



**ESPE**  
UNIVERSIDAD DE LAS FUERZAS ARMADAS  
INNOVACIÓN PARA LA EXCELENCIA

**VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y  
TRANSFERENCIA DE TECNOLOGÍA**

**CENTRO DE POSGRADOS**

**MAESTRÍA EN ENSEÑANZA DE LA MATEMÁTICA**

**TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DE TÍTULO  
DE MAGISTER EN ENSEÑANZA DE LA MATEMÁTICA**

**TEMA: ANÁLISIS EXPLORATORIO DEL COMPORTAMIENTO DE  
DATOS CAMBIANTES EN EL TIEMPO USANDO TÓPICOS**

**AVANZADOS DE ÁLGEBRA LINEAL**

**AUTOR: OÑA ROCHA, OMAR RICARDO**

**DIRECTOR: PhD. PELUFFO ORDOÑEZ, DIEGO HERNÁN**

**SANGOLQUÍ**

**2019**



VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y  
TRANSFERENCIA DE TECNOLOGÍA  
CENTRO DE POSGRADOS

**CERTIFICACIÓN**

Certifico que el trabajo de titulación, “*ANÁLISIS EXPLORATORIO DEL COMPORTAMIENTO DE DATOS CAMBIANTES EN EL TIEMPO USANDO TÓPICOS AVANZADOS DE ÁLGEBRA LINEAL*” fue realizado por el señor *Oña Rocha Omar Ricardo* el mismo que ha sido revisado en su totalidad, analizado por la herramienta de verificación de similitud de contenido; por lo tanto cumple con los requisitos teóricos, científicos, técnicos, metodológicos y legales establecidos por la Universidad de Fuerzas Armadas ESPE, razón por la cual me permito acreditar y autorizar para que lo sustente públicamente.

Sangolquí, junio 2019

---

PhD Diego Hernán Peluffo Ordoñez

C.C.: 1757278443



**VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y  
TRANSFERENCIA DE TECNOLOGÍA  
CENTRO DE POSGRADOS  
AUTORÍA DE RESPONSABILIDAD**

Yo, *Oña Rocha Omar Ricardo*, con cédula de ciudadanía n° 1001615473 declaro que el contenido, ideas y criterios del trabajo de titulación: *“Análisis exploratorio del comportamiento de datos cambiantes en el tiempo usando tópicos avanzados de álgebra lineal”* es de mi autoría y responsabilidad, cumpliendo con los requisitos teóricos, científicos, técnicos, metodológicos y legales establecidos por la Universidad de Fuerzas Armadas ESPE, respetando los derechos intelectuales de terceros y referenciando las citas bibliográficas.

Consecuentemente el contenido de la investigación mencionada es veraz.

Sangolquí, junio 2019

---

Omar Ricardo Oña Rocha  
C.C.: 1001615473



**VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y  
TRANSFERENCIA DE TECNOLOGÍA  
CENTRO DE POSGRADOS**

**AUTORIZACIÓN**

*Yo, **Oña Rocha Omar Ricardo** autorizo a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de titulación: “Análisis exploratorio del comportamiento de datos cambiantes en el tiempo usando tópicos avanzados de álgebra lineal” en el Repositorio Institucional, cuyo contenido, ideas y criterios son de mi responsabilidad.*

Sangolquí, junio 2019

---

**Omar Ricardo Oña Rocha**  
C.C.: 1001615473

## **DEDICATORIA**

A mis padres, hijos y nietos que con su inocencia, alegría, amor y comprensión estuvieron en todo momento presentes.

## **AGRADECIMIENTO**

En este momento de mi vida al culminar una etapa más de estudio doy gracias a mi papito Dios por brindarme la vida completa y permitir que llegue hasta este logro, además agradezco a mi gran amigo, compañero y mentor Diego; una gran persona y calidad humana inconmensurable, por su paciencia, esfuerzo y perseverancia para que este trabajo llegue a feliz término. De igual manera a todas las personas que de una u otra manera contribuyeron durante el tiempo de estudio y elaboración de esta tesis con mucho cariño y afecto.

**INDICE DE CONTENIDOS**

<b>CAPÍTULO I</b> .....	15
<b>INTRODUCCIÓN</b> .....	15
1.1 Contexto y antecedentes.....	15
1.2 Planteamiento del problema.....	18
1.3 Objetivos .....	19
1.3.1 Objetivo general .....	19
1.3.2 Objetivos específicos .....	20
1.4 Contribuciones de la tesis .....	20
1.5 Organización del documento.....	21
<b>CAPITULO II</b> .....	22
<b>ESTADO DEL ARTE DE ANALISIS DE DATOS DINÁMICOS</b> .....	22
2.1 Análisis de datos dinámicos usando técnicas de álgebra lineal .....	22
2.2 Análisis de datos dinámicos usando agrupamiento espectral .....	22
2.2.1 Modificación de algoritmos .....	23
2.2.2 Agrupamiento basado en kernel .....	28
<b>CAPITULO III</b> .....	32
<b>MARCO TEÓRICO</b> .....	32
3.1 Introducción .....	32

3.2 Elementos de algebra lineal .....	32
3.2.1 Matriz .....	32
3.2.2 Orden de una matriz.....	32
3.2.3 Matrices fila y columna.....	33
3.2.4 Matriz cuadrada .....	33
3.2.5 Matriz diagonal.....	34
3.2.6 Matriz identidad.....	34
3.2.7 Matriz triangular .....	34
3.2.8 Matriz escalonada .....	35
3.2.10 Matrices simétrica y antisimétrica .....	36
3.2.11 Matriz ortogonal .....	36
3.2.12 Traza de una matriz.....	36
3.2.13 Inversa de una matriz .....	37
3.2.14 Independencia lineal .....	37
3.2.15 Rango de una matriz .....	38
3.2.16 Espacio vectorial.....	38
3.2.17 Subespacio vectorial .....	39
3.2.18 Combinación lineal .....	39
3.2.19 Operador matemático.....	39

3.2.20 Producto Interno .....	39
3.2.21 Norma.....	40
3.2.21.1 Normas de vectores.....	41
3.2.21.2 Normas inducidas de matrices.....	41
3.2.21.3 Norma de Frobenius.....	42
3.2.22 Estimación de valores y vectores propios.....	42
3.2.23 Introducción a la optimización.....	43
3.2.23.1 Multiplicadores de Lagrange.....	43
3.2.23.2 Funciones reales cuadráticas.....	47
3.3 Agrupamiento espectral.....	47
3.4 Teoría de grafos.....	49
3.4 Función y matriz kernel.....	50
3.4.1 Función kernel.....	50
3.4.2 Matriz kernel.....	51
3.4.3 Tipos de funciones kernel.....	52
3.5 Formulación y resolución del problema de optimización.....	53
CAPITULO IV.....	58
MODELO PROPUESTO PARA EL ANÁLISIS DE DATOS DINÁMICOS.....	58
4.1 Introducción.....	58

4.2 Datos cambiantes en el tiempo.....	59
4.3 Matrices kernel acumuladas (agrupamiento espectral de kernel).....	59
4.4 Vector de seguimiento .....	60
4.5 Identificación de segmentos de movimiento.....	64
CAPITULO V .....	66
RESULTADOS EXPERIMENTALES Y DISCUSIÓN.....	66
5.1 Introducción .....	66
5.2 Marco experimental.....	66
5.2.1 Bases de datos para realizar las pruebas .....	66
5.2.1.1 Datos sintéticos: Curvas de nivel.....	66
5.2.1.2 Datos reales: Objetos girando sobre su propio eje.....	67
5.2.1.3 Aplicación del método KSC en el caso: Salto vertical de humanos.....	68
5.2.2 Métodos de comparación .....	69
5.2.3 Medidas de desempeño .....	70
5.3 Resultados .....	71
5.3.1 Datos sintéticos: Curvas de nivel.....	71
5.3.2 Giro de objeto 360 grados.....	75
5.3.3 Aplicación del método KSC en el caso: Salto vertical de humanos.....	79
5.4 Discusión.....	80

CAPITULO VI.....	81
CONCLUSIONES Y TRABAJO FUTURO .....	81
6.1 Conclusiones .....	81
6.2 Trabajo futuro.....	83
BIBLIOGRAFIA.....	84

**INDICE DE TABLAS**

<b>Tabla 1</b> <i>Ejemplo de funciones Kernel</i> .....	53
<b>Tabla 2</b> <i>Rendimiento de agrupamiento Sujeto #2 en términos de medidas métricas</i> .....	71

## INDICE DE FIGURAS

<b>Figura 1.</b> Ilustración de Karush-Kuhn-Tucker para el caso de una restricción de igualdad y dos variables .....	45
<b>Figura 2.</b> Grafo ponderado.....	50
<b>Figura 3.</b> Mapeo de alta dimensión de espacio de entrada X usando la función $\phi$ .....	52
<b>Figura 4.</b> Explicación gráfica de la segmentación a partir del vector de seguimiento original. ...	65
<b>Figura 5.</b> 2-D curva de movimiento .....	67
<b>Figura 6.</b> Base de datos Coil20 .....	67
<b>Figura 7.</b> Los cuadros del objeto #4 de la base de datos COIL20.....	68
<b>Figura 8.</b> Sujeto # 2 de la Base de Datos captura de movimientos .....	69
<b>Figura 9.</b> Agrupamiento de la curva de movimiento 2-D en agrupamientos $K = 3$ con cuadros $N_f = 100$ y $N = 100$ muestras por cuadro .....	73
<b>Figura 10.</b> Agrupamiento de la curva de movimiento 2-D en agrupamientos $K$ con cuadros $N_f = 100$ y $N = 100$ muestras por cuadro.....	74
<b>Figura 11.</b> Trazado de los vectores original y codificado.....	76
<b>Figura 12.</b> Objeto #4 rastreo de cuadros originales (5, 15, 20, 30, 35) y rastreo de vectores.....	77
<b>Figura 13.</b> Objeto #4 rastreo de cuadros originales (40, 50, 60, 65, 70) y rastreo de vectores.....	78
<b>Figura 14.</b> Cuadro 71 y representación de resultados finales .....	78
<b>Figura 15.</b> Seguimiento del Sujeto #2 .....	79

## RESUMEN

En la actualidad, el análisis de datos dinámicos o variantes en el tiempo es de altísimo interés en la ciencia y la tecnología, y es de gran utilidad en diversas aplicaciones, tales como: predicción, análisis de vídeo, segmentación automática de movimiento, entre otras. Particularmente, las técnicas de reconocimiento de patrones, en especial, aquellas basadas en análisis espectral y álgebra de matrices han mostrado ser una buena alternativa. Sin embargo, aún existe un amplio abanico de problemas sin resolver relacionados con la precisión y la interpretación de los segmentos de movimiento. Esta tesis de maestría presenta un estudio sobre el uso en análisis de datos cambiantes en el tiempo de una técnica de reconocimiento de patrones no supervisada, denominada agrupamiento espectral basado en *kernels* (*kernel spectral clustering*, en inglés). Específicamente, se estudia la posibilidad de construir un vector de seguimiento que se encarga de segmentar automáticamente movimientos en una secuencia de cuadros de un vídeo, el cual comprueba su utilidad en la identificación de inicio y fin de movimientos en objetos rotativos y curvas de nivel en movimiento. Con el desarrollo de este trabajo se comprueba también el beneficio de las propiedades, la optimización y el álgebra de funciones con matrices para el análisis de datos dinámicos.

### **PALABRAS CLAVE:**

- **AGRUPAMIENTO ESPECTRAL BASADO EN KERNELS**
- **DATOS DINÁMICOS**
- **KERNELS**

## **ABSTRACT**

Nowadays, the analysis of dynamic or time-varying data is of great interest in science and technology, as it is very useful in several applications, such as: time-series forecasting, video analysis, automatic movement segmentation, among others. In particular, pattern recognition techniques, especially those based on spectral analysis and matrix algebra, have proven to be a suitable alternative. Notwithstanding, there is still a wide range of open issues related to the accuracy and interpretation of the movement segments. This master's thesis presents a study on the use -in time-varying data analysis- of an unsupervised pattern recognition technique, so-called kernel spectral clustering. Specifically, it is of interest the possibility of creating a tracking vector able to automatically segmenting movements in a sequence of video frames, which shows its usefulness in the identification of starting and ending points of movements in rotating objects and moving-contour-lines. Also, this work proves the benefit of the properties, the optimization and the algebra of functions with matrices for the analysis of dynamic data.

### **KEYWORDS:**

- **KERNEL SPECTRAL CLUSTERING**
- **DYNAMIC DATA**
- **KERNELS**

# CAPÍTULO I

## INTRODUCCIÓN

### 1.1 Contexto y antecedentes

Actualmente, el análisis de datos dinámicos o variantes en el tiempo es un tema que toma cada vez mayor interés en la comunidad científica, especialmente, en las áreas de reconocimiento de patrones y aprendizaje de máquina (*Machine Learning*). Existen diversas aplicaciones en donde el análisis de datos dinámicos ha mostrado ser una buena alternativa, tales como: el análisis de video (Arróspide & Salgado, 2015 ; Lu, Zhang, Xu, & Shang, 2015) , al igual que la identificación de movimiento para la supervisión (Yadav, Dubey, & Ahmed, 2014) y la clasificación del movimiento de cuerpos (Saripalle et al., 2014) algunas de sus aplicaciones más notables. En este contexto surgieron los enfoques de análisis con base en un kernel y de matriz espectral como alternativas apropiadas

Una de las alternativas matemáticas para desarrollar métodos y modelos para procesar datos dinámicos es la aplicación de tópicos avanzados de álgebra lineal, específicamente, el análisis matricial de tipo espectral (es decir, análisis que implica cálculo de valores y vectores propios de una matriz cuadrada de gran cantidad de elementos). Las técnicas espectrales, principalmente aquellas basadas en matrices *kernel*, han demostrado su alta aplicabilidad en diversos aspectos del reconocimiento de patrones y aprendizaje de máquina, incluso cuando los datos son variantes en el tiempo, tales como la estimación del número de grupos o eventos, agrupamiento y clasificación de datos.

Por ejemplo, (Shanmao, Yunlong, Lijun, & Ni, 2015) propone un esquema continuo de análisis de componentes principales del kernel ponderado (*WKPCA - weighted kernel principal component analysis*) que se encuentre en capacidad de capturar el comportamiento dinámico de los datos. Otro estudio (Langone, Alzate, & Suykens, 2013) toma ventajas de propiedades de agrupamiento espectral para explorar la naturaleza variable en el tiempo mediante la adición de Efecto de memoria en un cuadro de agrupación espectral del kernel (KSC) (C. Alzate & Suykens, 2010) .

La mayoría de los métodos espectrales han sido diseñados para el análisis de datos estáticos, descartando la información temporal, es decir, omitiendo el comportamiento y la evolución de los datos a lo largo del tiempo. La mayor parte de los trabajos actuales consideran el efecto de la variación en el tiempo, sin embargo, el diseño de un método que permita seguir la dinámica de los datos y agrupar los mismos en ambientes de tiempo real, con alta fidelidad y precisión, es aún un problema abierto. Por tanto, la segmentación del movimiento ha estado tradicionalmente ligada con la detección de movimiento, en la que cada región corresponde a un modelo particular de movimiento que explica los cambios temporales en dicha región. También se han realizado diferentes propuestas para segmentación de movimiento basadas en clustering. Sin embargo, en cualquiera de los casos anteriores es necesario tener un conocimiento a priori del número de objetos en movimiento presentes en la escena, lo cual no siempre es posible. En la mayoría de los trabajos sobre el tema, proponen un algoritmo de segmentación de objetos con modelos de movimiento afín basado en la técnica matemática GPCA (Generalized Principal Components Analysis) que permite segmentar objetos en movimiento sin conocimiento previo del número de objetos presentes en la escena. Hay que indicar que esta técnica matemática está basada en el álgebra lineal, pero en la literatura también se le conoce por PCA generalizado.

Se tiene conocimiento de que tanto a nivel internacional y nacional existen grupos de investigación que desarrollan investigaciones en torno a estos temas y entre ellos tenemos:

En el ámbito internacional el grupo MLG (Machine Learning Group) de la Universidad Católica de Louvain, investiga en áreas relacionadas con el machine learning tales como: Reducción de dimensión, análisis multilabel, datos ruidosos, modelos estadísticos y kernel, clustering, clasificación supervisada, predicción en series de tiempo, minería de datos, predicción de información financieros, análisis de imágenes, control de calidad de productos, entre otros. Sitio web <https://mlg.info.ucl.ac.be/Main/HomePage>.

También se encuentran trabajos realizados por grupos de investigación como Grupo de Control y Procesamiento Digital de Señales GCPDS de la Universidad Nacional de Colombia, sede Manizales. Abarca líneas de investigación en procesamiento de audio, imágenes y vídeo, reconocimiento de patrones, incluyendo asuntos de representación, reducción de dimensión y generalización, también temas emergentes como balance de clases y análisis con múltiples expertos; análisis de movimiento a partir de técnicas de reducción de dimensión no lineal; Otras aplicaciones como meteorología y dinámica de suelos.

Sitio web <http://gta.manizales.unal.edu.co/signal/investigacion.php>

Grupo GRIAS, Universidad Nariño, investiga en técnicas de descubrimiento de conocimiento en bases de datos, visualización geo referenciada y desarrollo de software. Sus líneas de investigación son: minería de datos, desarrollo de software.

Sitio web: [http://sistemas.udenar.edu.co/?page\\_id=101](http://sistemas.udenar.edu.co/?page_id=101)

En el ámbito nacional se cuenta con investigaciones realizadas en diferentes universidades, tales como:

El Grupo de Sistemas de Análisis de Datos Inteligentes (Grupo SDAS) fue fundado por el Prof. Diego Peluffo en 2018. Cuenta con el apoyo de expertos y jóvenes investigadores de varias instituciones académicas de Ecuador, Colombia, Brasil, Argentina y España. El Grupo de Investigación SDAS se ocupa de los problemas relacionados con el análisis de datos al proporcionar soluciones basadas en software y sistemas integrados. El ámbito de sus aplicaciones de interés abarca: análisis de datos complejos, variables en el tiempo, métodos formales de aprendizaje automático, redes de sensores, enfoques de bajo costo computacional para el análisis de datos y visualización de datos, entre otros. Además, el Grupo SDAS ofrece servicios de análisis de datos, redes de sensores y diseño de sistemas integrados, diseño asistido por computadora de plantas de producción, sistemas de soporte de diagnóstico médico computarizados, procesamiento de imágenes, desarrollo de software, diseño eléctrico, diseño web, sistemas domóticos y biométricos, y aplicaciones Matemáticas.

Sitio web: <https://sdas-group.com/>

Desarrollo del software de integración del proyecto Sistema de Análisis de Movimiento (SAM) de la Universidad Politécnica Salesiana sede Cuenca, Diseño e implementación de un sistema para el análisis de movimiento corporal humano de la Universidad de las Fuerzas Armadas,

## **1.2 Planteamiento del problema**

En general, el agrupamiento no supervisado o clustering ha mostrado ser útil en el análisis exploratorio de los datos y muy versátil. En particular, en el contexto del agrupamiento espectral, se han desarrollado diferentes métodos de agrupamiento que atienden a diversos problemas como sensibilidad a los parámetros iniciales (Zelnik-Manor & Perona, 2005), convergencia a un óptimo

local distante del global (Meila & Shi, 2001b), (Meila & Shi, 2001a), clases difícilmente separables (Stella & Shi, 2003), entre otros. Sin embargo, la selección de un método no es una tarea trivial, es necesario considerar la naturaleza de los datos y las condiciones del problema con el fin de agrupar patrones homogéneos o similares, de tal forma que se realice un efectivo reconocimiento de las clases.

Entre los principales problemas abiertos que presenta el agrupamiento espectral, se encuentra que: Los métodos de agrupamiento espectral fueron concebidos para la clasificación de datos estáticos con clases difícilmente separables por métodos tradicionales. Entonces, hasta el momento no se han desarrollado métodos dinámicos que consideren el cambio de los datos a través del tiempo. Recientemente, se han desarrollado algunos trabajos teniendo en cuenta la información temporal en el agrupamiento (Takács, Butler, & Demiris, 2007), (Zhou, De la Torre, & Hodgins, 2008), pero en general no existe un método de agrupamiento estrictamente espectral basado en kernel que analice datos dinámicos.

En suma, no existe una metodología de agrupamiento espectral basada en kernel que contemple la información temporal de los datos.

## **1.3 Objetivos**

### **1.3.1 Objetivo general**

Desarrollar un análisis exploratorio sobre información de datos variantes en el tiempo, usando matrices kernel y agrupamiento de tipo espectral, con el fin de proponer un modelo de segmentación automática de movimientos en una secuencia de cuadros con un patrón de giro.

### 1.3.2 Objetivos específicos

- Establecer el método de agrupamiento espectral para ser extendido a análisis de datos dinámicos considerando un enfoque de matrices kernel.
- Modelizar la dinámica de los datos variantes en el tiempo a partir de formulaciones con matrices kernel acumuladas capaz de segmentar automáticamente movimientos en vídeos que incluyen objetos girando 360 grados.
- Demostrar el uso de fundamentos de matrices y funciones kernel en la enseñanza de los conceptos, resolución de problemas de optimización de variable latente, y derivadas parciales de funciones matriciales a partir de simulaciones y resultados teóricos.

### 1.4 Contribuciones de la tesis

Dado que tradicionalmente, el análisis espectral de datos ha sido desarrollado para escenarios de datos estáticos, existe la necesidad científica y tecnológica para extender este tipo de análisis al uso de datos variantes en el tiempo. En este sentido, los métodos basados en matrices kernel representan una buena alternativa.

Este trabajo tiene como motivación, obtener un resultado que represente un aporte a la comunidad académica y científica además de servir como una herramienta para la enseñanza de conceptos de álgebra lineal esto es, resolución de problemas matriciales de optimización de variable latente, y derivadas parciales de funciones matriciales, asimismo, se cuantificará la relación de contribución de cada columna de una matriz al cálculo de la norma de ésta.

Específicamente, a continuación, se mencionan los alcances teóricos y metodológicos esperados de este trabajo.

- Metodología de segmentación automática de movimientos: Se desarrollará una metodología de segmentación automática de movimientos basada en agrupamiento espectral y matrices kernel.
- Estrategia de enseñanza de conceptos de álgebra lineal: Es una estrategia para comprender el concepto de norma y traza a partir de la cuantificación de cuánto aporta cada columna de la matriz en la minimización de una función de coste o función objetivo que se planteará para resolver el problema de segmentación de movimiento.

### **1.5 Organización del documento**

El presente documento está estructurado por seis capítulos, en el Capítulo I se establece antecedentes, problema, objetivos y aporte de la tesis. En el capítulo II: Estado del arte que contempla análisis de datos dinámicos, agrupamiento de datos dinámicos, agrupamiento espectral tradicional y agrupamiento espectral usando kernel. El capítulo III: Marco Teórico, donde se describe los elementos básicos del algebra lineal, agrupamiento espectral, teoría de grafos, funciones kernel y matrices kernel. Capitulo IV: Modelo Propuesto, en el que se establece los datos cambiantes en el tiempo, matrices kernel acumuladas, vector de seguimiento y la identificación de movimiento de objetos. El capítulo V: Resultados y discusión, allí se tiene el marco experimental donde se describe las bases de datos que se utilizan para los experimentos: curvas de nivel, objeto girando 360 y humano en salto vertical; de igual manera los resultados de los tres experimentos antes mencionadas. Finalmente, el capítulo VI se detalla las conclusiones del trabajo realizado y futuras investigaciones. En apéndices se presenta el Anexo A con el código de programación realizado, Anexo B se muestra la presentación del sitio web y la dirección electrónica del mismo y en el Anexo C las publicaciones realizas respecto al tema. Y serrando el documento está la bibliografía utilizada.

## CAPITULO II

### ESTADO DEL ARTE DE ANALISIS DE DATOS DINÁMICOS

#### 2.1 Análisis de datos dinámicos usando técnicas de álgebra lineal

En general, los enfoques de agrupamiento de tipo espectral, particional y basados en grafos son representados a partir de formulaciones matrices y vectoriales, por tanto, el álgebra lineal es una herramienta fundamental.

Recientemente se han realizado enfoque basado en agrupamiento difuso en varias ocasiones requiere lidiar con la complejidad de la información tanto espacial como temporal. En particular, se debe considerar la naturaleza espacial de las unidades a agrupar, las características del espacio de las trayectorias de tiempo multivariadas (D'Urso, De Giovanni, Disegna, & Massari, 2019).

La detección de patrones en series de tiempo multivariadas es una tarea relevante, especialmente para grandes conjuntos de datos. Para hacer frente a la complejidad de las características de cada serie de tiempo multivariable y la incertidumbre de asignación asociada, se adopta un enfoque de agrupamiento difuso (D'Urso, De Giovanni, & Massari, 2018).

No obstante, estos enfoques carecen de interpretabilidad matemática y versatilidad como la que si poseen los métodos espectrales y basados en kernel.

#### 2.2 Análisis de datos dinámicos usando agrupamiento espectral

Al igual que cualquier técnica no supervisada, el agrupamiento de tipo espectral, es un método discriminativo, es decir, no requiere de conocimientos previos sobre las clases para su clasificación. Como tal, realiza el proceso de agrupación utilizando sólo la información dentro de los datos y, en general, algunos parámetros de inicio tales como el número de grupos resultantes o cualquier otra pista sobre la partición inicial.

En este caso, dado que la agrupación espectral está respaldada por la teoría de los gráficos, los parámetros de entrada para los algoritmos de agrupamiento son el número de grupos y la matriz de similitud. La inicialización es una etapa importante en los métodos no supervisados ya que la mayoría de ellos son sensibles a los parámetros de partida; esto significa que, si los parámetros iniciales no son adecuados, la convergencia del algoritmo puede fallar al caer sobre una solución subóptima bastante distante del óptimo global. Por otro lado, como se mencionó anteriormente, los métodos no supervisados llevan a cabo el agrupamiento con la información directa dada por los datos. Por lo tanto, si el espacio de representación inicial no es suficientemente discriminatorio bajo los criterios de agrupamiento, en algunos casos pueden requerirse etapas de extracción y / o selección de características.

### **2.2.1 Modificación de algoritmos**

A continuación, se presenta los trabajos relacionados con la temática de interés, organizados en secciones así: los enfoques de análisis de datos dinámicos, agrupamiento basado en kernels y agrupamiento espectral tradicional.

#### **Enfoques para el análisis dinámico**

Existen enfoques que incluyen el diseño de kernels dinámicos para la agrupación ((A. B. Chan & Vasconcelos, 2005), (Keshet & Bengio, 2009) o un modelo basado en el análisis de componentes principales del kernel dinámico (KPCA) (Maestri, Cassanello, & Horowitz, 2009), (Choi & Lee, 2004). Otro estudio (Langone et al., 2013) modifica la función primaria de una formulación de KPCA para la agrupación espectral para agregar el efecto memoria.

Hay otra alternativa conocida como Multiple Kernel Learning (MKL), que ha surgido para tratar diferentes problemas en el aprendizaje automático, principalmente, con respecto a las

máquinas de vectores de soporte (SVM) (González, Bermeo, Ramos, & Nasraoui, 2012), (Huang, Chuang, & Chen, 2012). La idea intuitiva de MKL es que el aprendizaje se puede mejorar cuando se usan kernels diferentes en lugar de un kernel único. Por tanto, el análisis local proporcionado por cada kernel es beneficioso para examinar la estructura de los datos completos cuando se tienen clusters complejos locales.

La agrupación basada en la teoría espectral es un enfoque relativamente nuevo para el análisis sin supervisión; aunque se ha utilizado en varios estudios que demuestran su eficacia en las tareas de agrupación, especialmente en los casos en que los conglomerados no son linealmente separables. Esta técnica de agrupamiento ha sido ampliamente utilizada en una gran cantidad de aplicaciones como el diseño de circuitos (Jianbo Shi & Malik, 2000), el equilibrio de carga computacional para la computación intensiva, el análisis del movimiento humano y la identificación de personas (P. K. Chan, Schlag, & Zien, 1994), (Hendrickson & Leland, 1995), la segmentación de imágenes (Ishikawa & Geiger, 1998), (Stella & Shi, 2003) y el análisis de vídeo (Chung, 1997), (Ng, Jordan, & Weiss, 2002a), entre otros. Lo que hace que el análisis espectral aplicado al agrupamiento de datos sea atractivo, es el uso del valor propio y de la descomposición del vector propio para obtener el óptimo local más cercana a la óptima continua global. Las técnicas de agrupamiento espectral se aprovechan de la topología de los datos de una representación gráfica no dirigida y ponderada. Este enfoque se conoce como partición de grafos; en él un problema inicial de la optimización se plantea generalmente formulado bajo algunas restricciones. A menudo, esta formulación es ligera y luego se convierte en un problema NP-completo (Meila & Shi, 2001a), (Ham, Lee, Mika, & Schölkopf, 2004). Además, la estimación de la óptima global en un dominio continuo se realiza a través de la descomposición del vector propio

(Hagen & Kahng, 1992), basada en el teorema de Perron-Frobenius, que establece que los valores propios estrictamente reales más grandes asociados a un valor positivo y una matriz irreducible para determinar su propia relación espectral (Wolf & Bileschi, 2005). En este caso, dicha matriz corresponde a la matriz de afinidad o similitud que contiene las similitudes entre los puntos de datos. En otras palabras, el espacio creado por los vectores propios en la matriz de afinidad está estrechamente relacionado con la calidad de la agrupación.

### **Agrupamiento datos dinámicos**

Los algunos enfoques más avanzados proponen minimizar una medida de factibilidad entre las soluciones del problema sin restricciones y el indicador de agrupamiento permitido (Zhou et al., 2008),(A. B. Chan & Vasconcelos, 2005), a través de la búsqueda de picos y valles de una función de coste que cuantifica el grupo de superposición (Keshet & Bengio, 2009). Después de dicha búsqueda, se lleva a cabo un proceso de discretización. El problema de la discretización de la solución continua se formula y se resuelve en (Ng, Jordan, & Weiss, 2002b). Existen métodos basados en grafos asociados con el Laplaciano normalizado y el no normalizado (el cual se hace sobre una matriz de similitud, también denominada matriz kernel), los cuales proporcionan información relevante contenida en el Laplaciano de los vectores propios y ésta, a su vez, es de gran utilidad para la tarea de agrupamiento. Además, a menudo, proporcionan una interpretación directa sobre la calidad del agrupamiento basada en una teoría de grafos aleatorios.

Es de gran importancia el análisis dinámico de datos, existen varias aplicaciones de gran interés tales como análisis de video (Shirazi, Harandi, Sanderson, Alavi, & Lovell, 2012) e identificación de movimiento (Sudha & Bhavani, 2012), entre otras. Haciendo uso de análisis espectral y la agrupación, algunos trabajos se han desarrollado teniendo en cuenta la información temporal para

la tarea de agrupación, principalmente en la segmentación del movimiento humano (Takács et al., 2007), (Zhou et al., 2008).

No obstante, el diseño de un método de análisis no supervisado basado en matrices kernel y capaz de seguir la dinámica de datos cambiantes en el tiempo es aún un problema abierto.

### **Agrupamiento espectral tradicional**

Bajo el principio fundamental del agrupamiento espectral, y, considerando la posibilidad de obtener una solución discreta a través de vectores propios, se han propuesto enfoques de múltiples agrupamientos como el método de corte normalizado k-way introducido en (Hagen & Kahng, 1992). En dicho método, para obtener la solución discreta, debe resolverse otro problema de optimización, aunque en un dominio de dimensión inferior. Además, los vectores propios líderes pueden considerarse como un nuevo conjunto de datos que pueden agruparse mediante un algoritmo de agrupamiento convencional, como K-means (Carlos Alzate & Suykens, 2006), (Meila & Shi, 2001a). A continuación, se logra un algoritmo híbrido en el que los algoritmos de análisis espectral pueden generar el espacio de vectores propios y establecer parámetros iniciales adecuados, mientras que los algoritmos de partición estarían a cargo del propio agrupamiento.

Esto significa que el análisis espectral puede soportar las técnicas convencionales que generan una inicialización que mejora la convergencia del algoritmo. Sin embargo, similar a los valores de membresía dura de algunos algoritmos de agrupación basada en el centro (Higham & Kibble, 2004), un esquema de agrupación espectral puede agrupar datos solos de salida de una matriz binaria de agrupaciones que denota el punto de datos de pertenencia a cada grupo. Para llevar a cabo este enfoque de agrupamiento, no sólo son los vectores propios necesarios, sino también es la transformación ortonormal de la solución del vector propio (solución propia). Esto lleva a la

meta de agrupamiento espectral que es el hallazgo de la mejor transformación ortonormal que genera una discretización adecuada de la solución continua (Carlos Alzate & Suykens, 2006).

En general, cuando se trata de problemas bi-cluster, la solución del problema relajado es a menudo un vector propio específico (Takács et al., 2007). Sin embargo, cuando el problema se encuentra dentro de una clase de cuadro de varias agrupaciones, la determinación de la información relevante para la tarea de clúster basada en la solución propia no es tan simple. Entre los enfoques propuestos para abordar el problema de agrupación espectral de múltiples vías, encontramos métodos recursivo re-clustering (cortes recursiva) para obtener el vector de indicador de agrupamiento, como se describe en (Carlos Alzate & Suykens, 2006). Este enfoque no es óptimo ya que sólo toma en consideración un vector propio y descarta la información de los vectores restantes, lo que también puede proporcionar una pista en cuanto a la conformación del conglomerado. El enfoque de re-clustering equivale a aplicar un método de agrupación de clústeres en el espacio de vectores propios, aunque esto sólo funciona correctamente cuando el espacio de los vectores propios tiene una estructura esférica (Takács et al., 2007). En (Meila & Shi, 2001b), los autores proponen un simple algoritmo de agrupamiento consistente en un método convencional para estimar los puntos de datos más representativos, uno por grupo. Esto se hace a través de la auto-descomposición de la matriz de similitud normalizada. Posteriormente, se obtiene una matriz de vectores propios normalizado de amplitud, que se agrupa en subconjuntos  $K$  usando el algoritmo de  $K$ -means. Finalmente, los puntos de datos se asignan a los agrupamientos correspondientes según la estructura geométrica de la matriz de datos inicial.

Un trabajo posterior (Hagen & Kahng, 1992) presenta un método de agrupamiento espectral de múltiples agrupaciones de una manera más detallada, en la que se plantea un problema de

optimización para garantizar que la convergencia del algoritmo se encuentra dentro de la región óptima global. Para ello se formulan dos sub-problemas de optimización: uno para obtener los valores óptimos globales en un dominio continuo sin restricciones y otro para obtener una solución discreta correspondiente al continuo. A diferencia del algoritmo introducido en (Meila & Shi, 2001b), este método no requiere un algoritmo de agrupación adicional, ya que genera una matriz binaria por sí sola que representa la pertenencia al punto de datos. De esta manera cada punto de datos pertenece a una agrupación única. En este estudio, se emplean transformaciones ortonormales y descomposición de valores singulares para determinar la solución discreta óptima, teniendo en cuenta el principio de la invariancia ortonormal y el criterio óptimo de auto-solución.

Otro estudio (Perona & Freeman, 1998) aplica los fundamentos de (Hagen & Kahng, 1992) y (Meila & Shi, 2001b) para explicar teóricamente la relación entre los K-medios de Kernel y la agrupación espectral, partiendo desde el punto de vista de la generalización de la función objetivo de K-means hacia la obtención de un caso especial de la función objetivo capaz de resolver el problema de corte normalizado (CN). Entonces, se asume una matriz de similitud definida positiva, así como algoritmos basados en búsquedas locales y se realizan formulaciones de optimización para mejorar el proceso de agrupación con respecto al kernel. Esto se hace de tal manera que podría satisfacerse la condición de optimización estableciendo que la función objetivo de agrupación basada en corte normalizado debe ser monótonamente decreciente.

### **2.2.2 Agrupamiento basado en kernel**

Bajo la suposición de que los datos son un grafo ponderado al azar, se demuestra que la agrupación espectral puede ser vista como una agrupación basada en la máxima similitud, la cual puede ser interpretada como un kernel. En (A. B. Chan & Vasconcelos, 2005), se analizan algunas

alternativas a la resolución de problemas abiertos en la agrupación espectral, como la selección de una escala de análisis adecuada, extensiones a la gestión de datos a múltiples escalas, la existencia de grupos irregulares y parcialmente conocidos y la estimación automática del número de grupos. Los autores proponen una escala local para calcular la afinidad o similitud entre cada par de puntos de datos.

Esta similitud escalada mejora los algoritmos de agrupación en términos tanto de convergencia como de tiempo de procesamiento. Además, se sugiere aprovechar la información subyacente dada por los vectores propios para establecer automáticamente el número de grupos. La salida de este algoritmo puede ser utilizada como punto de partida para inicializar algoritmos de particionamiento tales como K-means. Hay otros enfoques que se han señalado para ampliar el modelo de agrupación a nuevos datos (datos de prueba), extensión fuera de la muestra. Por ejemplo, los métodos presentados en (Rodríguez-Sotelo, Peluffo-Ordonez, Cuesta-Frau, & Castellanos-Domínguez, 2009) permiten ampliar el proceso de agrupación a nuevos datos, aproximando la función propia (una función basada en la auto descomposición) utilizando el método de Nyström. En ellos, un método de agrupación y un criterio de búsqueda se eligen típicamente de una manera heurística.

Otra perspectiva de agrupamiento espectral es el Análisis de Componentes Principales de Kernel (KPCA) (Maestri et al., 2009), (Choi & Lee, 2004). KPCA es una técnica no supervisada para la extracción característica no lineal y la reducción de la dimensionalidad. Este método es una versión no lineal de PCA utilizando matrices definidas positivas cuyo objetivo es encontrar el espacio proyectado en un espacio de característica del kernel inducido preservando la varianza máxima explicada (Choi & Lee, 2004). La relación entre KPCA y la agrupación espectral se

explica en (Maestri et al., 2009), (Choi & Lee, 2004). De hecho, el trabajo presentado en (Choi & Lee, 2004) demuestra que la agrupación espectral clásica, como el corte normalizado (Carlos Alzate & Suykens, 2010b), el algoritmo NJW (Zelnik-Manor & Perona, 2005) y el modelo de caminata aleatoria (Meila & Shi, 2001a) pueden verse como casos particulares de KPCA ponderada, en la matriz del kernel.

En (Takács et al., 2007), se propone un modelo de agrupación espectral que se basa en un KPCA esquema sobre la base de la propuesta en (Choi & Lee, 2004), con el fin de extender a multi-cluster de agrupaciones, en el que la codificación y decodificación de etapas están involucrados. Con este fin, una formulación basada en el método de los mínimos cuadrados basado en el apoyo vector de máquinas (LSSVM) teniendo en cuenta las versiones ponderadas.

Esta formulación apunta a un problema de optimización restringida en un cuadro primal-dual que permite extender el modelo a nuevos datos aislados sin necesidad de técnicas adicionales (por ejemplo, el método de Nyström). A respecto, se propone una matriz de similitud modificada cuyos vectores propios son duales con respecto a un problema de optimización primitiva formulado en un espacio de alta dimensión. Además, está probado que tales vectores propios contienen información relevante subyacente útil para propósitos de agrupación y muestran una estructura geométrica especial cuando los grupos resultantes están bien formados en términos de compacidad y / o separabilidad.

Las extensiones fuera de muestra se realizan proyectando indicadores de clúster en los vectores propios. Como criterio de selección de modelo, se propone el denominado criterio de equilibrio de línea (BLF). Este criterio explora la estructura de los vectores propios y sus correspondientes proyecciones, que pueden utilizarse para establecer los parámetros iniciales del modelo.

Del mismo modo, otro modelo de kernel para espectral agrupación se presenta. Este método se basa en la incompleta descomposición de Cholesky. Es capaz de manejar eficientemente los problemas de agrupación a gran escala. La puesta en marcha se realiza sobre la base de un esquema WPCA basado en kernels (WKPCA) a partir de una formulación primal-dual. En ella, la matriz de similitud se estima utilizando la descomposición incompleta de Cholesky.

## CAPITULO III

### MARCO TEÓRICO

#### 3.1 Introducción

El objetivo de esta sección es describir algunos aspectos básicos y fundamentos con respecto algebra lineal, sección 3.2; el agrupamiento espectral convencional sección 3.3; la sección 3.4 presenta en términos generales la definición de kernel, matriz kernel y algunas propiedades y conceptos relacionados. Finalmente, en la sección 3.5 se describe la formulación y resolución del problema de optimización

#### 3.2 Elementos de algebra lineal

##### 3.2.1 Matriz

Una matriz es una tabla de  $m \times n$  elementos ordenados en filas y columnas. Se representa por letras mayúsculas:  $A, B, C, \dots$  etc. Y a sus elementos como  $a_{ij}$  donde el primer subíndice indica la fila y el segundo la columna a la que pertenece dicho elemento (Perry, Sánchez Trujillo, & others, 1990).

Así, una matriz  $A = (a_{ij})$  con  $1 \leq i \leq m$  ;  $1 \leq j \leq n$  es de la forma:

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

##### 3.2.2 Orden de una matriz

Una Matriz de  $m$  filas y  $n$  columnas tiene dimensión u orden  $m \times n$ , y al conjunto de todas las matrices de este tipo se denota por  $\mathbb{R}^{m \times n}$ .

Dos matrices  $A, B \in \mathbb{R}^{m \times n}$ , son iguales si y solo si:  $a_{ij} = b_{ij} \forall i = 1, 2, \dots, m$  y  $\forall j = 1, 2, \dots, n$

$$\mathbf{A}_{m \times n} = \mathbf{B}_{m \times n} \leftrightarrow (a_{ij}) = (b_{ij})$$

### 3.2.3 Matrices fila y columna

Se denomina *matriz fila* a aquella que tiene una sola fila.

$$\mathbf{A} = (a_1, a_2, \dots, a_n) \in \mathbb{R}^{1 \times n}$$

De igual manera, se tiene una *matriz columna* cuando únicamente tiene una columna.

$$\mathbf{A} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \in \mathbb{R}^{n \times 1}$$

A este tipo de matrices también se les conoce como *Vectores*.

### 3.2.4 Matriz cuadrada

Una matriz cuadrada es la que tiene el mismo número de filas que de columnas. Se dice que una matriz cuadrada  $n \times n$  es de orden  $n$  y se denomina matriz  $n$ -cuadrada.

$$\mathbf{A} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}$$

Ejemplo: Sean las matrices

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 5 & 6 & 7 \\ 4 & 8 & 9 \end{pmatrix} \quad \text{y} \quad \mathbf{B} = \begin{pmatrix} 3 & 7 \\ 5 & 8 \end{pmatrix}$$

Entonces  $\mathbf{A}$  y  $\mathbf{B}$  son matrices cuadradas de orden 3 y 2 respectivamente.

Se denomina *diagonal principal* de una matriz cuadrada a la formada por los elementos

$a_{ii}$  donde  $i = 1, 2, 3, \dots, n$ .

### 3.2.5 Matriz diagonal

Se dice que una matriz cuadrada es *diagonal* cuando todos los elementos situados fuera de la diagonal principal son nulos. Es decir, en una matriz diagonal se verifica que  $a_{ij} = 0$  para  $i$  distinto de  $j$ . Así, la siguiente matriz es diagonal:

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 6 \end{pmatrix}$$

Aquella matriz diagonal es la cual los elementos de la diagonal son iguales denomina matriz Escalar.

### 3.2.6 Matriz identidad

Sea  $A = (a_{ij})$  una matriz  $n$ -cuadrada. La diagonal principal de  $A$  consiste en los elementos  $a_{11}, a_{22}, a_{nn}$ . La matriz  $n$ -cuadrada con 1 en la diagonal principal y ceros en cualquier otra posición, denotada por  $I$ , se conoce como matriz identidad. Para cualquier matriz  $A$ ,

$$A \cdot I = I \cdot A = A$$

$$I = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}_{n \times n}$$

### 3.2.7 Matriz triangular

Se denomina matriz triangular superior e inferior a aquella matriz cuadrada cuyos elementos situados por debajo o arriba de su diagonal principal son todos nulos.

Triangular superior:  $a_{ij} = 0$  si  $i > j$

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ 0 & a_{22} & a_{23} & \dots & a_{2n} \\ 0 & 0 & a_{33} & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & a_{nn} \end{pmatrix}$$

Triangular inferior:  $a_{ij} = 0$  si  $i > j$

$$\begin{pmatrix} a_{11} & 0 & 0 & \cdots & 0 \\ a_{21} & a_{22} & 0 & \cdots & 0 \\ a_{31} & a_{32} & a_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{pmatrix}$$

### 3.2.8 Matriz escalonada

El equivalente para matrices rectangulares de una matriz triangular es denominado matrices escalonadas que son aquellas matrices que  $a_{ij} = 0$  si  $i > j$ . El orden de una matriz escalonada no necesariamente es cuadrada, en caso de ser una matriz cuadrada, se tendría una matriz triangular superior.

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1m-1} & \cdots & a_{1n} \\ 0 & a_{22} & a_{23} & \cdots & a_{2m-1} & \cdots & a_{2n} \\ 0 & 0 & a_{33} & \cdots & a_{3m-1} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & a_{mm} & \cdots & a_{mn} \end{pmatrix} \quad \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & a_{22} & a_{23} & \cdots & a_{2n} \\ 0 & 0 & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn} \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

### 3.2.9 MATRIZ TRASPUESTA

Sea  $A \in \mathbb{R}^{n \times n}$  se denomina matriz traspuesta de  $A$  y se denota por  $A^T$  a la matriz resultante de cambiar, ordenadamente, las filas por las columnas de la matriz  $A$  de tal manera, que si llamamos  $A = (a_{ij})$  y  $A^T = (a'_{ij})$  tenemos:

$$a_{ij} = a_{ji} \quad \text{con } 1 \leq i \leq m \text{ y } 1 \leq j \leq n$$

por lo que si  $\mathbf{A} \in \mathbb{R}^{m \times n} \Rightarrow \mathbf{A}^T \in \mathbb{R}^{n \times m}$

Propiedades de la matriz transpuesta.

1.  $(\mathbf{A}^T)^T = \mathbf{A}$
2.  $(k\mathbf{A})^T = k\mathbf{A}^T, \quad k \in \mathbb{R}$
3.  $(\mathbf{A} + \mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$
4.  $(\mathbf{A} * \mathbf{B})^T = \mathbf{B}^T * \mathbf{A}^T$

### 3.2.10 Matrices simétrica y antisimétrica

Se dice que una matriz cuadrada es simétrica, si  $\mathbf{A}^T = \mathbf{A}$

$$\mathbf{A} = \begin{pmatrix} 3 & 4 & -1 \\ 4 & 5 & 2 \\ -1 & 2 & 1 \end{pmatrix}$$

Además, es antisimétrica, si  $\mathbf{A}^T = -\mathbf{A}$ .

### 3.2.11 Matriz ortogonal

Una matriz cuadrada y no singular se dice ortogonal si su transpuesta coincide con su inversa, es decir, si  $\mathbf{A}^T = \mathbf{A}^{-1}$  o lo que es lo mismo:

$$\mathbf{A} \text{ ortogonal} \Leftrightarrow \mathbf{A} \mathbf{A}^T = \mathbf{A}^T \mathbf{A} = \mathbf{I}$$

### 3.2.12 Traza de una matriz

En una matriz cuadrada de orden  $n$  la diagonal principal está formada por los elementos  $a_{ii}$  ( $i=1, 2, \dots, n$ ). La traza de una matriz cuadrada  $\mathbf{A}$ , a la que designa por  $tr(\mathbf{A})$ , o por  $traza(\mathbf{A})$ , es la suma de los elementos de la diagonal principal. Por lo tanto,

$$\text{tr}(\mathbf{A}) = \sum_{i=1}^n a_{ii}$$

Propiedades

$$1) \text{tr}(\mathbf{A} + \mathbf{B}) = \text{tr}(\mathbf{A}) + \text{tr}(\mathbf{B})$$

$$2) \text{tr}(\alpha\mathbf{A}) = \alpha \text{tr}(\mathbf{A})$$

### 3.2.13 Inversa de una matriz

Poole (2011) expresa la definición de matriz inversa como: “Si  $\mathbf{A}$  es una matriz de orden  $n \times n$  esta tendrá una inversa denominada por  $\mathbf{A}^{-1}$  de  $n \times n$  y que cumple la siguiente propiedad:

$$\mathbf{A} * \mathbf{A}^{-1} = \mathbf{I} \quad \text{y} \quad \mathbf{A}^{-1} * \mathbf{A} = \mathbf{I}$$

Ejemplo:

Si  $\mathbf{A} = \begin{pmatrix} 3 & -5 \\ -1 & 2 \end{pmatrix}$ , entonces  $\mathbf{A}^{-1} = \begin{pmatrix} 2 & 5 \\ 1 & 3 \end{pmatrix}$ , debido a que:

$$\mathbf{A} * \mathbf{A}^{-1} = \begin{pmatrix} 3 & -5 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 2 & 5 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{y}$$

$$\mathbf{A}^{-1} * \mathbf{A} = \begin{pmatrix} 2 & 5 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} 3 & -5 \\ -1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

### 3.2.14 Independencia lineal

Sea un conjunto de  $m$  vectores  $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m\}$ , de orden  $m \times 1$ . Si la única solución de la ecuación

$$\gamma_1 \mathbf{x}_1 + \gamma_2 \mathbf{x}_2 + \dots + \gamma_m \mathbf{x}_m = \mathbf{0}$$

es  $\gamma_1 = \gamma_2 = \dots = \gamma_m = 0$  los vectores  $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m\}$  son *linealmente independientes*. Si existen otras soluciones se dice que son *linealmente dependientes*.

Las filas de una matriz pueden ser consideradas como vectores. Es posible que sean linealmente independientes (L.I.) o es posible que unos dependan linealmente de otros. Por ejemplo:

$\mathbf{A} = \begin{pmatrix} 2 & 3 & 2 & 5 \\ 1 & 3 & 4 & 2 \end{pmatrix}$  Sus filas son *linealmente independientes*.

$\mathbf{B} = \begin{pmatrix} 1 & 3 \\ 2 & 1 \\ 0 & 5 \\ 3 & 4 \end{pmatrix}$  Las dos primeras son filas L.I., las demás, dependen de las dos primeras.

### 3.2.15 Rango de una matriz

El concepto de rango se encuentra ligado al de “independencia lineal” de filas o columnas de una matriz. Por lo tanto, el rango de una matriz está definido como máximo número de filas o columnas linealmente independientes.

- El rango por filas de una matriz es el número de filas linealmente independientes.
- El rango por columnas de una matriz es el número de columnas linealmente independientes.

Se puede demostrar que el rango de una matriz por filas coincide con el rango por columnas en cualquier matriz. A este valor en común se le llama rango de una matriz y se representa por  $Rg(\mathbf{A})$ .

Si el rango de una matriz cuadrada  $\mathbf{A}^{n \times n}$  es  $n$ , se dice que es de máximo rango o rango completo. En este caso la matriz  $\mathbf{A}$  es no singular y, por tanto,  $|\mathbf{A}| \neq 0$ .

### 3.2.16 Espacio vectorial

Un espacio vectorial  $\mathbb{V}$  es una reunión de vectores  $\{\alpha_i\}_{i=1}^n$ . Estos vectores deben satisfacer la condición de sumarse entre sí y multiplicarse por un escalar cada uno. Donde  $\{\alpha_i\}_{i=1}^n$  es un conjunto de escalares (números reales). Para una mejor comprensión de esta definición se utiliza otra nomenclatura, pero considerando los elementos del espacio vectorial, esto es:

Dados los vectores  $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{V}$  y el escalar  $\alpha \in \mathbb{R}$ ; los espacios vectoriales satisfacen los siguientes axiomas:

i)  $\mathbf{a} + \mathbf{b} \in \mathbb{V}$  (Propiedad de adición)

ii)  $\alpha \mathbf{a} \in \mathbb{V}$  (Propiedad multiplicación por un escalar)

iii)  $(\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c}) = \mathbf{a} + \mathbf{b} + \mathbf{c}$  (Propiedad asociativa de la adición)

### 3.2.17 Subespacio vectorial

El conjunto  $\mathbb{E} \subset \mathbb{V}$  es un subespacio vectorial si este es un espacio vectorial bajo las operaciones definidas en  $\mathbb{V}$ .

### 3.2.18 Combinación lineal

Dados los conjuntos de vectores  $\{\mathbf{v}_i\}_{i=1}^n$  y un conjunto de escalares  $\{\alpha_i\}_{i=1}^n$ , una combinación lineal está definida como:

$$\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \dots + \alpha_n \mathbf{v}_n = \sum_{i=1}^n \alpha_i \mathbf{v}_i$$

Si todo  $\mathbf{v}_i$  son Linealmente Independientes (LI), pueden generar un espacio vectorial, como:

$$\mathbb{V} = \sum_{i=1}^n \alpha_i \mathbf{v}_i$$

### 3.2.19 Operador matemático

Un operador matemático binario  $n$  - dimensional  $O(\cdot)$  sobre el dominio de  $\mathbb{D}$  puede ser definido de la siguiente manera:

$$\begin{aligned} O(\cdot, \cdot): \mathbb{D}^n \times \mathbb{D}^n &\rightarrow \mathbb{K} \\ a, b &\mapsto O(a, b), \end{aligned}$$

Donde  $O(a, b) = \phi(a, b)$  y  $\phi(\cdot, \cdot)$  es una función dada.

### 3.2.20 Producto Interno

Dados  $x, y, z \in \mathbb{E}$  y  $\alpha, \beta \in \mathbb{R}$ . En general un producto interno está definido así:

$$\begin{aligned} \langle \cdot, \cdot \rangle : \mathbb{E} \times \mathbb{E} &\rightarrow \mathbb{K} \\ a, b &\mapsto \langle a, b \rangle \end{aligned}$$

El mismo que satisface los axiomas siguientes:

$$i) \langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle ,$$

$$ii) \langle x, x \rangle \geq 0 ,$$

$$iii) \langle x, y \rangle = \overline{\langle y, x \rangle} .$$

Producto interno para espacio Euclidiano

Se define de la siguiente manera:

$$\begin{aligned} \langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n &\rightarrow \mathbb{R} \\ \mathbf{u}, \mathbf{v} &\mapsto \langle \mathbf{u}, \mathbf{v} \rangle \end{aligned}$$

Donde

$$\langle \mathbf{u}, \mathbf{v} \rangle = \sum_{i=1}^n u_i v_i = \mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T \mathbf{v}$$

y

$$\mathbf{u} = (u_1, \dots, u_n)^T \text{ y } \mathbf{v} = (v_1, \dots, v_n)^T$$

### 3.2.21 Norma

La norma de un vector  $\mathbf{x} \in \mathbb{E}$  es denotado como  $\|\mathbf{x}\|$  y definido de la forma siguiente:

$$\begin{aligned} \|\cdot\| : \mathbb{E} &\rightarrow \mathbb{R}^+ \cup \{0\} \\ \mathbf{x} &\mapsto \|\mathbf{x}\| \end{aligned}$$

Además, cumple los axiomas:

$$i) \|\mathbf{x}\| \geq 0,$$

$$ii) \|\alpha \mathbf{x}\| = |\alpha| \|\mathbf{x}\|$$

$$iii) \|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$$

Norma para un espacio Euclidiano

Para el caso de un espacio Euclidiano  $n$ -dimensional ( $\mathbb{R}$ ), se tiene:

$$\begin{aligned} \|\cdot\|_{L_2} : \mathbb{R}^n &\rightarrow \mathbb{R}^+ \\ \mathbf{x} &\mapsto \|\mathbf{x}\|_{L_2} \end{aligned}$$

Donde

$$\|\mathbf{x}\|_{L_2} = \left( \sum_{i=1}^n x_i^2 \right)^{1/2}$$

Y, se conoce que  $\mathbf{x} = (x_1, \dots, x_n)^T$

### 3.2.21.1 Normas de vectores

Una norma es una función  $\|\cdot\|: \mathbb{R}^n \rightarrow \mathbb{R}$  que asigna un número real no negativo (tamaño) a cada vector.

Sean  $x, y \in \mathbb{R}^n$  y escalar  $\beta$ , se tiene:

i)  $\|\mathbf{x}\| \geq 0$ , y  $\|\mathbf{x}\| = 0$  solo si  $\mathbf{x} = 0$ ,

ii)  $\|\beta\mathbf{x}\| = |\beta|\|\mathbf{x}\|$ ,

iii)  $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ .

Tres normas de vectores importantes en optimización

$$\|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i| \quad , \quad \|\mathbf{x}\|_2 = \left( \sum_{i=1}^n x_i^2 \right)^{1/2} \quad , \quad \|\mathbf{x}\|_\infty = \max_{1 \leq i \leq n} |x_i|$$

### 3.2.21.2 Normas inducidas de matrices

Si  $\|\cdot\|$  es una norma en  $\mathbb{R}^n$ , la norma  $\|\cdot\|$  de matrices inducida por ella, viene definida por:

$$\|A\| = \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|} = \max_{\|x\|=1} \|Ax\|$$

$$\|A\|_1 = \max_{1 \leq j \leq n} \|A_j\|_1, \text{ (columnas)}$$

$$\|A\|_\infty = \max_{1 \leq i \leq m} \|a_i^T\|_1, \text{ (filas)}$$

$$\|A\|_2 = \sqrt{\lambda_{\max} \operatorname{tr}(A^T A)}$$

Propiedad de normas inducidas

$$\|Ax\| \leq \|A\| \|x\|$$

### 3.2.21.3 Norma de Frobenius

Una famosa norma no inducida es la norma de Frobenius:

$$\|A\|_F = \left( \sum_{i=1}^m \sum_{j=1}^n a_{ij}^2 \right)^{\frac{1}{2}}$$

$$\|A\|_F = \left( \sum_{j=1}^n \|A_j\|^2 \right)^{\frac{1}{2}}$$

$$\|A\|_F = \sqrt{\operatorname{tr}(A^T A)} = \sqrt{\operatorname{tr}(AA^T)}.$$

### 3.2.22 Estimación de valores y vectores propios

**Teorema 1** (Raíces del polinomio característico). Sean  $A \in \mathbb{R}^{n \times n}$ ,  $x \in \mathbb{C}^n$ ,  $x \neq 0$ , y  $\lambda \in \mathbb{C}$ .  $\lambda$  es valor propio de  $A$  si y solo si es raíz del polinomio característico.

$$P_n(\lambda) = \det(\lambda I - A)$$

$$Ax = \lambda x \Leftrightarrow (\lambda I - A)x = 0 \Leftrightarrow A \text{ es singular} \Leftrightarrow \det(\lambda I - A) = 0.$$

Los vectores propios asociados a valores propios distintos son Linealmente Independientes (LI).

Si existe un conjunto de  $n$  vectores propios LI, tenemos la descomposición matricial  $\mathbf{A} = \mathbf{X} \mathbf{\Lambda} \mathbf{X}^{-1}$ , siendo  $\mathbf{\Lambda} = \text{diag}(\lambda_i)$ .

**Teorema 2** (Teorema Fundamental del Algebra). *Se tiene que  $P_n(\lambda)$  tiene una única factorización:*

$$\det(\lambda \mathbf{I} - \mathbf{A}) = P_n(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_n).$$

**Teorema 3** (Existencia y unicidad).  $\mathbf{A} \in \mathbb{R}^{n \times n}$  posee  $n$  valores propios contando multiplicidad en las raíces del polinomio característico. Los valores propios y sus multiplicidades son únicas.

**Teorema 4.** *Propiedades que se desprenden:*

$$a) \det(\mathbf{A}) = \prod_{j=1}^n \lambda_j, \text{tr}(\mathbf{A}) = \sum_{j=1}^n \lambda_j.$$

b) Si  $\lambda = (a + bi)$  es valor propio, entonces  $\bar{\lambda} = (a - bi)$  es valor propio.

c) Si  $n$  es impar, al menos un valor propio es real.

d) Si  $\mathbf{A}$  es triangular entonces  $\lambda_i(\mathbf{A}) = a_{ii}$ .

### 3.2.23 Introducción a la optimización

#### 3.2.23.1 Multiplicadores de Lagrange

El método de los multiplicadores de Lagrange esencialmente proporciona un conjunto de condiciones necesarias para identificar candidatos al óptimo en problemas de optimización con restricciones de igualdad. Esto se hace convirtiendo el problema con restricciones de igualdad en un problema equivalente sin restricciones con la ayuda de ciertos parámetros no especificados conocidos como Multiplicadores de Lagrange.

Un problema de optimización con restricciones tiene la siguiente forma:

$$\begin{array}{ll} \min & f(x) \\ \text{s.a.} & h(x)=0, \end{array} \quad \text{problema (1)}$$

donde  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $h : \mathbb{R}^n \rightarrow \mathbb{R}$ , y  $x \in \mathbb{R}^n$ .

El método de los multiplicadores de Lagrange consiste en convertir el *problema (1)* en el siguiente problema sin restricciones:

$$\begin{array}{ll} \min & L(x, \lambda) \\ & x \in \mathbb{R}^n, \lambda \in \mathbb{R} \end{array}$$

La función  $L(x, \lambda)$  es llamada función Lagrangiana y  $\lambda$  es una constante no especificada llamada multiplicador de Lagrange. No hay restricciones de signo para el valor de  $\lambda$ .

**Definición 1** (Función Lagrangiana). La función lagrangiana-asociada al *problema (1)* es:

$$L(x, \lambda) = f(x) + \lambda g(x),$$

donde  $\lambda \in \mathbb{R}$  es el multiplicador de Lagrange.

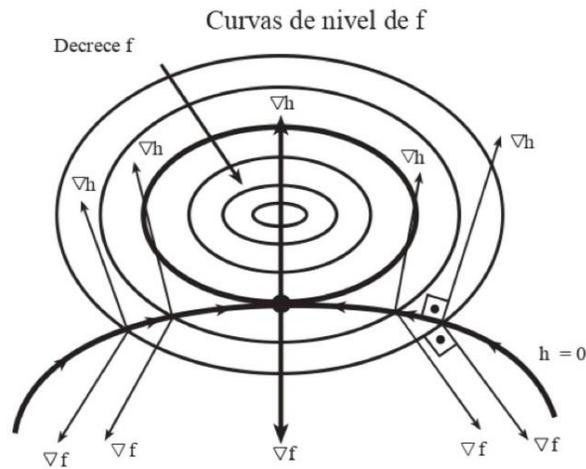
Gráficamente, el método busca puntos donde las curvas de nivel de  $f$  y  $h$  sean tangentes entre sí. Esto es lo mismo que encontrar puntos donde los vectores de los gradientes de  $f$  y  $h$  sean paralelos entre sí. Todo el proceso puede reducirse al hacer el gradiente de la función Lagrangiana igual al vector cero.

**Teorema 5** (condiciones de optimalidad). Las condiciones necesarias de optimalidad del *problema (1)* son las siguientes:

$$1) \quad \nabla f(x^*) - \lambda^* \nabla h(x^*) = 0$$

$$2) \quad h(x^*) = 0,$$

Estas condiciones son llamadas condiciones Karush-Kuhn-Tucker o simplemente condiciones KKT (*Figura 1*).



**Figura 1.** Ilustración de Karush-Kuhn-Tucker para el caso de una restricción de igualdad y dos variables

Ahora, considerando problemas de optimización con  $k$  restricciones de igualdad,

$$\begin{array}{ll} \min & f(x) \\ \text{s.a.} & h_i(x)=0, \\ & \text{para } i=1,\dots,k. \end{array} \quad \text{Problema (2)}$$

donde  $f: \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $h_i: \mathbb{R}^n \rightarrow \mathbb{R}$ , con  $i = 1, \dots, k$  y  $x \in \mathbb{R}^n$ .

La función Lagrangiana-asociada al Problema (2) es:

$$L(x, \lambda) = f(x) + \sum_{i=1}^k \lambda_i h_i(x), \quad \text{para } i = 1, \dots, k$$

donde  $\lambda = (\lambda_1, \dots, \lambda_k)$ .

**Teorema 6** (Condiciones de optimalidad). Las condiciones necesarias de optimalidad del Problema (2) son:

- 1)  $\nabla f(x^*) - \sum_{i=1}^n \lambda_i^* \nabla h_i(x^*) = 0$
- 2)  $h(x^*) = 0$ , para  $i = 1, \dots, k$ .

**Ejemplo 1.**

$$\begin{array}{ll} \min & (z_1 - 1)^2 + (z_2 - z_3)^2 + (z_4 - z_5)^2 \\ \text{s. a.} & z_1 + z_2 + z_3 + z_4 + z_5 = 0 \\ & z_3 - 2(z_4 + z_5) + 3 = 0 \end{array}$$

Sean:

$$f(z) = (z_1 - 1)^2 + (z_2 - z_3)^2 + (z_4 - z_5)^2,$$

$$h_1(z) = z_1 + z_2 + z_3 + z_4 + z_5,$$

$$h_2(z) = z_3 - 2(z_4 + z_5) + 3,$$

$$z = (z_1 + z_2 + z_3 + z_4 + z_5) \in \mathbb{R}^5 \text{ y } \lambda = (\lambda_1, \lambda_2). (n = 5 \text{ y } k = 2)$$

la función Lagrangiana asociada al problema es:

$$L(x, \lambda) = f(x) + \sum_{i=1}^2 \lambda_i h_i(x)$$

Las condiciones de optimalidad son:

- $\nabla f(x^*) - \sum_{i=1}^2 \lambda_i^* \nabla h_i(x^*) = 0$
- $h(x^*) = 0$ , con  $i = 1, 2$ .

**Observación 1.** Una solución al sistema aumentado de  $n + k$  ecuaciones en  $n + k$  variables proporciona un punto estacionario  $x^*$  de la función lagrangiana  $L$ . Puede verificarse si este punto  $x^*$  es un mínimo o un máximo verificando que la matriz hessiana de  $L$  con respecto a  $x^*$  es PD.

### 3.2.23.2 Funciones reales cuadráticas

Las formas Cuadráticas son aplicaciones que transforman a vectores de  $n$  dimensiones en números reales, premultiplicando una matriz cuadrada simétrica por dicho vector. Toda forma cuadrática se puede escribir como:

$$q(x) = \frac{1}{2}x^T Ax - b^T x + c,$$

donde  $A \in \mathbb{R}^{n \times n}$  es una matriz simétrica, los vectores  $x, b \in \mathbb{R}^n$  y la constante  $c \in \mathbb{R}$

**Ejemplo 2.** Para

$$q(x_1 + x_2 + x_3) = x_1^2 + 3x_3^2 + 2x_1x_2 - 5x_2x_3 + x_1 - x_3 - 4,$$

Se obtiene que:

$$A = \begin{pmatrix} 2 & 2 & 0 \\ 2 & 0 & -5 \\ 0 & -5 & -6 \end{pmatrix}, \quad b = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \quad c = -4.$$

## 3.3 Agrupamiento espectral

En los últimos años el reconocimiento y clasificación de patrones ha adquirido cierta popularidad gracias a la automatización de las soluciones a muchos problemas de la vida real y en particular, los métodos de agrupamiento (clustering), debido a que cada vez es más necesario tratar grandes volúmenes de datos que requieren herramientas nuevas para descubrir información relevante y relación entre los datos. Cada una de las estrategias de agrupamiento existentes intenta resolver el problema de la clasificación de los puntos teniendo en cuenta la presencia o no de determinadas características en la base de datos.

Las técnicas no supervisadas de reconocimiento y clasificación de patrones son ampliamente usadas en problemas que requieren herramientas para descubrir información relevante y relaciones

entre los datos. Aunque existe una gran variedad de técnicas, muchas de ellas tienen como desventaja que cuando los grupos están muy solapados no son capaces de separarlos, más aún si hay ruido y tienen formas muy irregulares. Los métodos de agrupamiento espectral tienden a determinar la estructura subyacente en un conjunto de patrones, donde otros métodos convencionales por la disposición y características particulares de los agrupamientos no obtienen los resultados esperados. Por ejemplo, el algoritmo k-Means construye grupos globulares, las estrategias jerárquicas tienden a construir grupos alargados, algunos algoritmos de agrupamiento no manejan el ruido. Todas estas consideraciones constituyen limitaciones de los algoritmos de agrupamiento ante la existencia de conjuntos de puntos de diversas formas y densidades, con determinado grado de solapamiento entre los grupos y en presencia de ruido. Entre los métodos de agrupamiento, el algoritmo de k-medias ha sido, tal vez, el más difundido.

Cuando la estructura de los datos no corresponde a regiones convexas, no es lineal o cuando los métodos de agrupamientos clásico, jerárquicos o particionales, no obtienen resultados satisfactorios, una alternativa son los métodos de agrupamiento espectral (spectral clustering) (Ng, Jordan, & Weiss, 2002c). Las técnicas de agrupamiento espectral utilizan los vectores propios de la matriz de afinidad, o de matrices derivadas, para generar una partición del conjunto de muestras en agrupamientos disjuntos, que presenten valores altos de la medida de semejanza adoptada para patrones en un mismo conjunto, y bajos para patrones de conjuntos diferentes. En general el valor de los correspondientes valores propios determina un criterio que permite establecer la prioridad de los vectores propios, esta no necesariamente genera la mejor partición del espacio muestral.

Últimamente, los algoritmos de agrupamiento basados en grafos y análisis espectral han ganado auge por su fácil implementación al utilizar conceptos de álgebra lineal (Sánchez, Osorio, &

Suárez, 2008) que son de uso extensivo y que has sido aplicados exitosamente en diseño de circuitos, segmentación de imágenes (Ishikawa & Geiger, 1998), entre otros. La mayor parte de estos métodos son discriminativos, es decir no requieren de suposiciones sobre la estructura de datos. Este tipo de análisis puede ser de manera fácil explicado mediante la teoría de grafos.

### 3.4 Teoría de grafos

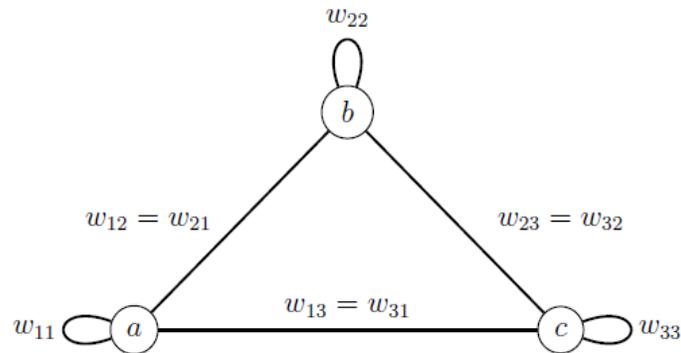
La teoría de gráficas o teoría de grafos es aplicada entre otras, en áreas tales como ciencias sociales, ciencias físicas, ingeniería de comunicación; pero, básicamente juega un papel importante en las ciencias de la computación, tales como inteligencia artificial, lenguajes formales, teoría de cambio y lógica de diseño, gráficos por computadora, sistemas operativos, compiladores, y organización y recuperación de información, en lo que respecta al modelado de problemas, indicando sus características de manera muy objetiva.

Para el análisis espectral es fundamental el conocimiento y uso de una geometría empleando bloques topológicos que comúnmente se les llama grafos. En términos elementales podríamos definir un grafo como un conjunto de puntos (llamados elementos, vértices, nudos o nodos) con líneas que unen pares de vértice de ellas. La teoría de grafos es usada, por lo general, en el modelado de problemas como tráfico de redes, circuitos eléctricos y redes de internet.

**Definición:** Grafo es una abstracción matemática que designaremos por  $\mathbb{G} = (\mathbb{V}, A)$ , donde  $\mathbb{V}$  es un conjunto de puntos o vértices  $v_i$ ,  $\mathbb{V} \neq 0$  y  $A$  es un conjunto de líneas o aristas que unen dos puntos de  $\mathbb{V}$ ;  $A$  puede ser vacío ( $\emptyset$ ), llamado conjunto de las aristas que están relacionados mediante la aplicación (Alvarez Nuñez, 2013).

Un grafo ponderado es un grafo en el cual, además del orden y el sentido, es importante el grado de afinidad entre los vértices, se define a través de una matriz de pesos  $\mathbf{W}$ , donde  $w_{ij}$  es un valor

positivo que corresponde al peso de la arista ubicada entre los vértices  $i$  y  $j$  (**Figura 2**). Además, un grafo ponderado puede describirse por la forma  $\mathbb{G} = (\mathbb{V}, A, \mathbf{W})$ , donde  $\mathbf{W}$  es una matriz semidefinida positiva y simétrica.



**Figura 2.** Grafo ponderado

Fuente: (Valencia, 2017)

### 3.4 Función y matriz kernel

#### 3.4.1 Función kernel

En forma general, el termino kernel es usado para definir una función que establece la similitud entre los elementos de entrada dados. Las funciones de kernel consideradas en este trabajo son definidos positivos.

**Definición:** Una función kernel es una función (López Díaz, 2018) dada por

$$\mathcal{K}(\cdot, \cdot) : \mathbb{K}^d \times \mathbb{K}^d \rightarrow \mathbb{K}$$

$$x_i \times x_j \rightarrow \mathcal{K}(x_i, x_j),$$

que asigna a cada par de objetos del espacio de entrada  $\mathbb{K}$ , un valor real correspondiente al producto

escalar de las imágenes de dichos objetos en un espacio, donde  $\mathbb{K} = \mathbb{C}$  o  $\mathbb{R}$  se evidencia que los elementos  $x_i$  son reales y de dimensión  $d$ .

**Definición:** Una función  $\mathcal{K}(\cdot, \cdot) : \mathbb{K}^d \times \mathbb{K}^d \rightarrow \mathbb{K}$  es simétrica si para cada par de objetos  $x_i, x_j \in \mathbb{K}$  se cumpla que:  $\mathcal{K}(x_i, x_j) = \mathcal{K}(x_j, x_i)$

**Definición:** Una función  $\mathcal{K}(\cdot, \cdot) : \mathbb{K}^d \times \mathbb{K}^d \rightarrow \mathbb{K}$  es semi-definida positiva si:

$$\sum_{i=1}^N \sum_{j=1}^N c_i c_j \mathcal{K}(x_i, x_j) \geq 0$$

### 3.4.2 Matriz kernel

Una matriz de Gram o matriz kernel se obtiene cuando, si en una función kernel hay un total de  $N$  objetos, se genera la matriz  $\mathbf{\Omega}$  de dimensión  $N \times N$ , con entradas  $ij$ , establecida por  $\Omega_{ij} = \mathcal{K}(x_i, x_j)$ . Por lo tanto, la matriz kernel cumple la definición anterior es definida positiva que satisface:

$$\sum_{i=1}^N \sum_{j=1}^N c_i \bar{c}_j \Omega_{ij} \geq 0 \quad (1)$$

para todo  $c_i \in \mathbb{C}$ , donde  $\bar{c}_i$ , denota el complejo conjugado de  $c_i$ . (Schölkopf, Smola, Bach, & others, 2002).

Además, una matriz es simétrica positiva definida si y sólo si todos sus valores propios son positivos.

Se establece que, para posteriores desarrollos matemáticos, la notación:

La función que mapea desde un espacio de dimensión  $d$  a un espacio de dimensión  $d_h$  es considerada de la forma  $\phi(\cdot)$ , tal que:

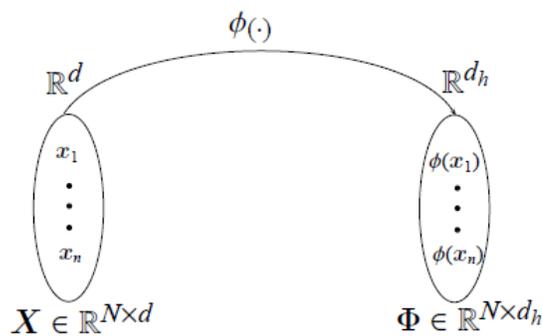
$$\begin{aligned} \phi(\cdot) : \mathbb{R}^d &\rightarrow \mathbb{R}^{d_h} \\ x_i &\rightarrow \phi(x_i) \end{aligned}$$

En consecuencia, con la función  $\phi$  puede obtenerse la matriz  $\Phi$  :

$$\Phi = [\phi(x_1)^T, \dots, \phi(x_N)^T]^T, \Phi \in \mathbb{R}^{N \times d_h},$$

y que consiste una representación de alta dimensión del espacio de entrada  $X \in \mathbb{R}^{N \times d}$  (Ver **Figura 3**).

Según los criterios de Mercer (Bhavsar & Ganatra, 2016), existe una propiedad muy útil que es el *truco kernel*. Esta propiedad es importante en la teoría de funciones kernel, en vista de que permite remplazar una función kernel positiva definida con otra función kernel que es finita y aproximadamente positiva definida.



**Figura 3.** Mapeo de alta dimensión de espacio de entrada  $X$  usando la función  $\phi$

Fuente: (Valencia, 2017)

Como ejemplo, si se dispone de un algoritmo formulado en términos de una función kernel positiva definida  $\mathcal{K}$ , se construye un algoritmo alternativo sustituyendo el algoritmo formulado por otra función kernel positiva definida  $\tilde{\mathcal{K}}$  (Schölkopf et al., 2002), por lo tanto que  $\Phi\Phi^T = \Omega$ . En consecuencia, la función kernel  $\Phi\Phi^T$  es considerada *truco kernel*.

### 3.4.3 Tipos de funciones kernel

Las funciones kernel de base radial (*Radial basis function -RBF*) se pueden escribir en términos de una medida de similitud o disimilitud, así:

$$\mathcal{K}(x_i, x_j) = f(d(x_i, x_j)),$$

Donde  $d(\cdot, \cdot)$  es una medida en el dominio de  $\mathbf{X}$ , esto es en el espacio  $\mathbb{R}^d$  así:

$$d(\cdot, \cdot) : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^+$$

$$x_i, x_j \mapsto d(x_i, x_j),$$

y  $f$  es una función definida en  $\mathbb{R}^+$ . Básicamente, dichas medidas se basan en el producto interior:

$$d(x_i, x_j) = \|x_i - x_j\| = \sqrt{\langle x_i - x_j, x_i - x_j \rangle}.$$

En la **Tabla 1** se describe algunas funciones kernels recomendadas por la literatura.

**Tabla 1**

*Ejemplos de funciones kernel*

Denominación	Definición	Dominio
Lineal	$\langle x_i, x_j \rangle$	$\mathbb{R}^d$
Polinomial	$\langle x_i, x_j \rangle^d$	$\mathbb{R}^d$
Racional Cuadrático	$1 - \frac{\ x_i - x_j\ ^2}{\ x_i - x_j\ ^2 + \sigma}, \sigma \in \mathbb{R}^+$	$\mathbb{R}^d$
exponencial	$\exp\left(-\frac{\ x_i - x_j\ }{2\sigma^2}\right), \sigma \in \mathbb{R}^+$	$\mathbb{R}^d$
Gaussiano	$\exp\left(-\frac{\ x_i - x_j\ ^2}{2\sigma^2}\right), \sigma \in \mathbb{R}^+$	$\mathbb{R}^d$

Fuente: (Belanche Muñoz, 2013)

### 3.5 Formulación y resolución del problema de optimización

Consideremos  $\mathbf{X} = [x_1^T, \dots, x_N^T]^T$  con  $\mathbf{X} \in \mathbb{R}^{N \times d}$  como la matriz de entrada de datos, la cual se debe dividir en  $K$  subconjuntos desagrupados, siendo  $x_i \in \mathbb{R}^d$  en punto dimensional de datos  $i$ -th  $d$ ,  $N$  el número de puntos de datos y  $K$  el número de grupos deseados. En el presente

trabajo se emplea el agrupamiento espectral de kernel (KSC)(C. Alzate & Suykens, 2010) que se basa en las máquinas vector de soporte – de los mínimos cuadrados (*LS-SVM least squares-Support Vector Machines*) y que se pueden ver como un análisis del componente principal ponderado del kernel, con formulación primaria dual en donde se asume el siguiente modelo de agrupamiento. Definamos  $e^{(l)} \in \mathbb{R}^N$  como el vector  $l$ -th de proyección como el modelo de variable latente en la forma:

$$e^{(l)} = \Phi \omega^{(l)} + b_l \mathbf{1}_N, l \in \{1, \dots, n_e\} \quad (2)$$

en donde  $\omega^{(l)} \in \mathbb{R}^{d_h}$  es el vector de ponderación  $l$ -th,  $b_l$  es un término de sesgo,  $n_e$  es el número de variables latentes consideradas, la notación  $\mathbf{1}_N$  representa un vector  $N$  todo unos, dimensional y la matriz  $\Phi = [\phi(x_1)^T, \dots, \phi(x_N)^T]^T$ ,  $\Phi \in \mathbb{R}^{N \times d_h}$ , es una representación de datos de alta dimensionalidad. Luego, el vector  $e^{(l)}$  representa las variables latentes desde un conjunto  $n_e$  de Indicadores binarios de agrupamiento obtenidos por  $\text{sign}(e^{(l)})$ , los cuales son más codificados con el fin de obtener los grupos resultantes  $K$ .

Por lo tanto, al apuntar a agrupar los datos de entrada y considerar unos mínimos cuadrados con formulación SVM (Joseph, Bart, & K, 2002) para la ecuación (2), podemos plantear, en términos matriciales, lo siguiente problema de optimización:

$$\max_{E, W, b} \frac{1}{2N} \text{tr}(E^T V E \Gamma) - \frac{1}{2} \text{tr}(W^T W) \quad (3)$$

$$\text{s. a. } E = \Phi W + \mathbf{1}_N \otimes \mathbf{b}^T \quad (4)$$

En donde  $\mathbf{b} = [b_1, \dots, b_{n_e}]$ ,  $\mathbf{b} \in \mathbb{R}^{n_e}$ ,  $\Gamma = \text{Diag}([\gamma_1, \dots, \gamma_{n_e}])$  es  $W = [\omega^{(1)}, \dots, \omega^{(n_e)}]$ ,  $W \in \mathbb{R}^{d_h \times n_e}$ ,  $\gamma_l \in \mathbb{R}^+$  el parámetro de regularización  $l$ -th,  $V \in \mathbb{R}^{N \times N}$  es

una matriz de ponderación para proyecciones y  $\mathbf{E} = [\mathbf{e}^{(1)}, \dots, \mathbf{e}^{(n_e)}]$ ,  $\mathbf{E} \in \mathbb{R}^{N \times n_e}$ . Las notaciones  $\text{tr}(\cdot)$  y  $\otimes$  representan el seguimiento y el producto Kronecker, respectivamente. El lagrangiano del problema ecuaciones (3) y (4) se puede expresar como se muestra a continuación:

$$\mathcal{L}(\mathbf{E}, \mathbf{W}, \Gamma, \mathbf{A}) = \frac{1}{2N} \text{tr}(\mathbf{E}^T \mathbf{V} \mathbf{E}) - \frac{1}{2} \text{tr}(\mathbf{W}^T \mathbf{W}) \text{tr}(\mathbf{A}^T (\mathbf{E} - \Phi \mathbf{W} - \mathbf{1}_N \otimes \mathbf{b}^T)) \quad (5)$$

En donde la matriz  $\mathbf{A} \in \mathbb{R}^{N \times n_e}$  está formada por los vectores multiplicadores lagrangianos, a saber,  $\mathbf{A} = [\boldsymbol{\alpha}^{(1)}, \dots, \boldsymbol{\alpha}^{(n_e)}]$ , y  $\boldsymbol{\alpha}^{(l)} \in \mathbb{R}^N$  el vector de multiplicadores lagrangianos  $l$ -th.

Mediante la resolución de las derivadas parciales de ecuación (5), se producen las siguientes condiciones de Karush-Kuhn-Tucker (KKT):

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \mathbf{E}} = 0 &\Rightarrow \mathbf{E} = N\mathbf{V}^{-1} \mathbf{A} \Gamma^{-1}, & \frac{\partial \mathcal{L}}{\partial \mathbf{W}} = 0 &\Rightarrow \mathbf{W} = \Phi^T \mathbf{A}, \\ \frac{\partial \mathcal{L}}{\partial \mathbf{A}} = 0 &\Rightarrow \mathbf{E} = \Phi \mathbf{W}, \quad \mathbf{y} & \frac{\partial \mathcal{L}}{\partial \mathbf{b}} = 0 &\Rightarrow \mathbf{b}^T \mathbf{1}_N = 0. \end{aligned}$$

Ahora, mediante la eliminación de las variables primales del problema inicial de las ecuaciones (3) y (4), se obtiene la siguiente solución dual basada en vectores propios:

$$\mathbf{A} \boldsymbol{\Lambda} = \mathbf{A} \mathbf{V} (\mathbf{I}_N + (\mathbf{1}_N \otimes \mathbf{b}^T) (\boldsymbol{\Omega} \boldsymbol{\Lambda})^{-1}) \boldsymbol{\Omega}$$

En donde  $\boldsymbol{\Lambda} = \text{Diag}(\lambda)$ ,  $\boldsymbol{\Lambda} \in \mathbb{R}^{N \times N}$ ,  $\lambda \in \mathbb{R}^N$  es el vector de valores propios con  $\lambda_l = N/\gamma_l$ ,  $\lambda_1 \in \mathbb{R}^+$  y  $\boldsymbol{\Omega} \in \mathbb{R}^{N \times N}$  es una matriz de kernel determinada, de forma que  $\Phi \Phi^T = \boldsymbol{\Omega}$ . En particular se emplea  $\mathbf{V} = \mathbf{D}^{-1}$  y ya que los  $K - 1$  vectores propios de  $\mathbf{A}$  son indicadores de la asignación de agrupamiento, se establece el valor  $n_e$  para que sea  $k - 1$  (C. Alzate & Suykens, 2010). Ya que la condición  $\mathbf{b}^T \mathbf{1}_N = 0$  se puede satisfacer centrando el vector  $\mathbf{b}$ , el

término de sesgo se escoge en la forma  $b_l = -1/(\mathbf{1}_N \mathbf{V} \mathbf{1}_N) \mathbf{1}_N^T \mathbf{V} \boldsymbol{\Omega} \boldsymbol{\alpha}^{(l)}$  Por consiguiente, la solución al problema de las ecuaciones (3) y (4) se reformula en el siguiente problema de vector propio:

$$\mathbf{A} \boldsymbol{\Lambda} = \mathbf{V} \mathbf{H} \boldsymbol{\Omega} \mathbf{A}$$

En donde la matriz  $\mathbf{H} \in \mathbb{R}^{N \times N}$  es la matriz de centrado que se define como  $\mathbf{H} = \mathbf{I}_N - 1/(\mathbf{1}_N^T \mathbf{V} \mathbf{1}_N) \mathbf{1}_N \mathbf{1}_N^T \mathbf{V}$ , ( $\mathbf{I}_N$  denota una matriz de identidad de  $N$  dimensión y  $\boldsymbol{\Omega} = [\Omega_{ij}]$ ,  $\boldsymbol{\Omega} \in \mathbb{R}^{N \times N}$ , siendo  $\Omega_{ij} = K(\mathbf{x}_i, \mathbf{x}_j)$ ,  $i, j \in 1, \dots, N$ ).

La notación  $K(\cdot, \cdot): \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$  representa la función kernel presentada.

De forma consecuente se puede calcular el conjunto de proyecciones como se muestra a continuación:

$$\mathbf{E} = \boldsymbol{\Omega} \mathbf{A} + \mathbf{1}_N \otimes \mathbf{b}^T$$

Finalmente, se codifican las proyecciones con el fin de determinar las asignaciones de agrupamiento.

Ya que cada agrupamiento se representa mediante un solo punto en el auto espacio  $K-1$ -dimensional, de manera que esos puntos solos se encuentren siempre en diferentes octantes, la codificación de los vectores propios se puede realizar considerando que dos puntos se encuentren en la misma agrupación cuando se ubiquen en el mismo octante en el espacio propio correspondiente (C. Alzate & Suykens, 2010).

De este modo se puede obtener un libro de códigos a partir de las filas de la matriz que contengan los vectores propio  $K-1$  preponderantes binarizados en las columnas, mediante el uso de  $\text{sign}(\mathbf{e}^{(l)})$ . Preponderantes binarizados en las columnas, mediante el uso de  $\tilde{\mathbf{E}} = \text{sign}(\mathbf{E})$  se

convierten en libro de códigos, siendo cada fila una palabra clave. Dentro del mismo cuadro, se pueden calcular las extensiones sin muestras para determinar la asignación de membresía del agrupamiento para los nuevos datos de prueba, de la siguiente manera. Definir  $\mathbf{z} \in \mathbb{R}^{n_e}$  como el vector de proyección de un punto de datos de prueba  $x_{test}$ .

Las proyecciones de prueba se pueden calcular como  $\mathbf{z} = \mathbf{A}^T \tilde{\boldsymbol{\Omega}}_{test} + \mathbf{b}$ , en donde  $\tilde{\boldsymbol{\Omega}}_{test} \in \mathbb{R}^{n_e}$  es el vector kernel, de forma que  $\tilde{\boldsymbol{\Omega}}_{test} = [\tilde{\Omega}_{test_1}, \dots, \tilde{\Omega}_{test_N}]^T$ , siendo  $\tilde{\Omega}_{test_i} = \mathcal{K}(x_i, x_{test})$ .

Una vez calculado el vector de proyección de prueba  $\mathbf{z}$ , se lleva a cabo una etapa de decodificación que consiste en realizar la comparación de las proyecciones binarizadas con respecto a las palabras clave del libro de códigos  $\tilde{\mathbf{E}}$  y en la asignación de membresía del agrupamiento con base en la distancia mínima de Hamming (Carlos Alzate & Suykens, 2010b).

**Algoritmo 1: Agrupamiento espectral de kernel:**  $[\mathbf{A}, \boldsymbol{\Lambda}, \tilde{\mathbf{E}}] = \text{KSC}(\mathbf{X}, \mathcal{K}(\cdot, \cdot), K)$

- 1: Entrada:  $\mathbf{X}, \mathcal{K}(\cdot, \cdot), K$
- 2: Forme la matriz del kernel  $\boldsymbol{\Omega}$  de forma que  $\Omega_{ij} = \mathcal{K}(y_i, x_i)$
- 3: Calcule la matriz  $\mathbf{H} = \mathbf{I}_N - 1/(\mathbf{1}_N^T \mathbf{V} \mathbf{1}_N) \mathbf{1}_N \mathbf{1}_N^T \mathbf{V}$  y  

$$b_l = -1/(\mathbf{1}_N^T \mathbf{V} \mathbf{1}_N) \mathbf{1}_N^T \mathbf{V} \boldsymbol{\Omega} \boldsymbol{\alpha}^{(l)}$$
- 4: Calcule la descomposición propia dual del problema:  $\mathbf{A} \boldsymbol{\Lambda} = \mathbf{V} \mathbf{H} \boldsymbol{\Omega} \mathbf{A}$
- 5: Determine  $\mathbf{E}$  a través de  $\mathbf{E} = \boldsymbol{\Omega} \mathbf{A} + \mathbf{1}_N \otimes \mathbf{b}^T$
- 6: Formar el libro de códigos de entrenamiento a través de un proceso de binarización, así:  $\tilde{\mathbf{E}} = \text{sgn}(\mathbf{E})$
- 8: Resultados:  $\mathbf{A}, \boldsymbol{\Lambda}, \tilde{\mathbf{E}}$

## CAPITULO IV

### MODELO PROPUESTO PARA EL ANÁLISIS DE DATOS DINÁMICOS

#### 4.1 Introducción

Los datos dinámicos o cambiantes en tiempo han tomado un sitio cada vez más importante en la investigación científica, principalmente en la automatización y el reconocimiento de patrones, la identificación de movimiento para la supervisión y la clasificación del movimiento de cuerpos. En este contexto surgieron los enfoques de análisis con base en un kernel y de matriz espectral como alternativas apropiadas.

El enfoque de la segmentación de movimiento con base en el KSC (KSC-MS) propuesto en el presente documento surge a partir de la combinación del KSC con una clasificación de relevancia para la selección de atributos. Dicha clasificación proviene de una definición de la relevancia de atributos orientada a la selección de un subconjunto de atributos basados en las propiedades espectrales del laplaciano de la matriz de datos, la cual se basa en una clasificación continua de atributos por medio de un problema de maximización de los mínimos cuadrado.

La maximización se realiza sobre un funcional de tal forma que las soluciones dispersas para los valores de la clasificación se obtengan mediante una descomposición espectral de una matriz no negativa. Lo que hace que este enfoque sea interesante dentro del contexto del análisis dinámico de datos, es la posibilidad de obtener un valor de clasificación para cada cuadro en el proceso del análisis de datos de variación de tiempo. Además, dicho enfoque de relevancia de atributos mide la relevancia de un subconjunto de atributos comparados con su influencia en la asignación de agrupamiento. El objeto del KSC-MS es realizar el seguimiento de muestras móviles o de cuadros, contrastando cada uno de ellos con un valor significativo único. Este

enfoque puede ser de ayuda debido a su naturaleza sin supervisión, ya que en la práctica el etiquetado a menudo no se encuentra disponible para la aplicación de rastreo de movimiento.

El enfoque propuesto opera de la siguiente forma: Inicia agrupando los datos de entrada por medio del KSC con un número de grupos, establecido de forma manual y una función de kernel determinada. Luego, se proyecta la representación dimensional alta de datos de entrada de forma lineal, con el fin de aplicar un proceso de clasificación de relevancia de muestras, tal como se propone en (Wolf & Shashua, 2005). Está probado que la matriz de proyección es la misma que la matriz del vector propio proporcionada por el KSC.

Finalmente se obtiene un vector de seguimiento por medio de una combinación lineal de dichos vectores propios de forma similar a la descrita en (Molina-Giraldo, Álvarez-Meza, Peluffo-Ordoñez, & Castellanos-Domínguez, 2012)

#### **4.2 Datos cambiantes en el tiempo**

Datos que cambian sus valores o información en intervalos de tiempo muy pequeños, es decir cambios rápidos en tiempo. Los datos en movimiento proporcionan el conjunto más robusto de capacidades analíticas para grandes volúmenes de datos, por lo que se trata de datos en tiempo real eso conlleva a que se lo que se tiene son datos dinámicos. Por tanto, se tendrá una analítica de información estructurada como video, imágenes, dispositivos móviles, redes sociales, sensores, entre otros.

#### **4.3 Matrices kernel acumuladas (agrupamiento espectral de kernel)**

Se considera una secuencia de la matriz de datos  $\{\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(N_f)}\}$ , en donde  $N_f$  es el número de cuadros y  $\mathbf{X}^{(t)} = [\mathbf{x}_1^{(t)T}, \dots, \mathbf{x}_N^{(t)T}]^T$  es la matriz de datos relacionada con la instancia de tiempo  $t$  en dimensión  $N \times d$ . También se considera la matriz de cuadro  $\mathcal{X} \in$

$\mathbb{R}^{N_f \times N_d}$  la cual está formada de tal manera que cada fila es un cuadro, permitiendo que  $\tilde{x}_t \in \mathbb{R}^{N_d}$  sea la vectorización de coordenadas representando el cuadro t-th. En otras palabras,  $\mathbf{X} = [\tilde{x}_1^T, \dots, \tilde{x}_{N_f}^T]^T$  y  $\tilde{x}_t = \text{vec}(X^{(t)})$ .

La matriz kernel correspondiente se puede expresar como  $\tilde{\Omega} \in \mathbb{R}^{N_f \times N_f}$  de forma que  $\tilde{\Omega}_{ij} = \mathcal{K}(\tilde{x}_i, \tilde{x}_j)$

#### 4.4 Vector de seguimiento

Al igual que el análisis de relevancia explicado en (Peluffo Ordoñez, Lee, Verleysen, Rodriguez, & Castellanos-Dominguez, 2015), en el cual se presenta un funcional relacionado con una matriz no negativa, se plantea un problema de optimización con la diferencia de que el enfoque considerado es la obtención de valores de clasificación para muestras en lugar de atributos.

La representación de la matriz de gran dimensionamiento  $\tilde{\Phi} \in \mathbb{R}^{N_f \times d_h}$  puede ser expresada como

$$\tilde{\Phi} = [\phi(\tilde{x}_1)^T, \dots, \phi(\tilde{x}_{N_f})^T]^T$$

en donde  $\phi(\cdot): \mathbb{R}^{N_d} \rightarrow \mathbb{R}^{d_h}$ . Se asume una proyección lineal en la forma  $\mathbf{Z} = \tilde{\Phi}^T \mathbf{U}$ , donde  $\mathbf{U}$  es una matriz ortonormal en dimensión  $N_f \times N_f$ . Además, se asume una representación de menor clasificación de  $\mathbf{Z}$  en la forma  $\hat{\mathbf{Z}} = \tilde{\Phi}^T \hat{\mathbf{U}}$ , en donde  $\hat{\mathbf{U}}$ , es en dimensión  $N_f \times c$  ( $c < N_f$ ). Entonces, se puede escribir un problema de maximización de energía de la siguiente manera:

$$\max_{\hat{\mathbf{U}}} \text{tr}(\hat{\mathbf{U}}^T \tilde{\Omega} \hat{\mathbf{U}})$$

$$s. a. \quad \hat{\mathbf{U}}^T \hat{\mathbf{U}} = \mathbf{I}_c$$

De hecho, mediante el uso de la estrategia del kernel tenemos

$$\text{tr}(\hat{\mathbf{Z}}^T \hat{\mathbf{Z}}) = \text{tr}(\hat{\mathbf{U}}^T \tilde{\Phi} \tilde{\Phi}^T \hat{\mathbf{U}}) = \text{tr}(\hat{\mathbf{U}}^T \tilde{\Omega} \hat{\mathbf{U}})$$

Luego, recordando el problema dual KSC de la ecuación  $\mathbf{A}\mathbf{A} = \mathbf{V}\mathbf{H}\Omega\mathbf{A}$ , se puede escribir la matriz de centrado para la matriz de marco  $\mathcal{X}$ , de la siguiente forma:

$$\tilde{\mathbf{H}} = \mathbf{I}_{N_f} - \frac{1}{\mathbf{1}_{N_f}^T \tilde{\mathbf{V}} \mathbf{1}_{N_f}} \mathbf{1}_{N_f} \mathbf{1}_{N_f}^T \tilde{\mathbf{V}}$$

En donde se escoge a  $\tilde{\mathbf{V}}$  como la matriz de grado proporcionada por  $\tilde{\mathbf{D}} = \text{Diag}(\tilde{\Omega} \mathbf{1}_{N_f})$ .

Luego, mediante la normalización del grado relacionado y el centrado de tanto  $\mathbf{Z}$  como de  $\hat{\mathbf{Z}}$ , lo cual representa multiplicar previamente  $\tilde{\Phi}$  por  $\tilde{\mathbf{L}}\tilde{\mathbf{V}}^{-1/2}$ , se puede inferir que

$$\text{tr}(\hat{\mathbf{U}}^T \tilde{\Omega} \hat{\mathbf{U}}) = \sum_{t=1}^c \tilde{\lambda}_t$$

en donde  $\tilde{\mathbf{L}}$  proviene de la descomposición de Cholesky de  $\tilde{\mathbf{H}}$  de tal forma que  $\tilde{\mathbf{L}}^T \tilde{\mathbf{L}} = \tilde{\mathbf{H}}$  y  $\tilde{\lambda}_l$  es el autovalor  $l$ -th obtenido por el KSC al aplicarlo sobre  $\mathcal{X}$  con un número determinado de agrupamientos  $\tilde{\mathbf{K}}$ . Por lo tanto, una solución viable del problema es  $\mathbf{U} = \tilde{\mathbf{A}}$ , siendo  $\tilde{\mathbf{A}} = [\tilde{\alpha}^{(1)}, \dots, \tilde{\alpha}^{(c)}]$  la matriz de vector propio correspondiente. Por consiguiente,  $c$  es el mismo número de vectores  $\tilde{n}_e$  de soporte que se han considerado.

De forma similar al enfoque MKL explicado en (Ramirez-Giraldo, Molina-Giraldo, Álvarez-Meza, Daza-Santacoloma, & Castellanos-Domínguez, 2012), se presenta un vector de seguimiento  $\boldsymbol{\eta} \in \mathbb{R}^{N_f}$  como la solución de la minimización del término dispar obtenido por

$$\|\tilde{\Phi} - \hat{\Phi}\|_F^2 \quad (8)$$

Sujeto a algunas condiciones de ortogonalidad, siendo  $\hat{\Phi}$  una representación de menor clasificación de  $\tilde{\Phi}$ . Entonces, el vector de clasificación se puede calcular mediante:

$$\eta = \sum_{\ell=1}^{\tilde{n}_e} \tilde{\lambda}_\ell \tilde{\alpha}^{(\ell)} \circ \tilde{\alpha}^{(\ell)}$$

en donde  $\circ$  denota el producto Hadamard (relacionado con los elementos). Por consiguiente, el factor de clasificación  $\eta_i$  es un solo valor que representa un cuadro único en una secuencia. La notación  $\tilde{\alpha}$  significa que la variable  $\alpha$  se encuentra relacionada con  $\tilde{\Omega}$ .

Ya que  $\eta$  proviene de una combinación lineal de los auto vectores cuadráticos, siendo los coeficientes los valores propios correspondientes, los cuales son positivos, la positividad del vector de seguimiento está garantizada. Además, se puede normalizar el vector multiplicándolo por  $1/\max|\eta|$  con el propósito de mantener las entradas de  $\eta$  clasificadas en el intervalo  $[0, 1]$ .

A continuación, se describe el enfoque de codificación propuesto para comparar el seguimiento de cuadros dado por el enfoque original. Este enfoque de codificación está inspirado en el procedimiento explicado en (Carlos Alzate & Suykens, 2010a).

Como se discutió en (Oña-Rocha et al., 2017), (Peluffo-Ordóñez et al., 2013), dadas las condiciones KKT aplicadas a la formulación dual del problema KSC, los grupos pueden ser reconocidos directamente, como la ubicación geométrica de los puntos de datos proyectados  $E$  en cada orante individual representa un clúster único. En otras palabras, los clusters pueden ser codificados con indicadores en binario expresados en la ecuación  $\tilde{E} = \text{sgn}(E)$ ,

consecuentemente, podemos obtener valores nítidos de los indicadores de grupo como las filas  $\tilde{e}_i (\forall i, i \in \{1, \dots, N\})$  de la matriz  $\tilde{\mathbf{E}}$  se puede convertir directamente de números binarios a decimales. Sin embargo, aquí es preferible restringir dicha conversión como el número máximo resultante será el número esperado de clusters. Entonces, los códigos binarios se convierten en números decimales según el orden de aparición, de 1 a  $K$  para llegar al vector de seguimiento codificado  $\tilde{\boldsymbol{\eta}} \in \mathbb{R}^d$ .

Por lo tanto, para ejemplificar el enfoque de codificación, se considera el siguiente ejemplo con  $n_e = 4$ :

$$\mathbf{E} = \begin{pmatrix} 2.7 & 2.1 & -0.4 & 4.1 \\ 4.3 & 2.5 & -0.5 & -1.3 \\ 2.3 & 1.5 & -0.5 & -1.3 \\ 1.3 & -1.5 & -0.5 & 2.3 \\ 1.3 & 2.5 & -0.5 & 4.3 \end{pmatrix}$$

Obteniéndose una matriz codificada de la forma:

$$\tilde{\mathbf{E}} = \begin{pmatrix} 1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & 1 \end{pmatrix}$$

por lo tanto, su  $\tilde{\boldsymbol{\eta}}$  será correspondientemente dada por:

$$\tilde{\boldsymbol{\eta}} = \begin{pmatrix} 1 \\ 1 \\ 2 \\ 3 \\ 1 \end{pmatrix}$$

Los pasos para realizar el cálculo de los vectores de la segmentación de movimiento con base en el KSC (KSC-MS) se resumen en el Algoritmo 2.

**Algoritmo 2:** KSCT:  $[\eta, \tilde{\eta}] = \text{KSCT}(\{\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(N_f-1)}\}, K)$

*Entrada:* N° de grupos  $K$ , una secuencia de cuadro  $\{\mathcal{X}^{(1)}, \dots, \mathcal{X}^{(N_f)}\}$ , una función del kernel  $\mathcal{K}(\cdot, \cdot)$

1. Formar la matriz del cuadro  $\mathbf{X} = [\mathbf{x}_1^T, \dots, \mathbf{x}_{N_f}^T]$  de forma que  $\mathbf{x}_t = \text{vec}(\mathcal{X}^{(t)})$
2. Aplicar KSC sobre  $\mathbf{X}$  con  $K$  para obtener los valores propios  $\mathbf{\Lambda} = \text{Diag}(\lambda_1, \dots, \lambda_{\tilde{n}_e})$  y los vector propios  $\mathbf{A} = [\alpha^{(1)}, \dots, \alpha^{(\tilde{n}_e)}]$ :  $\mathbf{A}, \mathbf{\Lambda}, \tilde{\mathbf{E}} = \text{KSC}(\mathbf{X}, \mathcal{K}(\cdot, \cdot), K)$
3. Calcular  $\eta = \sum_{\ell=1}^{\tilde{n}_e} \lambda_{\ell} \alpha^{(\ell)} \circ \alpha^{(\ell)}$  con  $\tilde{n}_e = K - 1$
4. Normalizar  $\eta$  como  $\eta \leftarrow \eta / \max |\eta|$
5. Obtener  $\tilde{\eta}$  codificando en números decimales  $\tilde{\mathbf{E}}$

*Resultados:* Vectores de seguimiento  $\eta, \tilde{\eta}$

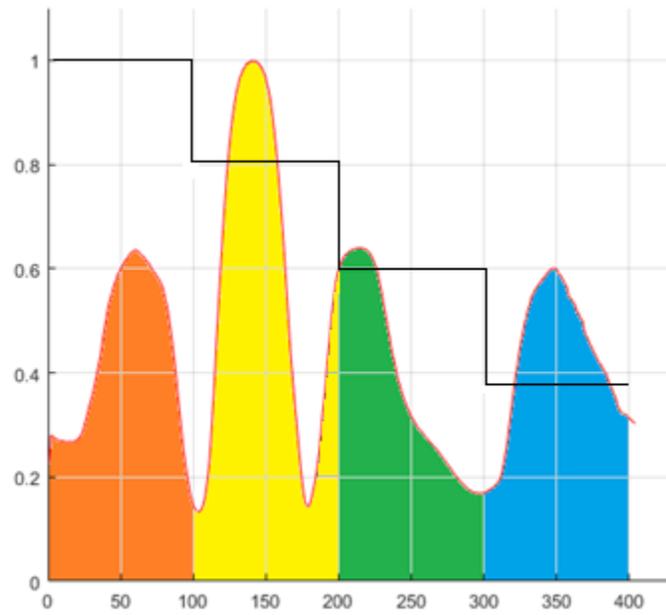
#### 4.5 Identificación de segmentos de movimiento

Cuando se tiene movimiento de objetos el vector de seguimiento puede describir dicho movimiento con una señal como se muestra en la

**Figura 4**, y podemos segmentarla en cuatro partes, la de color naranja que está en el intervalo 0 a 100, la de color amarillo entre 100 y 200, la de color verde en el intervalo 200 a 300 y finalmente la de color azul que esta entre 300 y 400.

El segmento de 0 a 100, representa un movimiento del objeto y en el punto 100 se produce un cambio de movimiento, y luego sigue el movimiento en el intervalo 100-200 y se tiene otro cambio de movimiento en el punto 200 y así sucesivamente se producen cambio de movimientos en los puntos 300 y 400.

Estos instantes de cambio de movimiento se pueden llevar a una interpretación de señal discreta, indicando una codificación que cada movimiento es un pulso, logrando con esto saber exactamente cuando se produce el cambio de movimiento, ya que en forma analógica las pequeñas variaciones o ruido no permite determinar el instante exacto de cambio de movimiento



*Figura 4.* Explicación gráfica de la segmentación a partir del vector de seguimiento original.

## CAPITULO V

### RESULTADOS EXPERIMENTALES Y DISCUSIÓN

#### 5.1 Introducción

En este capítulo se describe el marco experimental sección 5.2 que hace referencia a las bases de datos utilizadas para los diferentes experimentos y la manera de su utilización, la comparación entre los métodos y la medida de desempeño de los mismos. En la sección 5.3 se describen los resultados de los experimentos realizados y finalmente la discusión en el numeral 5.4.

#### 5.2 Marco experimental

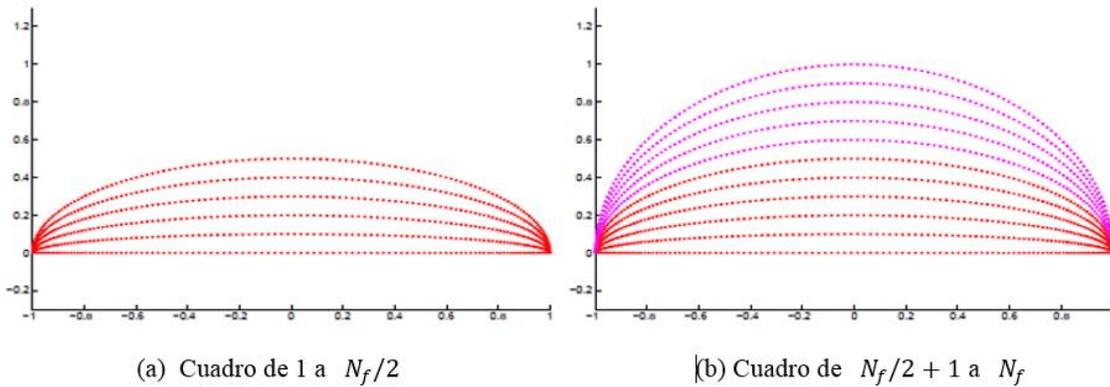
##### 5.2.1 Bases de datos para realizar las pruebas

###### 5.2.1.1 Datos sintéticos: Curvas de nivel

Con el fin de ilustrar el desempeño del vector de seguimiento, consideremos al siguiente modelo para demostración de una curva de movimiento. En la instancia de tiempo  $t$ , el efecto de una curva 2-D desplazándose en un arco en ascenso, se emula mediante las coordenadas  $X Y$  :

$$\mathbf{x}^{(t)} = \begin{pmatrix} |\cos(2\pi\tau)|^T \\ -|\cos(2\pi\tau)|^T \end{pmatrix} \quad y \quad \mathbf{y}^{(t)} = \begin{pmatrix} |t \sin(2\pi\tau)|^T \\ |t \sin(2\pi\tau)|^T \end{pmatrix}$$

En donde cada entrada del vector  $\tau$  es  $\tau_n = n/N$  donde  $n \in \{1, \dots, N/2\}$  siendo  $N$  número de muestras por cuadro. Entonces se puede formar la matriz de datos correspondiente  $\mathbf{X}^{(t)} \in \mathbb{R}^{N \times 2}$  como  $\mathbf{X}^{(t)} = [\mathbf{x}^{(t)}, \mathbf{y}^{(t)}]$  igual que la secuencia del cuadro  $\{\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(N_f)}\}$  Entonces el efecto del video hasta un cierto cuadro  $T$  se realiza manteniendo los cuadros previos para mostrar el seguimiento del trayecto seguido por la curva. La **Figura 5** representa el efecto del movimiento del arco, al considerar  $N = 100$ , y  $N_f = 10$



**Figura 5.** 2-D curva de movimiento

### 5.2.1.2 Datos reales: Objetos girando sobre su propio eje

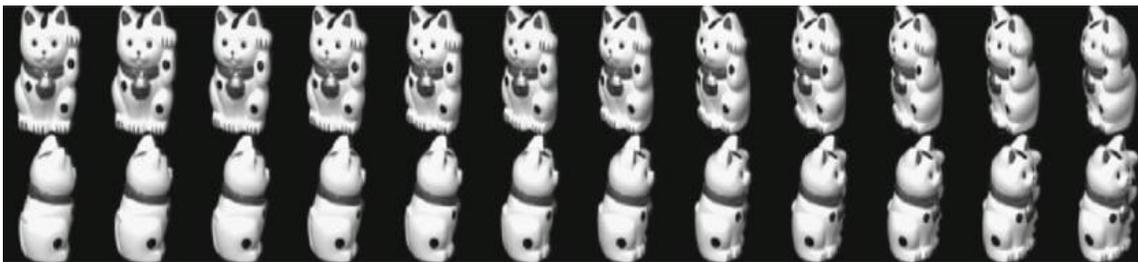
Para llevar a cabo de mejor manera los experimentos, se utilizó un objeto de la conocida base de datos *Columbia Object Image Library* (COIL 20) obtenida (Nene, Nayar, & Murase, 1996), que es un banco de imágenes de 20 objetos a nivel de grises como se ve en la **Figura 6**. Base de datos Coil20y que cada objeto a su vez consta de 72 imágenes.



**Figura 6.** Base de datos Coil20

Fuente: (S. A. Nene, 1996)

Los objetos están colocados en diferentes ángulos (72) que giran a cada 5 grados. Específicamente, seleccionó el objeto # 4 como se muestra en la **Figura 7**. Los cuadros del objeto #4 de la base de datos COIL20 (Jun Shi, Jiang, & Feng, 2014). Las 72 imágenes (un ángulo / pose), tienen el tamaño 128 x 128 píxeles, que en primer lugar se redimensionan al 50%, dando como resultado final imágenes en RGB expresadas como:  $\mathcal{X}^{(t)} \in \mathbb{R}^{64 \times 64}$ , siendo  $t \in \{1, \dots, 72\}$ . Posteriormente, una matriz de datos es formada por la vectorización de las imágenes en RGB. Por lo tanto, el número de datos es  $N = 64 \times 64 \times 3 = 12288$ , así como el número de variables es  $d = 72$  (siendo el mismo número que  $N_f$ ), lo que significa que la matriz de datos a agrupar es  $X \in \mathbb{R}^{12288 \times 72}$ .



**Figura 7.** Los cuadros del objeto #4 de la base de datos COIL20

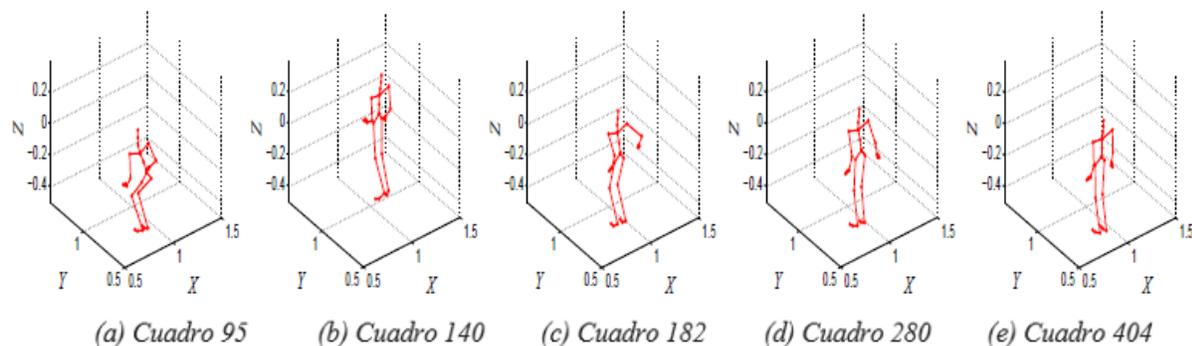
*Fuente:* (Jun Shi, 2013)

### 5.2.1.3 Aplicación del método KSC en el caso: Salto vertical de humanos

Algunos experimentos se llevan a cabo considerando dos bases de datos: Captura de Movimiento y un Gaussiano artificial de tres movimientos. El seguimiento se evalúa con respecto al rendimiento de agrupamiento por KSC y kernel K-means (Zha, He, Ding, Gu, & Simon, 2002).

Para cuantificar el rendimiento del agrupamiento, se utilizan información mutua normalizada y métricas de índice aleatorio ajustadas.

Los datos utilizados en este experimento se obtuvieron de la base mocap.cs.cmu.edu. La base de datos fue creada con fondos de NSF ETA-OI96217. Dicha base de datos se conoce como Graphics Lab Motion Capture Database de la Universidad Carnegie Mellon. Aquí, se llama captura de movimiento. Para experimentos, se uso el número de prueba 4 (02\_04) de la base de datos de captura de movimiento, en particular, el sujeto #2 (salto, balance). Cada cuadro tiene un tamaño de  $38 \times 3$  que representa las coordenadas X, Y y Z de los puntos del cuerpo del sujeto, por lo tanto, cada  $\mathbf{X}^{(t)}$  es de 114 dimensiones. Dado que se consideran 484 cuadros, la matriz de datos en el momento en que  $t$  es  $\mathbf{X}^{(t)}$  tiene el tamaño  $484 \times 114$ . Algunos cuadros de muestra están en la **Figura 8**.



**Figura 8.** Sujeto # 2 de la Base de Datos captura de movimientos

### 5.2.2 Métodos de comparación

Todos los experimentos se realizan bajo parámetros iniciales específicos, a saber, el Número de clusters  $K$  y función del kernel. Para la base de datos Motion Caption (*Sujeto # 2*), el parámetro  $K$  se establece en 5. En caso de movimiento gaussiano, empíricamente determina que  $K = 4$ . El

kernel considerado para ambos casos es el escalado local. Kernel gaussiano (Zelnik-Manor & Perona, 2005) definido como:

$$\Omega_{ij} = \exp(-\|x_i - x_j\|_2^2 / (\sigma_i \sigma_j))$$

Donde  $\|\cdot\|$  denota La norma euclidiana y el parámetro de escala  $\sigma_i$  se eligen como la distancia euclidiana entre la muestra  $x_i$  y su correspondiente m-ésimo vecino más cercano.

El parámetro libre  $m$  se establece empíricamente variándolo dentro de un intervalo y luego es elegido como aquel que alcanza el mayor valor de criterio de Fisher. En el caso de Motion Caption, obtenemos  $m = 35$ ; mientras que  $m = 10$  para el movimiento gaussiano. Para propósitos de comparación, las kernel K-means (KKM) y Min-Cuts también se consideran (Guo, Zheng, Xie, & Hao, 2012).

### 5.2.3 Medidas de desempeño

El rendimiento de la agrupación se cuantifica mediante dos métricas: Normalizada Información Mutua (NMI) (Normalized Mutual Information) (Strehl & Ghosh, 2002) e Índice Aleatorio Ajustado (ARI) (Adjusted Random Index) (Hubert & Arabie, 1985). Ambos los valores de retorno de las métricas se ubicaron en el intervalo  $[0, 1]$ , siendo el más cercano a 1 cuando mejor es el rendimiento de agrupamiento.

El efecto de seguimiento también se puede verificar en la **Tabla 2** En términos de las medidas consideradas, podemos observar que KSC supera a KKM y Min-Cuts, lo que significa que el método de agrupamiento podría ser coherente con el procedimiento de seguimiento. A este respecto, KSC es el más adecuado ya que  $\eta$  proviene de los vectores propios.

**Tabla 2***Rendimiento de agrupamiento para Sujeto #2 en términos de medidas métricas*

METODO DE AGRUPAMIENTO			
MEDIDA	KSC	KKM	Min Cuts
NMI	0.8410	0.6982	0.6974
ARI	0.8163	0.5732	0.5400

Fuente: (D. Peluffo-Ordóñez, García-Vega., 2013)

### 5.3 Resultados

#### 5.3.1 Datos sintéticos: Curvas de nivel

Para mostrar la capacidad de nuestro enfoque para segmentar movimientos de forma automática dentro de un cuadro dinámico, se realizan experimentos en un ejemplo de modelo para demostración de curva de movimiento. Como resultado, se demostró que nuestro enfoque puede determinar el inicio y el final de los movimientos al igual que su robustez ante los medios ruidosos.

El agrupamiento se realiza mediante el uso del KSC con el objeto de identificar dos movimientos naturales o agrupamientos  $\tilde{K} = 2$ . Para la ejecución del agrupamiento, se utiliza el KSC con un kernel Gaussiano adaptado  $K(\cdot, \cdot)$  mediante la selección del 7th ( $m = 7$ ) vecino más cercano como parámetro de adaptación. Toda la matriz del cuadro se utiliza para la captación, entonces  $\tilde{\mathbf{q}}_{\text{train}}$  son las asignaciones de agrupación que darán color a los cuadros de acuerdo con los grupos base, siendo

$$\tilde{\mathbf{q}}_{\text{train}} = \text{KSC}(\mathcal{X}, K(\cdot, \cdot), \tilde{K}) \quad (6)$$

Para incorporar ruido al modelo de la curva de movimiento, se considera la aplicación de ruido adicional sobre la coordenada Y en la forma  $v\mathbf{n}$ , en donde  $v$  es la cantidad de ruido y  $\mathbf{n} \in \mathbb{R}^N$  es el ruido introducido siguiendo una distribución gaussiana  $\mathbf{n} \sim N(0, 1)$ , así:

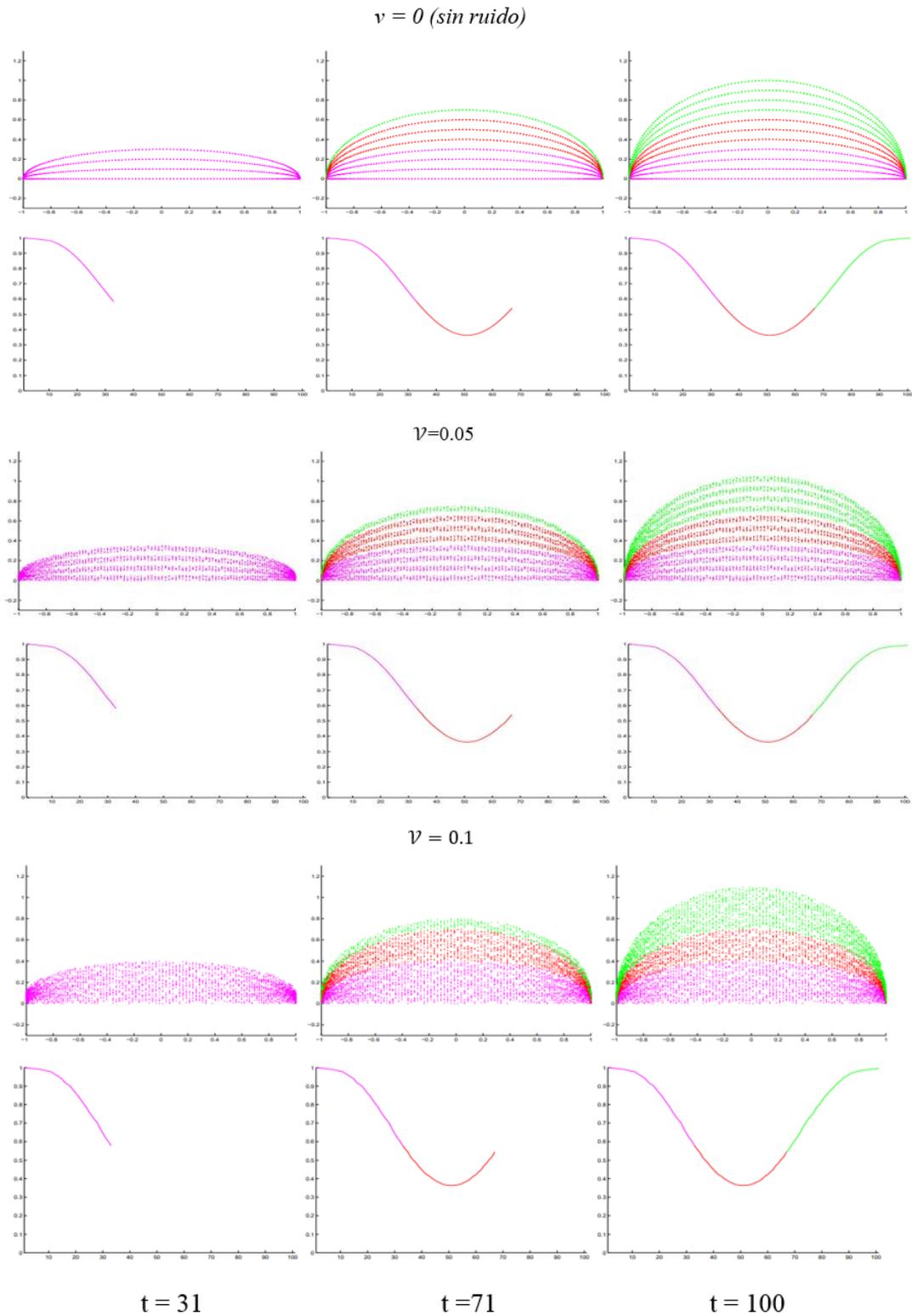
$$\mathbf{y}_n^{(t)} = \mathbf{y} + \mathbf{v}\mathbf{n} \quad (12)$$

Algunos ejemplos del desempeño de seguimiento se muestran en las *Figura 9* y *Figura 10* al establecer el número de cuadros para que sea  $N_f = 100$  y el número de muestras por cuadro para que sea  $N = 100$ , representando dos experimentos cambiando el número de grupos y condiciones de ruido. La *Figura 9* muestra el comportamiento del vector de seguimiento para el agrupamiento de la matriz del cuadro en tres agrupamientos ( $\tilde{K} = 3$ ) mientras que la *Figura 10* es para  $\tilde{K} = 4$ .

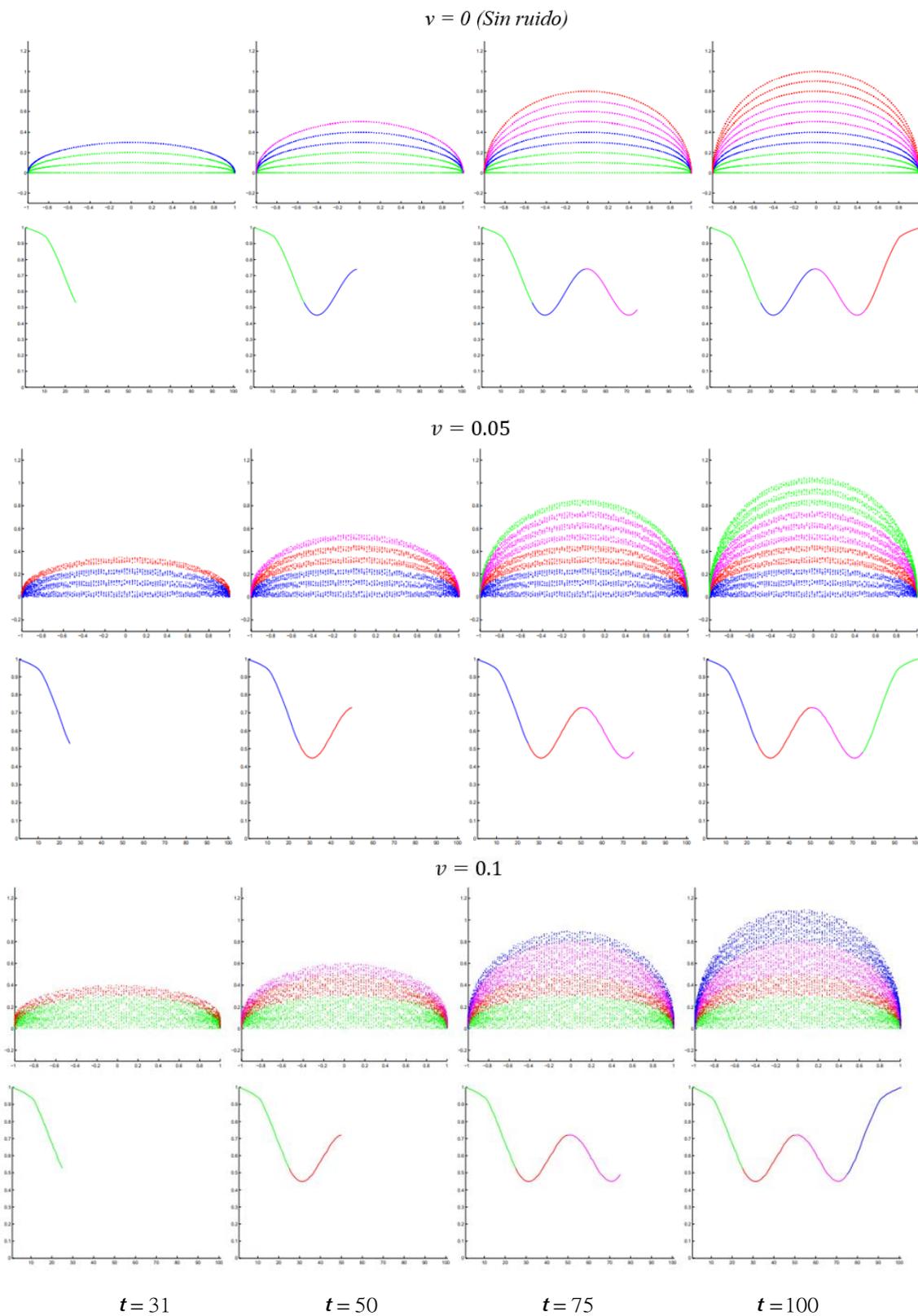
El vector de seguimiento se calcula tanto en un medio ruidoso como no ruidoso mediante la variación de la cantidad de ruido  $v$  al asumir que la matriz de cuadro  $X$  se debe separar en agrupaciones  $\tilde{K}$ . En ambas figuras y para cada cantidad de ruido considerada, la secuencia agrupada en la instancia de tiempo  $t$  se muestra en la fila superior al igual que el comportamiento del vector correspondiente en la fila inferior.

Con el fin de evaluar la robustez de este enfoque, se evalúa el vector de seguimiento sobre la cantidad de ruido  $v \in \{0.05, 0.1\}$ . También se considera el caso sin ruido ( $v = 0$ ). Para propósito de visualización, se seleccionan algunos cuadros significativos de ambos experimentos. Para el primer experimento los cuadros seleccionados son  $t \in \{31, 71, 100\}$ . De igual forma se seleccionaron los cuadros  $t \in \{31, 50, 75, 100\}$  para el segundo experimento.

Además, con el fin de observar el comportamiento del vector de seguimiento entre los agrupamientos, se agrupan las matrices del cuadro con el KSC, tal como se muestra en la ecuación (6) considerando las mismas condiciones mencionadas previamente.



**Figura 9.** Agrupamiento de la curva de movimiento 2-D en agrupamientos  $\tilde{K} = 3$  con cuadros  $N_f = 100$  y  $N = 100$  muestras por cuadro



**Figura 10.** Agrupamiento de la curva de movimiento 2-D en agrupamientos  $\tilde{K}$  con cuadros  $N_f = 100$  y  $N = 100$  muestras por cuadro

### 5.3.2 Giro de objeto 360 grados

El elemento seleccionado objeto # 4 de la base COIL20, se utilizó para realizar el experimento, para ello se estableció la configuración del cluster y kernel, definiendo que el número de clusters sea  $K = 4$ . La función kernel considerada es el kernel Gaussiano

convencional definido como  $\Omega_{ij} = \exp\left(-\frac{\|x_i - x_j\|_2^2}{2\sigma^2}\right)$ , donde  $\|\cdot\|$  denota la norma Euclidiana y

le parámetro de escala  $\sigma$  se define empíricamente como 30.

Para analizar la secuencia de cuadros organizados en la matriz  $\mathbf{X}$ , primero aplicamos KSC.

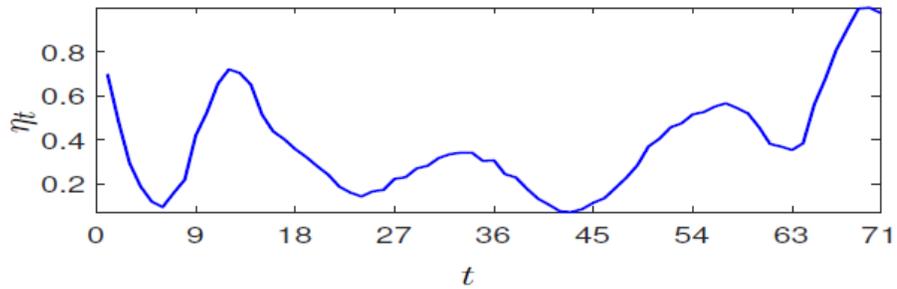
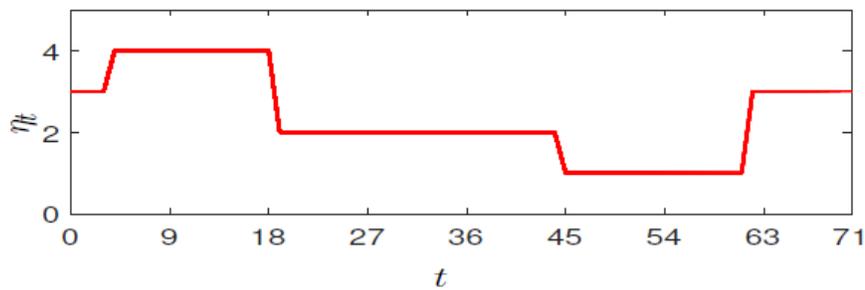
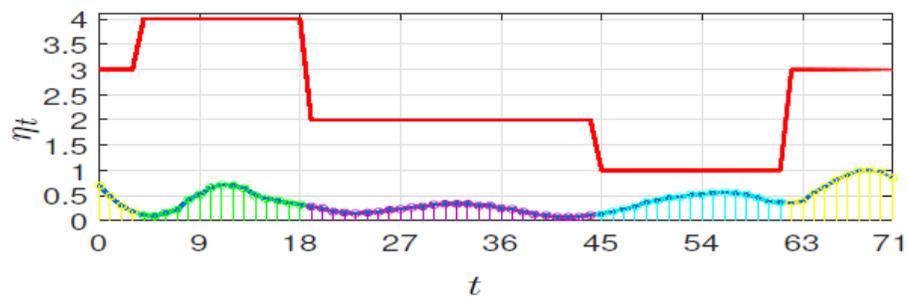
Luego, con los resultados, el vector  $\eta$  es calculado usando la ecuación:

$$\eta = \sum_{\ell=1}^{\tilde{n}_e} \tilde{\lambda}_\ell \tilde{\alpha}^{(\ell)} \circ \tilde{\alpha}^{(\ell)}.$$

En la **Figura 11** se puede ver el proceso del comportamiento dinámico capturado por el seguimiento basado en KSC, como se muestra en **Figura 11** (a) y **Figura 11**. Trazado de los vectores *original* y *codificado* (b), se tiene el trazado del vector de seguimiento original  $\eta$  y la versión codificada  $\tilde{\eta}$  respectivamente.

El trazado de vectores  $\eta$  y  $\tilde{\eta}$  es a lo largo de los 72 cuadros para el objeto #4 de la base de datos COIL20. En la **Figura 11** (c) se observa el vector de etiquetado de referencia, que se obtiene directamente de los valores de  $\tilde{\eta}$ , es la representación sobrepuesta de los vectores  $\eta$  y  $\tilde{\eta}$ , mientras que el área bajo la curva que esta coloreada se resalta el etiquetado basado en la etapa de movimiento con respecto a  $\tilde{\eta}$ .

Del trazado de  $\eta$  se puede ver que su forma es multimodal. Para comparar el vector  $\eta$  con  $\tilde{\eta}$ , se puede notar fácilmente que cada modo del trazado corresponde a un grupo diferente de grupo (cluster), es decir una etapa de movimiento en el contexto de análisis de video.

(a) Vector de seguimiento  $\eta$  original(b) Vector de seguimiento  $\tilde{\eta}$  codificado

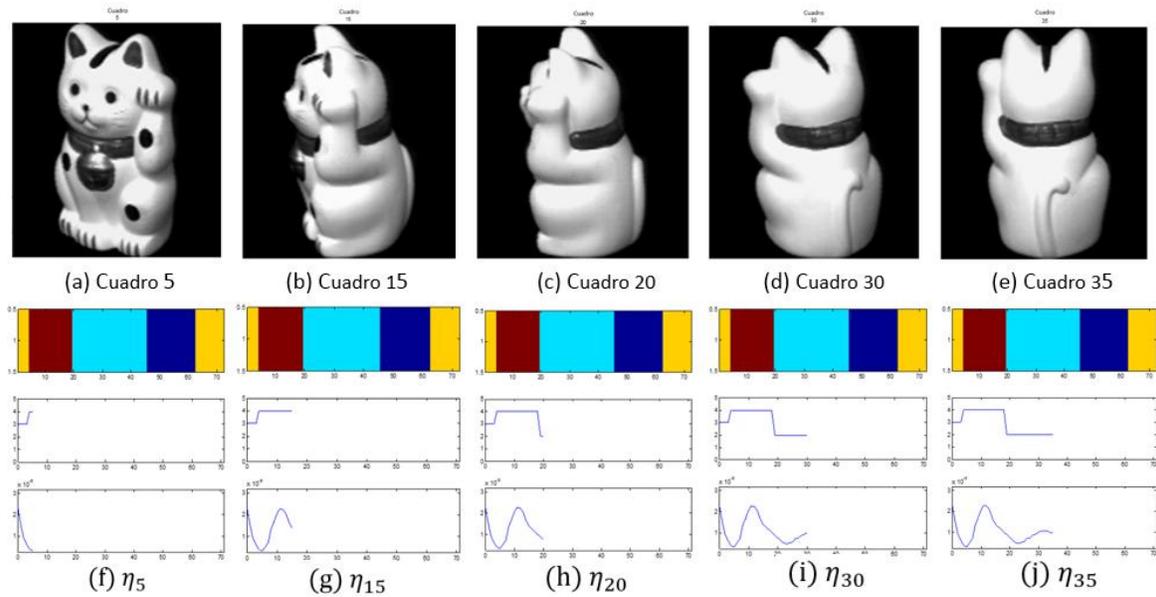
(c) Niveles establecidos en base a los vectores codificados

**Figura 11.** Trazado de los vectores original y codificado

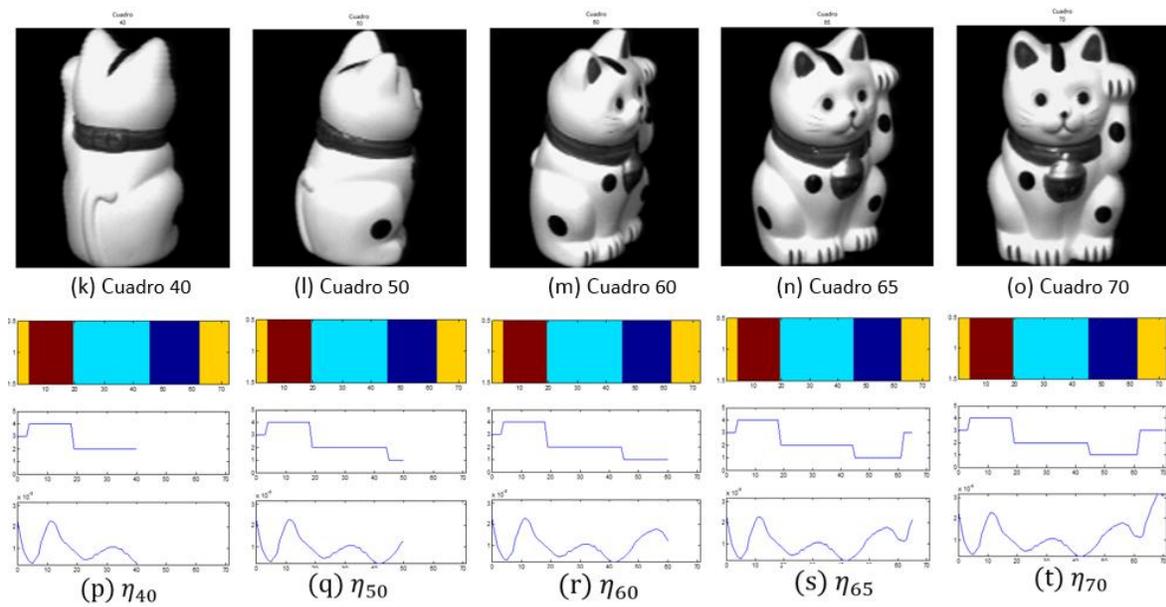
Dicha correspondencia puede atribuirse al hecho de que los vectores propios  $\alpha^{(l)}$  señalan la dirección donde las muestra exhiben más variabilidad medida en términos del producto interno generalizado  $(\Phi^T \Phi)$ . En esta conexión, las funciones kernel tiene lugar y permiten la estimación del interior de espacios de representación de alta dimensión, donde se suponen agrupaciones

resultantes para ser linealmente separables. La conexión directa entre el vector de seguimiento  $\eta$  y la partición de los movimientos naturales del Objeto #4 puede ser apreciado claramente en la **Figura 12** y **Figura 13** donde la fila superior muestra cuadros representativos por cluster mientras

qu



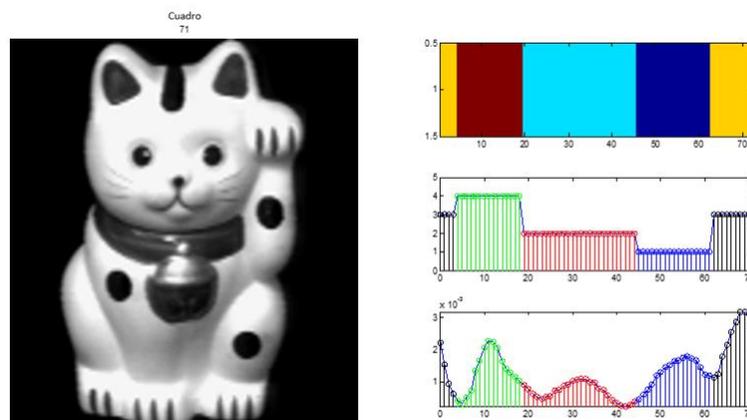
**Figura 12.** Objeto #4 rastreo de cuadros originales (5, 15, 20, 30, 35 y rastreo de vectores



**Figura 13.** Objeto #4 rastreo de cuadros originales (40, 50, 60, 65, 70) y rastreo de vectores

Como se notó, cada modo entre inflexiones forma una curva en el trazado, los que significa que ha aparecido otro grupo natural dentro de la secuencia. Dicha división de grupos puede incluso ser determinada mediante simple inspección. Junto al vector codificado permite validar la premisa de que el vector  $\eta$  es capaz de dividir la secuencia de cuadros en etapas de movimiento natural (grupos), cuando los ajustes de agrupamiento son apropiados.

En la **Figura 14** se tiene el ultimo cuadro, es equivalente al cuadro 1, después de haber girado 360 grados, se puede observar que en la gama de colores de la parte superior derecha de la imagen, el color amarillo representa cuando el objeto esta de frente y gira hacia la izquierda, el color café es cuando el objeto gira y se encuentra de lado izquierdo, el color celeste es el giro cuando se encuentra de espaldas, el azul representa el giro hacia el lado derecho y finalmente el amarillo cuando vuelve al frente.. En la segunda fila está el trazado de vector  $\tilde{\eta}$  es decir la señal codificada y en la tercera franja la representación del trazado del vector  $\eta$  a lo largo de los 72 cuadros.

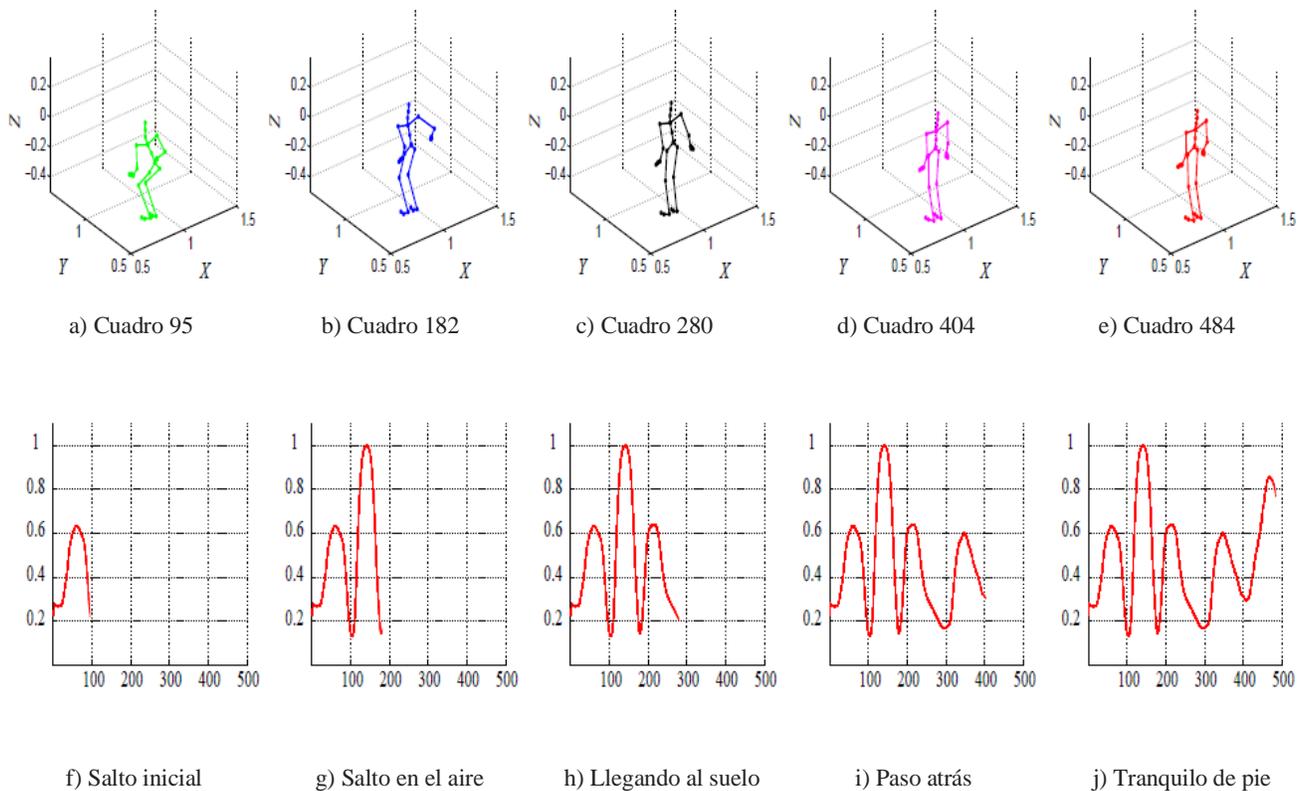


**Figura 14.** Cuadro 71 y representación de resultados finales

En general, los resultados obtenidos muestran claramente que existe una relación directa entre el vector de seguimiento propuesto y los datos dinámicos analizados.

### 5.3.3 Aplicación del método KSC en el caso: Salto vertical de humanos

La relación directa entre el vector de seguimiento  $\eta$  y la partición de los movimientos naturales del *Sujeto # 2* se puede apreciar claramente en la **Figura 15**, donde la fila superior muestra un Cuadro representativo por grupo (en diferente color) mientras que la fila inferior representa la curva  $\eta$  hasta el último cuadro por cluster correspondiente, como era de esperar, cada modo - cuadros entre las inflexiones que forman una curva cóncava - del trazado  $\eta$  corresponde a un grupo natural, que incluso puede determinarse mediante una simple inspección.



**Figura 15.** Seguimiento del *Sujeto #2*

## **5.4 Discusión**

Como se puede apreciar, la forma del vector de seguimiento tiene inflexiones en la transición desde un agrupamiento al siguiente. Entonces, el trazado de los vectores está directamente relacionada con los cambios a lo largo de la secuencia del cuadro.

Además, ya que no existen cambios en el vector de seguimiento al incorporar el ruido, se puede decir que es menos sensible a los datos de entrada ruidosos.

## CAPITULO VI

### CONCLUSIONES Y TRABAJO FUTURO

#### 6.1 Conclusiones

Al analizar una secuencia de cuadros representados por una matriz de un solo dato, con el objeto de identificar los eventos dinámicos subyacentes, los enfoques basados en kernels representan una alternativa apropiada. Sin duda las funciones de kernel provienen de una estimación del producto interno de espacios de representación dimensional alta, en donde se asume que los agrupamientos son separables y son a menudo definidos como una medida de similitud entre los puntos de datos del espacio original. Dicha similitud está asignada para un análisis de datos locales. En otras palabras, los kernels permiten una exploración de datos por partes por medio de una estimación de una varianza generalizada. Por lo tanto, se puede deducir que se puede hacer el seguimiento del comportamiento evolutivo de la secuencia mediante el uso de algunos valores de clasificación derivados a partir de una formulación con base en el kernel. De hecho, en este trabajo se demuestra que se puede lograr un enfoque de seguimiento viable mediante la maximización de un término de energía en relación con una versión aproximada del espacio de representación dimensional alto. Se utiliza un modelo ortonormado lineal, siendo el objetivo del problema de maximización, el cálculo de una proyección óptima de baja clasificación o matriz de rotación. Finalmente, al aprovechar la descomposición de un valor único de la matriz de kernel, se deduce un vector de seguimiento como una combinación lineal de los vectores propios cuadráticos. El seguimiento se lleva a cabo con el objeto de

encontrar un valor único que represente de forma adecuada a cada cuadro único. El enfoque (KSC-MS) propuesto en este documento determina un vector que tiene relación directa con el subsiguiente comportamiento dinámico de la secuencia analizada, permitiendo aún la estimación del número de grupos al igual que la verdad fundamental.

El punto de vista dinámico del amplio campo de análisis de datos implica un problema complejo y difícil de abordar, ya que los datos de entrada están ingresando a lo largo del tiempo. Más aún, las dinámicas, involucradas durante en si se suman más complejidad para la posterior tarea de procesamiento de datos. Referente a esto, uno de los temas abiertos desafiantes es la segmentación automática de movimiento, que se puede evaluar fácilmente sobre objetos giratorios. En este sentido, hemos demostrado que el método KSC ofrece una herramienta potente y adecuada.

En este trabajo, se prefiere el uso de enfoques no supervisados ya que en el mundo real aplicaciones de video una cantidad suficiente de etiquetado son inviables o prohibitivas. No obstante, la desventaja de trabajar en el análisis de objetos rotativos dentro de las configuraciones no supervisadas establece que no se puede realizar una segmentación automática del movimiento puede generarse directamente mediante una función de seguimiento (aquí, el llamado vector de seguimiento). En este punto, para superar este obstáculo, se ha introducido un procedimiento de codificación impulsado por indicadores de agrupación de modo que el vector de seguimiento multimodal original puede ser validado.

## **6.2 Trabajo futuro**

En estudios adicionales, el significado real de los valores de amplitud de los vectores de seguimiento y el comportamiento dinámico de los datos será revelado de manera formal. De igual forma, otras alternativas y propiedades de kernel serán exploradas con el fin de alcanzar un bajo costo de cálculo y enfoques eficientes para la segmentación del movimiento y seguimiento.

**BIBLIOGRAFIA**

Alvarez Nuñez, M. F. (2013). *Teoría de grafos*.

Alzate, C., & Suykens, J. A. K. (2010). Multiway Spectral Clustering with Out-of-Sample Extensions through Weighted Kernel PCA. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(2), 335–347. <https://doi.org/10.1109/TPAMI.2008.292>

Alzate, Carlos, & Suykens, J. A. (2006). A weighted kernel PCA formulation with out-of-sample extensions for spectral clustering methods. *Neural Networks, 2006. IJCNN'06. International Joint Conference on*, 138–144. IEEE.

Alzate, Carlos, & Suykens, J. A. (2010a). Highly sparse kernel spectral clustering with predictive out-of-sample extensions. *ESANN*, 235–240. Citeseer.

Alzate, Carlos, & Suykens, J. A. (2010b). Multiway spectral clustering with out-of-sample extensions through weighted kernel PCA. *IEEE transactions on pattern analysis and machine intelligence*, 32(2), 335–347.

Arróspide, J., & Salgado, L. (2015). Video based vehicle detection and tracking for driver assistance systems. *Securitas Vialis*, 7(1), 41–49. <https://doi.org/10.1007/s12615-014-9080-0>

Belanche Muñoz, L. A. (2013). Developments in kernel design. *ESANN 2013 proceedings: European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning: Bruges (Belgium), 24-26 April 2013*, 369–378.

- Bhavsar, M. H., & Ganatra, A. (2016). Radial Basis Polynomial Kernel (RBPK): A Generalized Kernel for Support Vector Machine. *International Journal of Computer Science and Information Security (IJCSIS)*, 14(4).
- Chan, A. B., & Vasconcelos, N. (2005). Probabilistic kernels for the classification of autoregressive visual processes. *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, 1, 846–851. IEEE.
- Chan, P. K., Schlag, M. D., & Zien, J. Y. (1994). Spectral k-way ratio-cut partitioning and clustering. *IEEE Transactions on computer-aided design of integrated circuits and systems*, 13(9), 1088–1096.
- Choi, S. W., & Lee, I.-B. (2004). Nonlinear dynamic process monitoring based on dynamic kernel PCA. *Chemical engineering science*, 59(24), 5897–5908.
- Chung, F. R. (1997). *Spectral graph theory*. American Mathematical Soc.
- D'Urso, P., De Giovanni, L., Disegna, M., & Massari, R. (2019). Fuzzy clustering with spatial-temporal information. *Spatial Statistics*, 30, 71–102.  
<https://doi.org/10.1016/j.spasta.2019.03.002>
- D'Urso, P., De Giovanni, L., & Massari, R. (2018). Robust fuzzy clustering of multivariate time trajectories. *International Journal of Approximate Reasoning*, 99, 12–38.  
<https://doi.org/10.1016/j.ijar.2018.05.002>
- González, F. A., Bermeo, D., Ramos, L., & Nasraoui, O. (2012). On the robustness of kernel-based clustering. *Iberoamerican Congress on Pattern Recognition*, 122–129. Springer.

- Guo, C., Zheng, S., Xie, Y., & Hao, W. (2012). A survey on spectral clustering. *World Automation Congress 2012*, 53–56. IEEE.
- Hagen, L., & Kahng, A. B. (1992). New spectral methods for ratio cut partitioning and clustering. *IEEE transactions on computer-aided design of integrated circuits and systems*, 11(9), 1074–1085.
- Ham, J., Lee, D. D., Mika, S., & Schölkopf, B. (2004). A kernel view of the dimensionality reduction of manifolds. *Proceedings of the twenty-first international conference on Machine learning*, 47. ACM.
- Hendrickson, B., & Leland, R. (1995). An improved spectral graph partitioning algorithm for mapping parallel computations. *SIAM Journal on Scientific Computing*, 16(2), 452–469.
- Higham, D. J., & Kibble, M. (2004). A unified view of spectral clustering. *University of Strathclyde mathematics research report*, 2.
- Huang, H.-C., Chuang, Y.-Y., & Chen, C.-S. (2012). Multiple kernel fuzzy clustering. *IEEE Transactions on Fuzzy Systems*, 20(1), 120–134.
- Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of classification*, 2(1), 193–218.
- Ishikawa, H., & Geiger, D. (1998). Segmentation by grouping junctions. *Computer Vision and Pattern Recognition, 1998. Proceedings. 1998 IEEE Computer Society Conference on*, 125–131. IEEE.
- Joseph, D. B., Bart, D. M., & K, S. J. A. (2002). *Least Squares Support Vector Machines*. World Scientific.

- Jun Shi, Z. J. (14 de septiembre de 2013). Adaptive Graph Embedding Discriminant Projections. Obtenido de <https://link.springer.com/article/10.1007/s11063-013-9323-8>
- Keshet, J., & Bengio, S. (2009). *Automatic speech and speaker recognition: Large margin and kernel methods*. John Wiley & Sons.
- Langone, R., Alzate, C., & Suykens, J. A. (2013). Kernel spectral clustering with memory effect. *Physica A: Statistical Mechanics and its Applications*, 392(10), 2588–2606.
- López Díaz, A. (2018). *Fundamentos matemáticos de los métodos Kernel para aprendizaje supervisado*.
- Lu, L., Zhang, X., Xu, X., & Shang, D. (2015). Video analysis using spatiotemporal descriptor and kernel extreme learning machine for lip reading. *Journal of Electronic Imaging*, 24, 053023. <https://doi.org/10.1117/1.JEI.24.5.053023>
- Maestri, M. L., Cassanello, M. C., & Horowitz, G. I. (2009). Kernel pca performance in processes with multiple operation modes. *Chemical Product and Process Modeling*, 4(5).
- Meila, M., & Shi, J. (2001a). *A random walks view of spectral segmentation*.
- Meila, M., & Shi, J. (2001b). Learning segmentation by random walks. *Advances in neural information processing systems*, 873–879.
- Molina-Giraldo, S., Álvarez-Meza, A. M., Peluffo-Ordoñez, D. H., & Castellanos-Domínguez, G. (2012). Image Segmentation Based on Multi-Kernel Learning and Feature Relevance Analysis. En J. Pavón, N. D. Duque-Méndez, & R. Fuentes-Fernández (Eds.), *Advances in Artificial Intelligence – IBERAMIA 2012* (pp. 501–510). Springer Berlin Heidelberg.
- Nene, S. A., Nayar, S. K., & Murase, H. (1996). *Columbia object image library (coil-20)*.

- Ng, A. Y., Jordan, M. I., & Weiss, Y. (2002a). On spectral clustering: Analysis and an algorithm. *Advances in neural information processing systems*, 849–856.
- Ng, A. Y., Jordan, M. I., & Weiss, Y. (2002b). On spectral clustering: Analysis and an algorithm. *Advances in neural information processing systems*, 849–856.
- Ng, A. Y., Jordan, M. I., & Weiss, Y. (2002c). On spectral clustering: Analysis and an algorithm. *Advances in neural information processing systems*, 849–856.
- Oña-Rocha, O., Sánchez-Manosalvas, O., Umaquina-Criollo, A. C., Rosero-Montalvo, P., Suárez-Zambrano, L., Rodríguez-Sotelo, J. L., & Peluffo-Ordóñez, D. H. (2017). Automatic Motion Segmentation via a Cumulative Kernel Representation and Spectral Clustering. *International Conference on Intelligent Data Engineering and Automated Learning*, 406–414. Springer.
- Peluffo Ordoñez, D. H., Lee, J. A., Verleysen, M., Rodriguez, J. L., & Castellanos-Dominguez, G. (2015). *Unsupervised relevance analysis for feature extraction and selection. A distance-based approach for feature relevance*. Presentado en 3rd International Conference on Pattern Recognition Applications and Methods (ICPRAM 2014). Recuperado de <https://dial.uclouvain.be/pr/boreal/object/boreal:171343>
- Peluffo-Ordóñez, D., García-Vega, S., Langone, R., Suykens, J. A., & Castellanos-Domínguez, G. (2013). Kernel spectral clustering for dynamic data using multiple kernel learning. *The 2013 International Joint Conference on Neural Networks (IJCNN)*, 1–6. IEEE.

- Peluffo-Ordóñez, D. H., García-Vega, S., Alvarez-Meza, A. M., & Castellanos-Domínguez, C. G. (2013). Kernel spectral clustering for dynamic data. *Iberoamerican Congress on Pattern Recognition*, 238–245. Springer.
- Perry, W. L., Sánchez Trujillo, C. M., & others. (1990). *Álgebra Lineal con aplicaciones*.
- Poole, D. (2011). *Álgebra lineal: una introducción moderna*. Cengage Learning Editores.
- Ramírez-Giraldo, D., Molina-Giraldo, S., Álvarez-Meza, A. M., Daza-Santacoloma, G., & Castellanos-Domínguez, G. (2012). Kernel based hand gesture recognition using kinect sensor. *2012 XVII Symposium of image, signal processing, and artificial vision (STSIVA)*, 158–161. IEEE.
- Rodríguez-Sotelo, J. L., Peluffo-Ordóñez, D., Cuesta-Frau, D., & Castellanos-Domínguez, G. (2009). Nonparametric density-based clustering for cardiac arrhythmia analysis. *Computers in Cardiology, 2009*, 569–572. IEEE.
- S. A. Nene, S. K. (February de 1996). Columbia University Image Library. Obtenido de <http://www.cs.columbia.edu/CAVE/software/softlib/coil-20.php>
- Sánchez, L. G., Osorio, G. A., & Suárez, J. F. (2008). Introduction to Kernel PCA and other Spectral Methods Applied to Unsupervised Learning. *Revista Colombiana de Estadística*, 31(1), 19–40.
- Saripalle, S. K., Paiva, G. C., Cliett III, T. C., Derakhshani, R. R., King, G. W., & Lovelace, C. T. (2014). Classification of body movements based on posturographic data. *Human movement science*, 33, 238–250.

- Schölkopf, B., Smola, A. J., Bach, F., & others. (2002). *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press.
- Shanmao, G., Yunlong, L., Lijun, L., & Ni, Z. (2015). Weighted principal component analysis applied to continuous stirred tank reactor system with time-varying. *2015 34th Chinese Control Conference (CCC)*, 6377–6381. <https://doi.org/10.1109/ChiCC.2015.7260643>
- Shi, Jianbo, & Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence*, 22(8), 888–905.
- Shi, Jun, Jiang, Z., & Feng, H. (2014). Adaptive Graph Embedding Discriminant Projections. *Neural Processing Letters*, 40(3), 211–226. <https://doi.org/10.1007/s11063-013-9323-8>
- Shirazi, S., Harandi, M. T., Sanderson, C., Alavi, A., & Lovell, B. C. (2012). Clustering on Grassmann manifolds via kernel embedding with application to action analysis. *Image Processing (ICIP), 2012 19th IEEE International Conference on*, 781–784. IEEE.
- Stella, X. Y., & Shi, J. (2003). Multiclass spectral clustering. *null*, 313. IEEE.
- Strehl, A., & Ghosh, J. (2002). Cluster ensembles—a knowledge reuse framework for combining multiple partitions. *Journal of machine learning research*, 3(Dec), 583–617.
- Sudha, L. R., & Bhavani, R. (2012). Performance comparison of SVM and kNN in automatic classification of human gait patterns. *Int. J. Comput*, 6(1), 19–28.
- Takács, B., Butler, S., & Demiris, Y. (2007). Multi-agent behaviour segmentation via spectral clustering. *Proceedings of the AAAI-2007, PAIR Workshop*, 74–81.

- Valencia, X. P. B. (2017). *Sistema genérico de razonamiento basado en casos (cbr) multi-clase como soporte al diagnóstico médico mediante técnicas de reconocimiento de patrones* (PhD Thesis). Universidad de Salamanca.
- Wolf, L., & Bileschi, S. (2005). Combining variable selection with dimensionality reduction. *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, 2, 801–806. IEEE.
- Wolf, L., & Shashua, A. (2005). Feature Selection for Unsupervised and Supervised Inference: The Emergence of Sparsity in a Weight-Based Approach. *Journal of Machine Learning Research*, 6(Nov), 1855–1887. Recuperado de <http://www.jmlr.org/papers/v6/wolf05a.html>
- Yadav, S., Dubey, R., & Ahmed, M. (2014). An Advanced Motion Detection Algorithm with Video Quality Analysis for Video Surveillance Systems. *International Journal of Advanced Research in Computer Science*, 5(8).
- Zelnik-Manor, L., & Perona, P. (2005). Self-tuning spectral clustering. *Advances in neural information processing systems*, 1601–1608.
- Zha, H., He, X., Ding, C., Gu, M., & Simon, H. D. (2002). Spectral relaxation for k-means clustering. *Advances in neural information processing systems*, 1057–1064.
- Zhou, F., De la Torre, F., & Hodgins, J. K. (2008). Aligned cluster analysis for temporal segmentation of human motion. *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*, 1–7. IEEE.