapta al ancho de banda y al flujo de señal.

Intel Indeo® Video 4.5, Es un codec que solo exporta video, permita variación en el factor, la profundidad de color es determinada por "millones de colores", esta variación fue desarrollada por Intel.

Microsoft Windows Media Video 9.0, Es un codec que reproduce audio y video una de las características notables es su propiedad de escalabilidad, no permite variación del factor de calidad.

[4] Unión Internacional de Telecomunicaciones

2. DESCRIPCIÓN GENERAL Y DIAGRAMA DE BLOQUES

Para llevar a cabo el objetivo general del trabajo se plantea un sistema basado en el diagrama de bloques de la figura 4, el cual ilustra de forma general los pasos del proyecto.

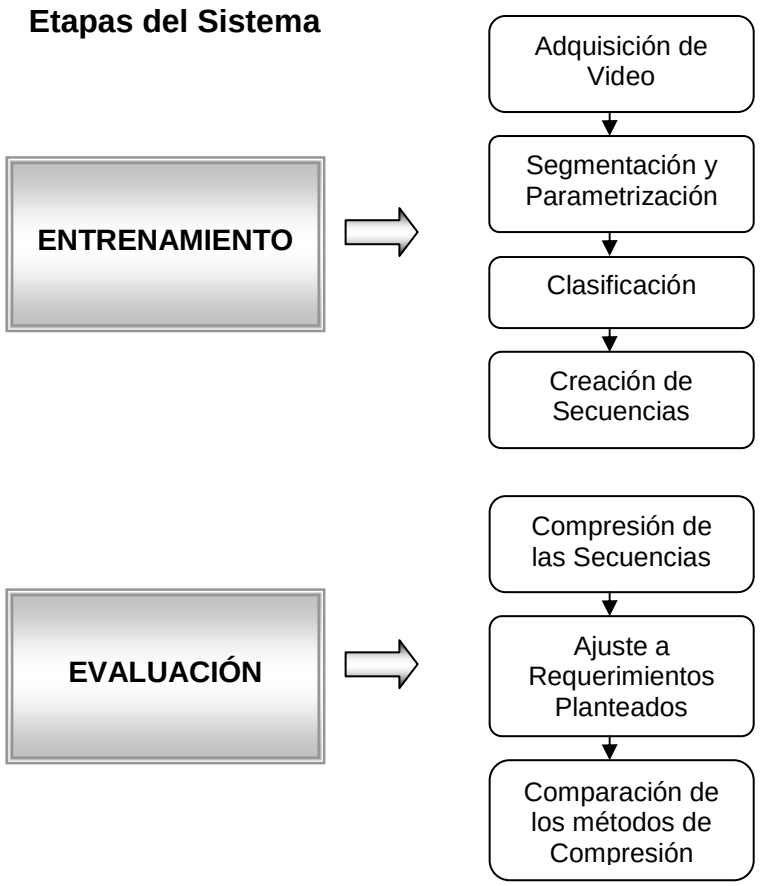


Figura 4. Diagrama de Bloques General

Existen dos etapas claramente definidas. La etapa de entrenamiento y la etapa de evaluación. La etapa de entrenamiento se encarga de crear un libro de códigos y armar las secuencias de video. En la etapa de evaluación se utiliza los diferentes métodos de compresión para video y se realizan los ajustes a los requerimientos mínimos planteados anteriormente.

2.1 Etapa de Entrenamiento

Durante esta etapa se realizan tres procesos muy importantes. La parametrización, generación del libro de código y creación de la secuencias.

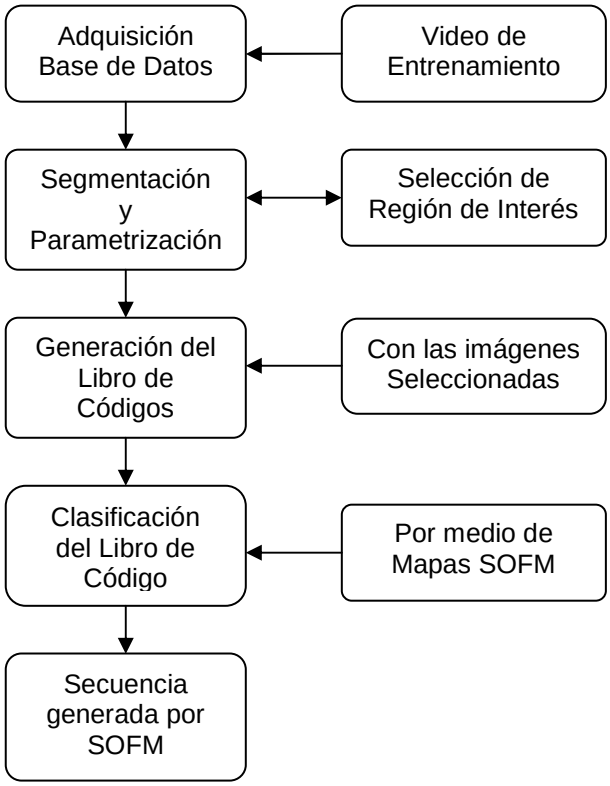


Figura 5. Diagrama de Bloques de la Etapa de Entrenamiento

Segmentación y Parametrización, Para reducir la complejidad y el costo computacional en el procesamiento; se seleccionó una región de interés (RI) común a todas las imágenes de la Base de Datos, sobre la cual se calculó la transformada discreta de coseno bidimensional (DCT), la cual comprimió y agrupo las imágenes que tienen la mayor cantidad de energía.

Generación del Libro de Código, A partir de las imágenes seleccionadas por la DCT se realizó el entrenamiento utilizando los Mapas de Autoorganización de Características (SOFM), los cuales clasificaron las imágenes teniendo en cuenta diferentes parámetros como: la topología del mapa, la distancia euclidiana y el tipo de grilla entre otros.

Creación de las Secuencias, El resultado de clasificación arrojada por SOFM se tomo como punto de partida (orden de las imágenes) sobre las cuales se realizaron posteriormente diferentes tipos de barridos y así se crearon otras nuevas secuencias.

2.2 Etapa de Evaluación

La Etapa de Evaluación se encarga de realizar lo más importante del trabajo que es encontrar las secuencias con transiciones suaves y comprimirlas a un tamaño que se ajuste a los requerimientos planteados en el proyecto.

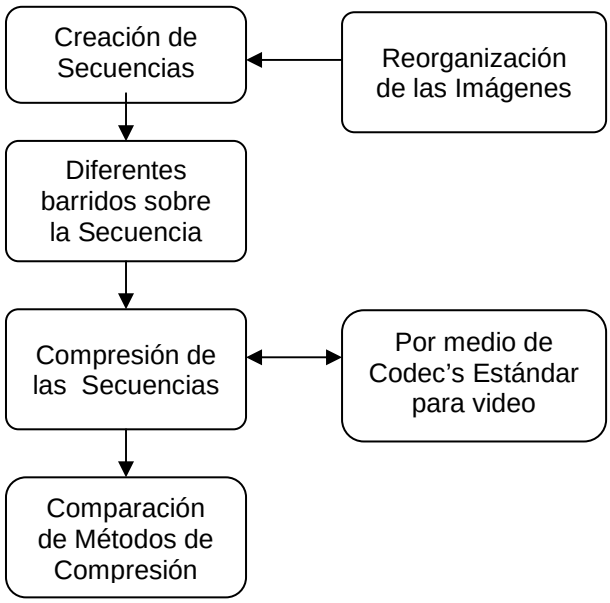


Figura 6. Diagrama de Bloques de la Etapa de Evaluación

Creación de las Secuencias, A partir de los resultados obtenidos del entrenamiento de con SOFM se creó una matriz de 8x8 sobre la cual se realizaron tres barridos diferentes con el fin de generar secuencias con transiciones más suaves y así se obtuvieron videos mas naturales.

Compresión de las Secuencias, Luego de armar las secuencias con los diferentes barridos se realizó la compresión de cada una utilizando diferentes compresores estándar para video. Los codec's que se utilizaron se encuentran en el software Adobe Premiere Pro vr 7.0, estos compresores permitieron la variación de diferentes parámetros como: el factor de calidad, el tamaño los frames, el numero de muestras por segundo (fps), la cantidad de colores del video y el tamaño de los píxeles. Por lo cual este programa fue muy versátil para lo que se busco durante la investigación.

Comparación de las Secuencias Comprimidas, Se compararon todas las secuencias comprimidas generadas por los tres barridos y dos adicionales sin algún tipo de barrido con el fin de comparar dos parámetros importantes la relación de señal a ruido (SNR) y el factor de calidad.

3. DESARROLLOS

Los desarrollos realizados en este trabajo de investigación se centran en dos puntos críticos para la Compresión de la Base Datos. En primer lugar se crearon las secuencias de video y en segundo lugar se realizó la compresión de los videos resultantes.

3.1 Creación de Secuencias de Video

Inicialmente se contaba con tres videos de entrenamiento diferentes de los cuales se escogió finalmente el desarrollado por Santa y Machado, ya que este video es el que mejor se ajustó a lo que se buscaba durante la investigación. Tomando como base de datos original el video de Santa y Machado ([2]), se seleccionó una Región de Interés (RI) la cual se segmentó y parametrizó, a estos parámetros se les aplicó la transformada discreta de coseno bidimensional (DCT), luego se realizó la clasificación de los parámetros por medio de los Mapas de Autoorganización de Características (SOFM) y finalmente se crearon las secuencias. Los desarrollos descritos a continuación y sus posteriores resultados muestran el proceso de creación del Libro de Código y la creación de las secuencias y como afectó notoriamente cada uno de los pasos realizados para encontrar secuencias más suaves y así realizar una compresión de la base de datos para síntesis de voz visual.

3.1.1 Base de Datos

La base de datos original en un video que se desarrolló bajo condiciones controladas con el fin de realizar transmisión de telefonía visual, el video muestra el rostro de una persona pronunciando frases específicas con el fin de tener diferentes posiciones de los labios, como se muestra en la Figura 7.

Figura 7. Imágenes de la Base de Datos

3.1.2 Selección de la Región de Interés (RI)

Para reducir la complejidad y el costo computacional se decidió seleccionar una Región de Interés (RI) que es común a todas las imágenes de la base de datos, la RI seleccionada fue los labios ya que ella registra mayores cambios con respecto a otras regiones del rostro.

3.1.3 Segmentación y Parametrización

De cada una de las imágenes de la base de datos se selecciono la RI la cual se represento como un conjunto reducido de parámetros, para esto de cada imagen se extrajo una sección de 128x64 píxeles y se convirtieron a escala de grises obteniendo una imagen como la que se observa en la Figura 8.

Figura 8. Imagen Segmentada en Escala de Grises

Sobre la RI se calcula la transformada discreta de coseno bidimensional DCT, de la cual se obtienen una matriz de coeficientes de 128x64 posiciones. De esta matriz, se escogen los elementos ubicados en las primeras 8x16 posiciones, puesto que en estas primeras posiciones está concentrada cerca del 96% de la energía total de las imágenes como se muestra en la Figura 9.

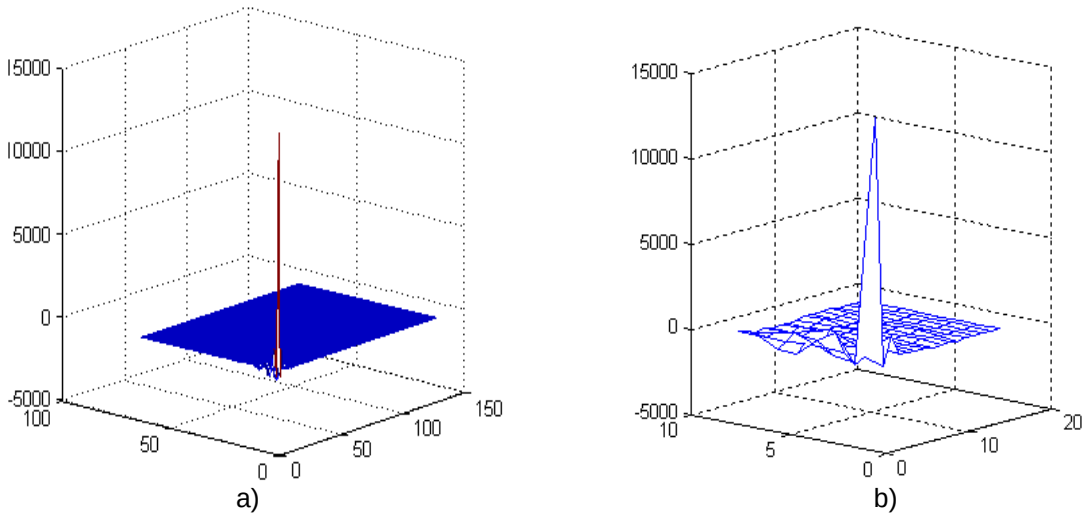


Figura 9. Graficas de la DCT bidimensional a) Matriz de 128x64 coeficientes b) Coeficientes de las primeras 8x16 posiciones

Cada imagen queda entonces representada como un vector de características de 128 posiciones. Este vector se construye de la siguiente forma:

$$v(8k+n) = c(k,n)\, 0 \le k \le 7, 0 \le n \le 15 \qquad [3.1]$$

En donde, c(k,n) representa la transformada coseno de la imagen en la posición (k,n), y v(8k + n) representa el vector de características final.

3.1.4 Clasificación de las Imágenes

La clasificación de las imágenes se realizó utilizando los Mapas de Autoorganización de Características SOFM o mapas de Kohonen, como método de clasificación no supervisada en el cual el sistema aprende a clasificar de forma automática los vectores de parámetros y dependiendo cómo se encuentren agrupados en el espacio original genera a la salida un conjunto de vectores representantes y un mapa topológico en el que dichos representantes son ubicados manteniéndose cerca de aquellos vectores de características similares ([11]). Luego se pasa a un espacio n-dimensional (espacio original) a un espacio de menos dimensiones y menor distancia.

Antes de realizar el entrenamiento del Libro de Código es necesario definir algunos parámetros de la red como son: la dimensión, el tamaño, la topología y la distancia.

Dimensión del Libro de Código, Durante la investigación se decidió trabajar con las primeras 64 imágenes, esta cantidad de frames facilita la creación de mapas simétricos y es suficiente para incluir las diferentes posiciones de los labios y algunas imágenes adicionales para generar transiciones suaves entre ellos.

Tamaño de la Grilla, Con esta cantidad de imágenes surgieron diferentes tipos de tamaños entre los cuales estaban 8x8, 7x9, 6x11 y 4x16, estos tamaño fueron evaluados anteriormente ([11]), a partir de estos resultados se decidió trabajar con el tamaño 8x8 por la simetría de los mapas resultantes.

Topología del Mapa, Es la estructura que se establece inicialmente para asociar las imágenes durante su entrenamiento, en SOFM existe 3 tipos de topología que son la cuadrada, hexagonal y la aleatoria de las cuales se selecciono la hexagonal por la mayor cantidad de conexiones que puede tener respecto a la cuadrada, lo cual le permite a la red tener mayor número de opciones para correlacionar las imágenes, esto es muy importante para la creación de secuencias suaves entre imágenes ya que se generarían secuencias mas naturales. La Figura 10 se muestra un ejemplo de la topología hexagonal, en la cual los puntos rojos representan cada una de las imágenes o índices de las imágenes en este caso del libro de código.

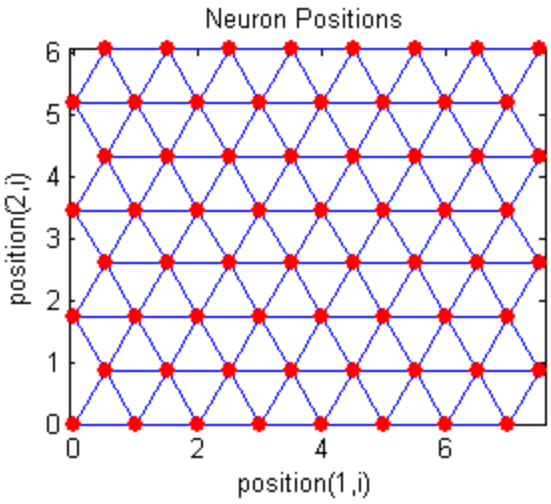


Figura 10. Ejemplo de topología hexagonal

Otros parámetros que se tuvieron en cuenta para el entrenamiento fueron: la entrada de los datos y la distancia de los parámetros y la grilla.

Entrada de los Datos, El orden de entrada de los datos a la red es de forma secuencial por lo cual influye directamente en su salida, entonces si la secuencia de entrada es siempre la misma tendremos el mismo resultado a la salida, por esta razón se organizaron las imágenes de tal manera que ingresen 150 secuencias aleatorias diferentes lo que genero 150 mapas diferentes.

Distancia, Al finalizar el entrenamiento se obtuvieron a la salida los vectores de los parámetros de las imágenes representantes y las coordenadas que indican la posición de ellos en el mapa, estos parámetros no necesariamente pertenecen a alguna de las imágenes del video de entrada, por lo cual se selecciono del video de entrenamiento las imágenes con los parámetros más parecidos a los arrojados por el mapa; para ello se midió la distancia Euclidiana entre los parámetros de los representantes con cada una de las imágenes del video de entrada y se seleccionaron las imágenes que arrojaron la menor distancia y con ellas se obtuvo el libro de código.

En la Figura 11 se muestra el Mapa resultante del Entrenamiento realizado por SOFM, que llamaremos Libro de Código, se puede observar cómo están agrupadas las imágenes semejantes: bocas cerradas, bocas semiabiertas y bocas abiertas. En este proceso se pasa de un espacio n-dimensional (espacio original) a un espacio con pocas dimensiones (mapa generado) facilitando la visualización y el análisis de la relación de las imágenes obtenidas ([11]).

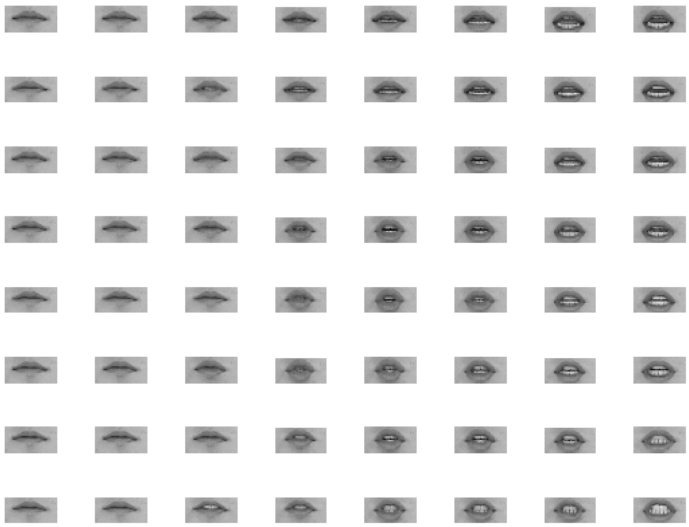


Figura 11. Mapa del Libro de Código generado por SOFM

El nuevo libro de código generado por SOFM tiene un tamaño de 37 MB, mostrando una reducción del 92.525% del tamaño de la base de datos. Este libro de código se tomo como punto de partida para crear nuevas secuencias por medio de diferentes barridos y así armar secuencias suaves.

3.1.5 Generación de Secuencias

El proyecto tiene como uno de sus objetivos específicos la generación de secuencias con transiciones suaves por lo cual se armaron las secuencias empleando 3 barridos diferentes (zig-zag diagonal, serpenteo y curva de peano). Cada uno de estos barridos tiene características comunes como son:

- No pasa dos veces por la misma imagen.
- Pasan por todas las imágenes de la matriz.
- Cada barrido realiza recorridos diferentes.

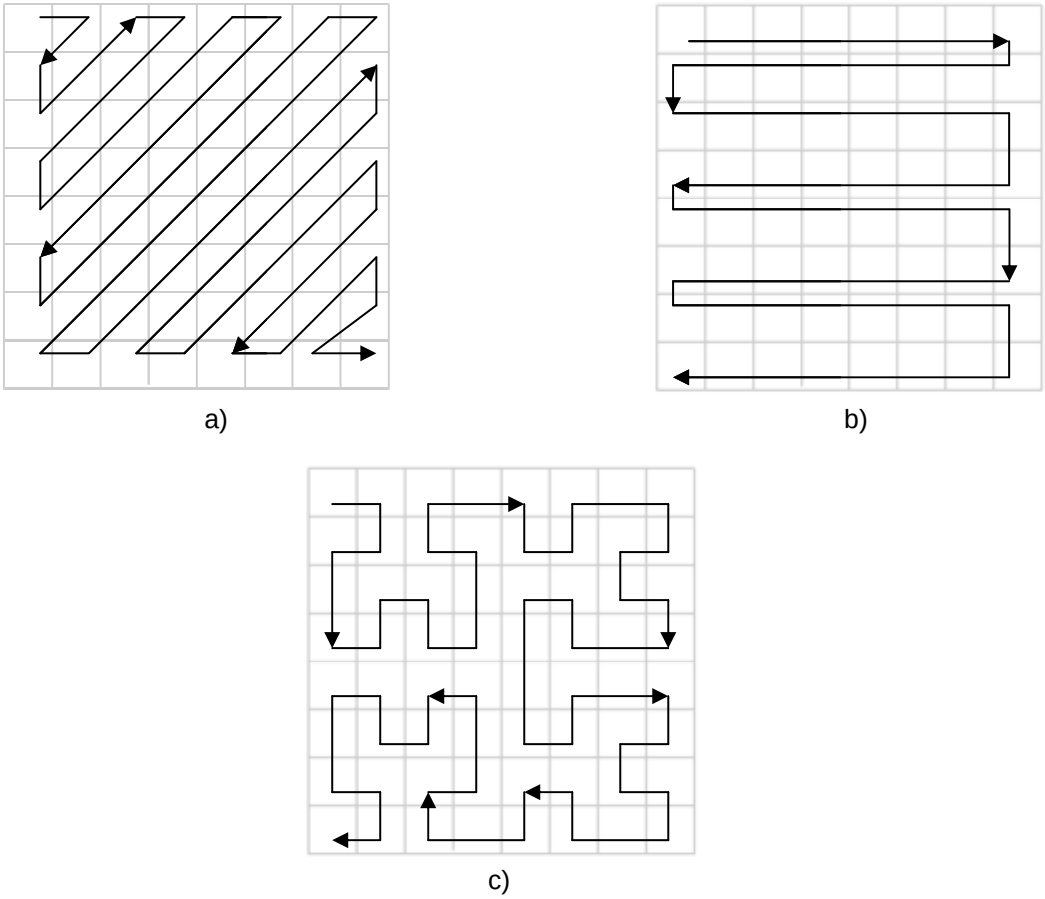


Figura 12. Barridos realizados sobre el Libro de Código a) Zig-Zag Diagonal. b) Serpenteo c) Curva de Peano

Los tres barridos se hicieron sobre los índices del Libro de Código como lo muestra la Figura 11. La finalidad de realizar estos barridos fue crear secuencias de video con transiciones más suaves entre cada imagen.

3.1.6 Mapas Resultantes y Creación de Secuencias

Una vez obtenido el libro de código se realizaron los diferentes barridos de los cuales se obtuvieron los mapas que se observan en la Figura 13.

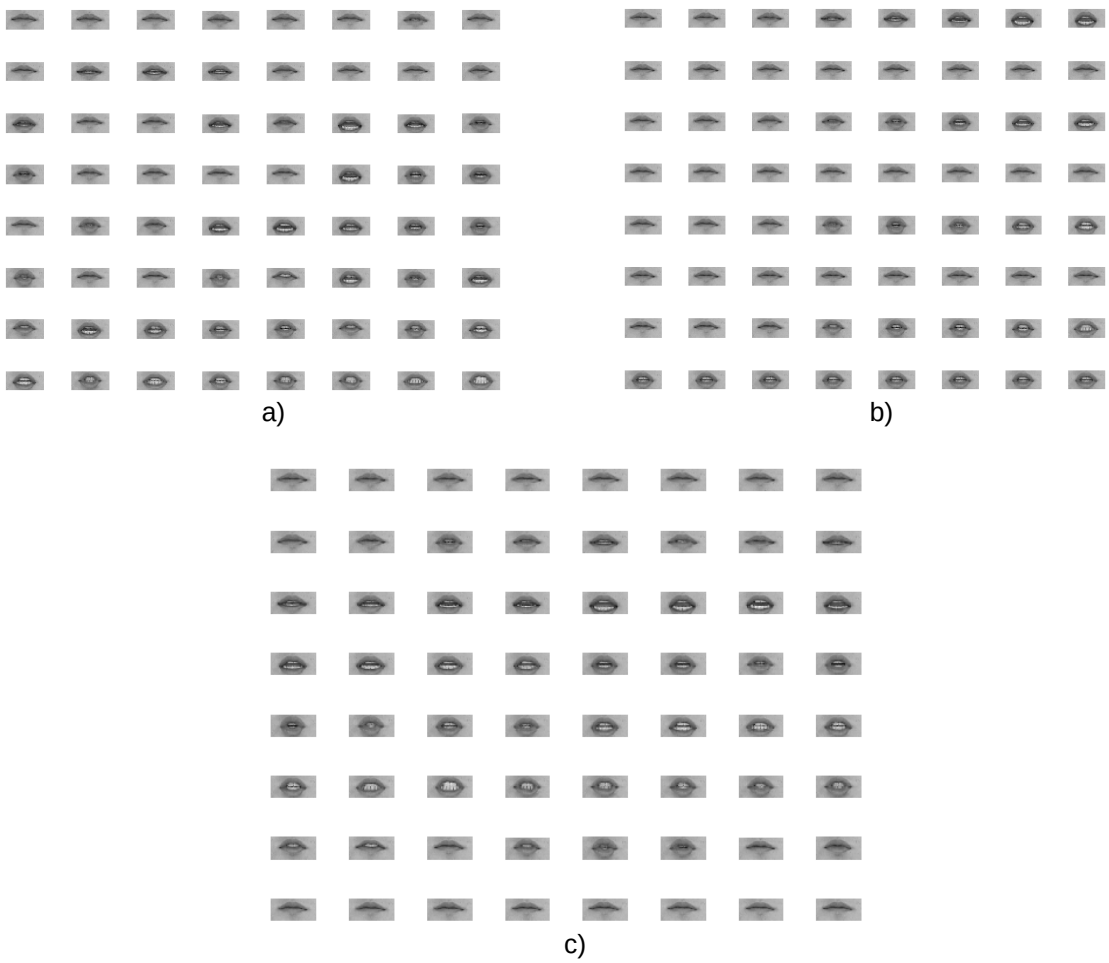
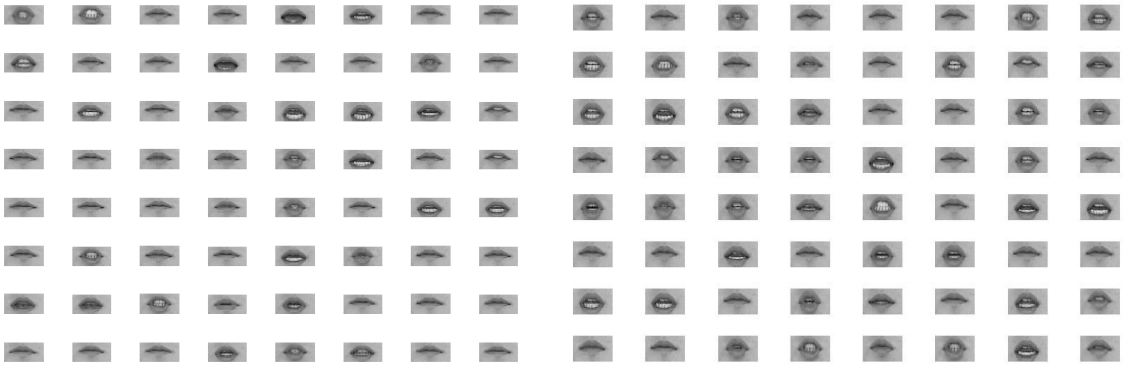


Figura 13. Mapas Resultantes a) Zig-Zag Diagonal b) Serpenteo c) Curva de Peano

Las imágenes generadas se organizaron en una matriz de 8x8 imágenes las cuales se leen de izquierda a derecha. Cada mapa muestra la nueva organización de las imágenes después de cada barrido, en ellos se observa como las imágenes bocas abiertas, semi abiertas y cerradas crean grupos pequeños y entre ellas están las bocas cerradas las cuales suavizan la secuencia.

Luego se generaron dos secuencias: la Aleatoria y la Natural, las imágenes de estas secuencias no tiene un orden determinado por algún tipo de barrido, la secuencia aleatoria como su nombre lo indica son imágenes se tomaron de forma desorganizada, por otro lado la secuencia natural es el resultado inicial del libro de código generado por SOFM. Estas secuencias no tienen un barrido determinado se generaron con el fin de evidenciar si el orden del las imágenes aumenta el factor de compresión significativamente. Los mapas generados, secuencia aleatoria y natural se muestra en la Figura 14.



a) b)

Figura 14. Mapas generados con SOFM: a) Aleatorio b) Natural

3.2 Compresión de las Secuencias

La compresión de las secuencias se realizó utilizando diferentes métodos de compresión para video estándar, los cuales se encontraban en el software de Adobe Premiere Pro vr 7.0; este es un programa que se emplea para edición de video, además entre sus aplicaciones encontramos 6 diferentes formatos de compresión para video que son:

- Cinepak by Radius 1.10
- Divx-4Mpeg Low-Motion 4.1
- Divx-4Mpeg Fast-Motion 4.1
- Microsoft Windows Media Video 9.0
- Indeo ® Video 5.1
- Intel Indeo ® Video 4.5

Todos permiten variar los siguientes parámetros antes de realizar la compresión: profundidad del color, tamaño de la trama, el número de tramas por segundo (fps), el tamaño de los píxeles y el factor de calidad. De estos parámetros los primeros 4 codec's no permiten variación en el factor de calidad, los otros dos permiten variación en el factor de calidad de 0% hasta 100%.

Los parámetros mencionados en el párrafo anterior, se variaron de la siguiente manera:

- **Profundidad del color**, se utilizó la máxima cantidad de colores que fue la opción de "millones de colores".
- **Tamaño de la trama**, se utilizó el tamaño original de las imágenes de la secuencia generado por el Libro de Código 408x496.

- **Numero de Tramas por segundo (fps)**, se utilizó el mismo valor que el video original 30 fps.
- **El tamaño de los píxeles**, se utilizó el mismo con el que se realizó el video píxeles cuadrados (1.0), aunque el programa da 7 opciones adicionales.
- **El Factor de calidad**, la variación del factor de calidad se realizó desde 100% (la mayor calidad) hasta un 50% de calidad.

3.2.1 Métodos de Compresión

Las abreviaturas para los métodos de compresión estándar para video que se utilizaron durante la investigación fueron:

- Cinepak by Radius 1.10 [5](cpk)
- Divx-4Mpeg Low-Motion 4.1 (divxl)
- Divx-4Mpeg Fast-Motion 4.1 (divxf)
- Microsoft Windows Media Video 9.0 (microsoft)
- Indeo ® Video 5.1 (indeo 5.1)
- Intel Indeo ® Video 4.5 (intel 4.5)

Se abreviaron los nombres de los compresores de la forma presentada con el fin de facilitar la escritura en las graficas resultantes.

3.2.2 Comparación de Secuencias comprimidas

Los métodos de compresión se compararon teniendo en cuenta dos parámetros muy importantes para imágenes y video son la relación señal a ruido (SNR) y factor de compresión, de la siguiente manera:

- La relación señal a ruido (SNR), es un parámetro que define el margen que hay entre la potencia del libro de código y la potencia del ruido del libro comprimido, tomando como valor optimo 20 dB.
- El factor de compresión, indica en cuánto se redujo el tamaño del Libro de código con diferentes barridos y métodos de compresión con respecto al Libro de código original, teniendo como referencia 10 a 1 mínimo.

Estos dos parámetros se analizaron durante la compresión de las secuencias generadas mediante seis métodos de compresión diferentes, dando como resultado las graficas presentadas en las figuras 15, 16, 17 y 18.

[5] Nombre abreviado del método de compresión

En la Figura 15 se visualiza la relación señal a ruido (SNR) de cada uno de los métodos de compresión utilizados en cada una de las secuencias generadas, lo cual dio como resultado valores por encima de los 20 dB, dando como resultado un margen aceptable planteados inicialmente.

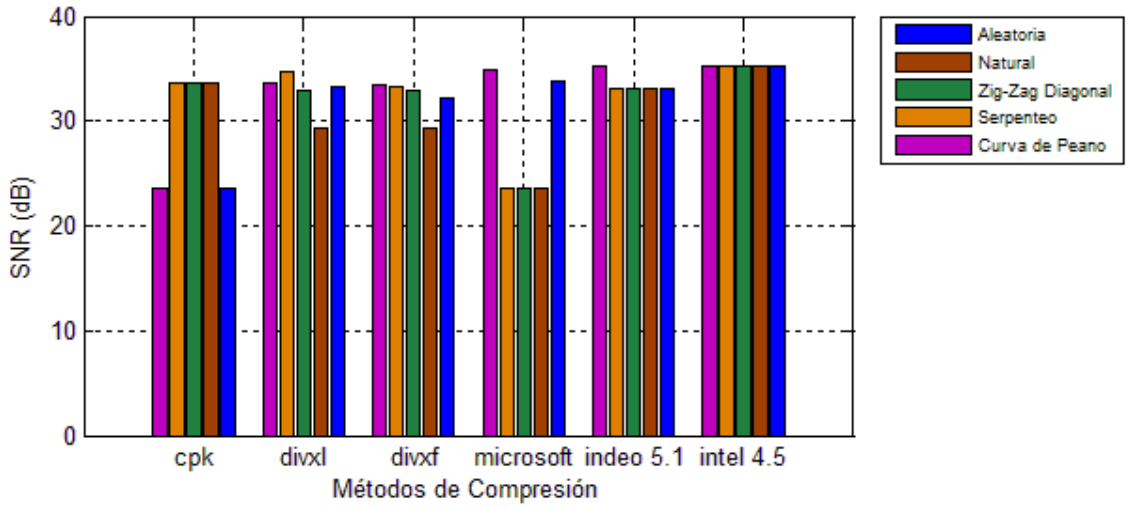


Figura 15. SNR vs Métodos de Compresión

En la siguiente Figura 16 se muestra la relación entre el Factor de Compresión con cada uno de los Métodos de Compresión utilizados de las cinco secuencias generadas durante la investigación, lo cual evidencio que para el caso de los compresores cinepak, indeo 5.1 e intel 4.5 el resultado para las cinco secuencias es muy similar.

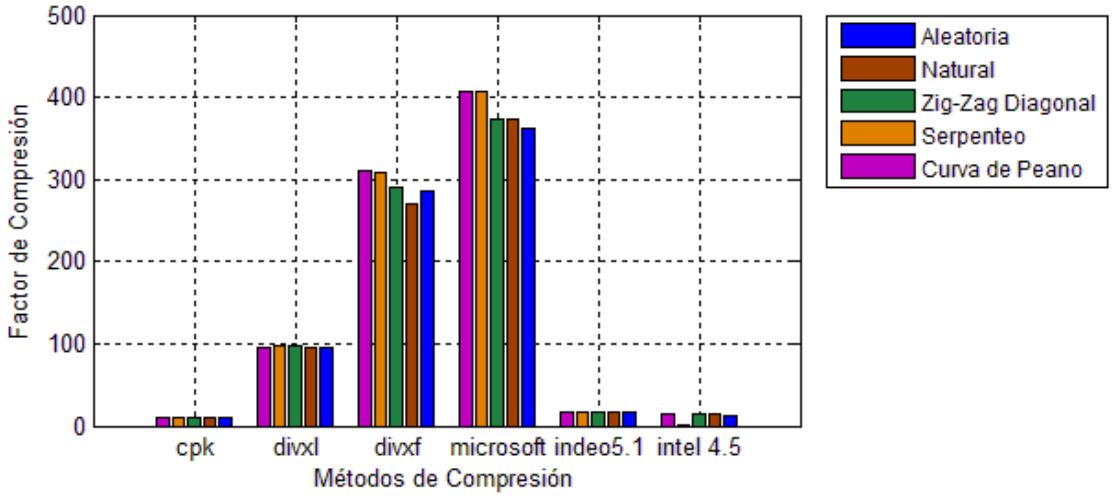


Figura 16. Factor de Compresión vs Métodos de Compresión

De la misma manera se realizó la comparación de SNR y el Factor de Compresión para la Región de Interés (RI) arrojando los siguientes resultados que se observan en las figuras 17 y 18.

En la figura 17 se evidencia que solo en cuatros secuencias en dos tipos de compresores no se cumple el valor mínimo para SNR de 20 dB, lo cual indica que la calidad del video resultante en estas secuencias no es la óptima.

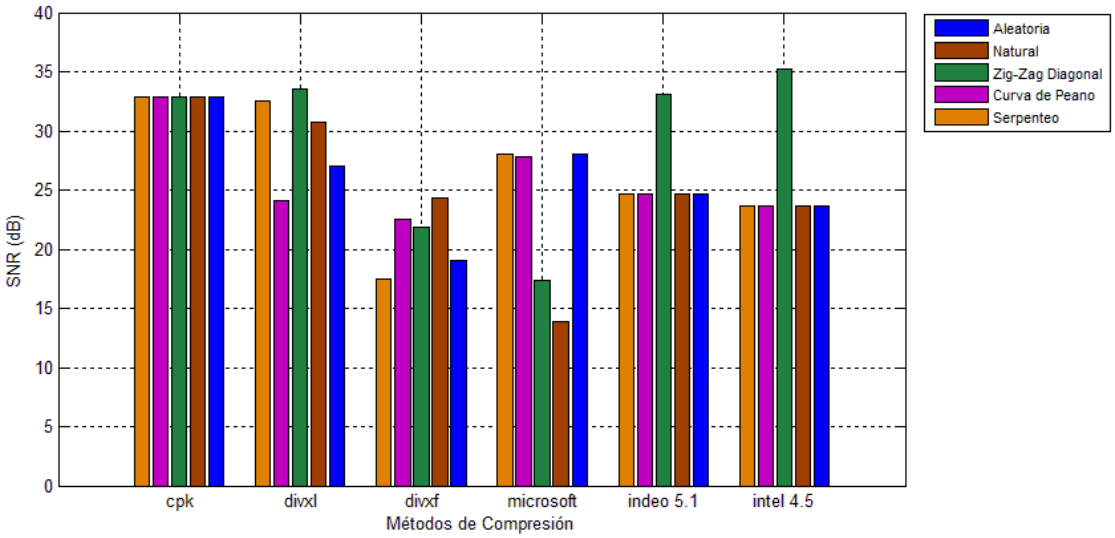


Figura 17. SNR y Métodos de Compresión

En la figura 18 se observa que el factor de compresión de las diferentes secuencias está relacionado directamente con el método de compresión y no con la secuencia, por lo cual el orden de las imágenes en cada una de las secuencias generadas no es relevante en este caso.

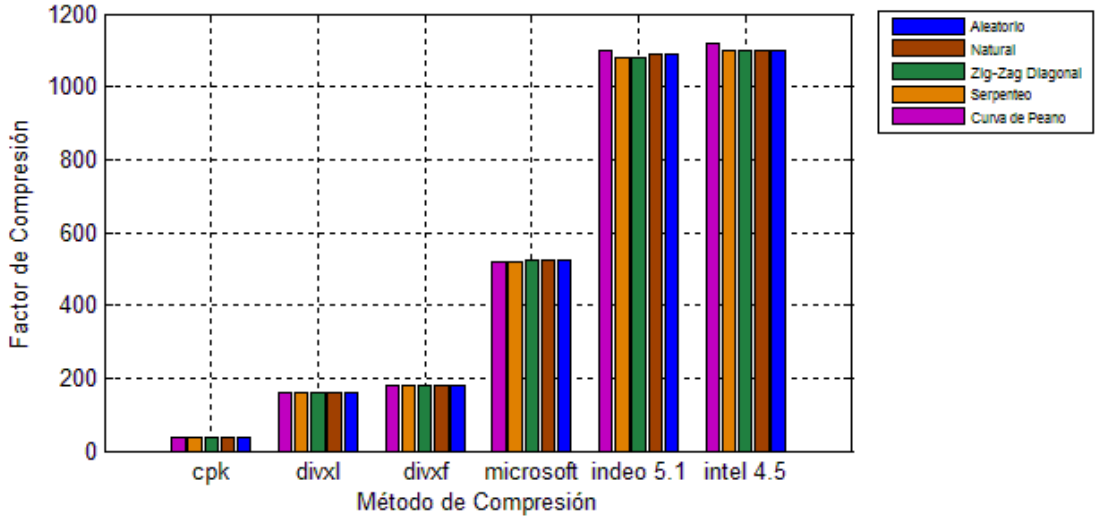


Figura 18. Factor de Compresión vs Métodos de Compresión en la RI

En la figura 19 se muestra el promedio de la desviación estándar de las imágenes diferencia lo cual evidencio que en efecto las secuencias generadas si algún tipo de

barrido como son la natural y la aleatoria tienen una mayor valor que las secuencias que tiene un barrido determinado.

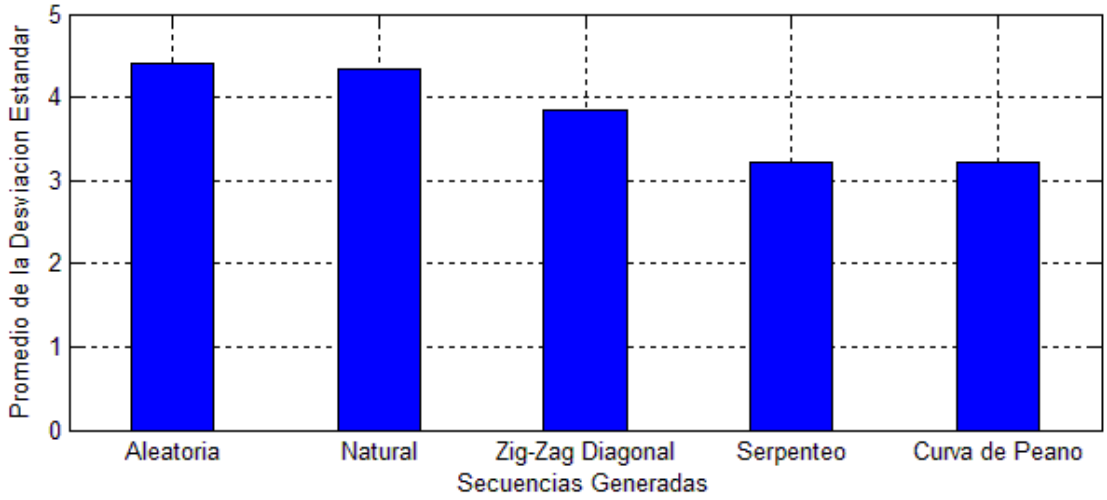


Figura 19. Promedio de la Desviación Estándar de las Imágenes Diferencia y las Secuencias Generadas

Se calculo el promedio de a desviacion estandar para la RI como se observa en la figura 20 la cual tiene la misma tendencia de la figura19.

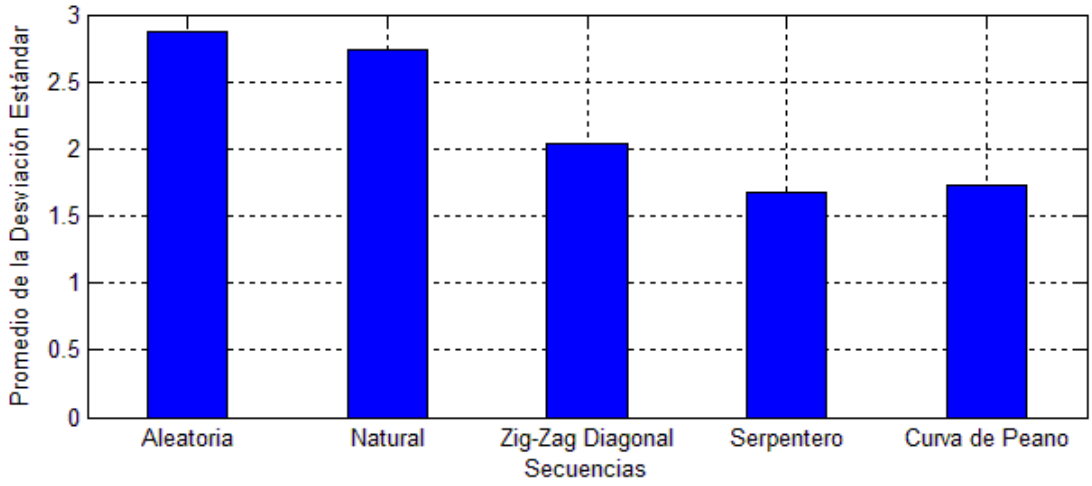


Figura 20. Promedio de la Desviación Estándar de la Región de la Interés de las imágenes diferencia

El cálculo de la Desviación Estándar de las imágenes diferencia se realizó hallando las 63 imágenes diferencia de las cuales se midió la distancia de separación de los datos de pixel a pixel de cada una de ellas y luego se calculó la desviación estándar.

Proceso que se describe en las siguientes ecuaciones:

$$R(X) = X_i\text{-}\mathrm{X}_{i+1} \quad [3.2]$$

Donde Xi es una matriz de 408 por 496 pixeles y Xi = (X1, X2,.., X64) son el numero de imágenes, adicionalmente $R(X)$ es el resultado de la resta punto a punto píxel a pixel resultante de cada par de imágenes la cual se denomino Imágen Diferencia (ID), en Matlab sería así: ID=imsubtract(Serpenteo(1,2).cdata,Serpenteo(1,3).cdata) siendo Serpenteo una de las secuencias generadas.

El cálculo de la desviación estándar de las Imágenes Diferencia se realizó utilizando la función de Matlab *std2* la cual tiene como característica principal realizar el cálculo de la desviación estándar de una matriz de elementos, la función se utilizó de la siguiente manera: DE=std2(ID), donde la ID es la imagen diferencia, así mismo se hallo la desviación estándar de las 63 imágenes diferencia de cada secuencia generada y luego se calculó la media de las 63 desviaciones estándar de la siguiente manera:

$$DESecuencia = mean(DE)$$

Donde DE es el vector de desviaciones estándar de las diferencias (63 elementos).

Los cálculos se realizaron para las secuencias de imágenes completas y para secuencias de la región de interés.

Los datos representados en las Figuras 19 y 20 se encuentran en la siguiente Tabla 1

Secuencias	Promedio de la Desviación Estándar Video Completo	Promedio de la Desviación Estándar Región de Interés
Base aleatoria	4,3388	2,8765
Natural	4,3954	2,6543
Zigzag Diagonal	3,8373	2,1002
Serpenteo	3,2228	1,6765
Curva de Peano	3,2105	1,7889

Tabla 1. Datos del Promedio de Desviación Estándar

4. ANÁLISIS DE RESULTADOS

Los resultados obtenidos en este trabajo de investigación plantean dos puntos importantes, la generación y compresión de secuencias de video con mínimos cambios entre imágenes de la base de datos de telefonía visual.

Generación de las Secuencias

El libro de código generado a partir de los mapas de Autoorganización de Kohonen (SOFM) se redujo de 495 MB a 37 MB, el cual contiene las imágenes con el aproximadamente el 96% de la información de la Base de Datos original.

La generación de las secuencias se realizó tomando como secuencia inicial la entregada por el entrenamiento realizado mediante los mapas de autoorganización la cual se tomó como secuencia natural y desde ahí se generaron las secuencias: aleatoria, zig-zag diagonal, serpenteo y curva de peano.

La clasificación de la imágenes de acuerdo a criterio de distancia, similitud libro de código generado por SOFM mostro que en la clasificación selecciono imágenes repetidas como fueron las que tiene los siguientes índices: 477 repetida 4 veces, 485 repetida 2 veces, 560 repetida 2 veces y 525 repetida 2 veces.

La generación de las tres secuencias con barrido determinado zig-zag diagonal, serpenteo y curva de peano organizo las imágenes de tal manera que se armaron videos con mínimos cambios entre imágenes, las dos secuencias sin un barrido determinado organizo las imágenes de forma desordenada en el caso de la aleatoria y buscando organizar las imágenes como estaba en la base de datos original (en el caso de la natural).

Compresión de las Secuencias

Los métodos de compresión que se emplearon durante la investigación mostraron que comprimen con un mínimo de 10 a 1 de las secuencias originales, pero no aprovechan la similitud entre imágenes.

El valor de la relación de señal a ruido (SNR) con respecto a los métodos de compresión aplicados a las diferentes secuencias generadas esta sobre el valor mínimo planteado durante el proyecto en todos los casos.

5. CONCLUSIONES

Las conclusiones en este trabajo de investigación plantean dos puntos importantes la generación y compresión de secuencias de las cuales se concluye:

A partir del libro de código que se obtuvo por SOFM se generaron 5 secuencias diferentes de las cuales las tres que tienen un barrido determinado formaron secuencias suaves lo cual se evidencio en las SOFM.

Los mapas de autoorganización SOFM solo verifica que sus vecinos no sean iguales, no verifica todas las imágenes. Esto es una conclusión bastante relevante dado que no se había evidenciado hasta este momento de la investigación.

La relación señal a ruido entre los diferentes compresores y las secuencias comprimidas da como resultado que los métodos de compresión no aprovechan la similitud entre imágenes, en algunos casos se observa que el valor del SNR entre una secuencia ordenada y una aleatoria es muy parecido.

Para proyectos futuros investigación se propone desarrollar y/o mejorar compresores de video existentes con el fin de aprovechar la similitud entre imágenes y de esta manera obtener el tamaño suficiente necesario para su almacenamiento.

BIBLIOGRAFIA

[1] Ayala O., Bárcenas E., Soto C., Carrillo R., Valderrama L.,Villegas J., Solano R. Vizcaya P., "Segmentación de Caracteristicas Faciales en Tiempo Real para Secuencias de Telefonía Visual," Pontificia Universidad Javeriana, Colombia, Manizales, Memorias del IX Simposio de Tratamiento de Señales, Imágenes y Visión Artificial Septiembre 2004.

[2] Santa D. Machado J. F., "Síntesis paramétrica de voz visual," Pontificia Universidad Javeriana, Bogóta, Colombia, Trabajo de Grado 2001.

[3] Sotomayor M. A., Vizcaya P. Baptiste C. A., "Segmentación y parametrización automática de rostros en video," Pontificia Universidad Javeriana, Bogotá, Colombia, Noviembre 2002.

[4] Vizcaya P. Soto C., "Generador de corpus para síntesis de voz visual," Pontificia Universidad Javeriana, Bogotá, Colombia, 2004.

[5] Ayala O., Bárcenas E., Soto C., Carrillo R., Valderrama L., Villegas J., Solano R. Vizcaya P., "Codificación y DEcodificación de Secuencias de Telefonía Visual," Pontificia Universidad Javeriana, Manizales, Memorias del IX Simposio de Tratamiento de Señales, Imágenes y Visión Artificial 2004.

[6] Internacional Telecommunicacion Union ITU, ITU-T Recommendation H.100 Visual Telephone Systems, 1998.

[7] Internacional Telecommunication Union ITU, Recommendation H.320 Narrow-ban Visual Telephone Systems and Terminal Equipment, 1999.

[8] I. Essa, "Analysis, interpretation and synthesis of facial expressions PhD thesis," MIT Department of Media Arts and Sciences, 1995.

[9] Allen W.G. Mark M. W., *Lip- Motion analysis for speech segmentation in noise*., 1994, vol. 14.

[10] Galán J., Soto C., Urbina J., y Vásquez S. Bárcenas E., "Análisis y síntesis de voz visual en el idionma español," Pontificia Universidad Javeriana, Bogotá, Trabajo de grado 2001.

[11] Quijano A. Vizcaya P., "Seleccion de Imágenes para síntesis de voz visual empleando SOFM," Pontificia Universidad Javeriana, Bogotá, 2006.

[12] Kohonen T., *Selt-organization and associative memory*. Helsinkin, Finlandia: Springer series in information sciences, 1984.

[13] Haykin S., *Neural networks a comprehensive foundation*. Englewood Cliffs, E. U.:

Macmillian publishing, 1994.

[14] Paniagua A., *Adobe Premiere Pro*.: Anaya, 2006.

ANEXOS

Anexo A: Las tres tablas con todos los datos obtenidos durante la investigación