



FACULTAD *de* CIENCIAS  
*de la* COMPUTACIÓN

SIMPLIFICACIÓN DE MODELOS BASADOS EN APARIENCIA  
PARA EL RECONOCIMIENTO DE OBJETOS  
TRIDIMENSIONALES

OSCAR ANTONIO DE LEÓN GÓMEZ

Tesis para alcanzar el Título de Licenciado en Ciencias de la Computación  
Facultad de Ciencias de la Computación  
Benemérita Universidad Autónoma de Puebla

Oscar Antonio De León Gómez: *Simplificación de modelos basados en apariencia para el reconocimiento de objetos tridimensionales*, , ©

ASESOR:

Dr. Luis Carlos Altamirano Robles

LUGAR:

H. Puebla de Z. Mayo 2012

A mis padres: Soledad y Antonio.



## RESUMEN

---

En este trabajo se propone un método para reducir el número de imágenes necesarias para representar un objeto bajo el paradigma basado en apariencia. Se presenta un algoritmo que toma ventaja de las características geométricas de los manifolds que representan a dichos objetos para construir un conjunto de modelos locales embebidos en subespacios lineales. Cada uno de estos modelos contiene un conjunto de imágenes cuya similaridad se mide mediante un criterio de linealidad que nos indica la pertenencia a un subespacio. Estas imágenes se promedian para construir un *ejemplar* del modelo que se puede utilizar para sustituir a todo el conjunto y se utiliza para efectuar tareas de reconocimiento de objetos y son requeridos para el cálculo de similaridad entre manifold. Dicho cálculo se usará posteriormente en la clasificación de objetos.



*Toda la credibilidad,  
toda buena conciencia,  
toda evidencia de verdad  
Viene sólo de los sentidos.*

Friedrich Nietzsche.

## AGRADECIMIENTOS

---

**A mis padres** por su infinito amor, por la confianza que siempre depositaron en mí y por el regalo más grande que me dieron: la vida.

**A mis hermanos** que siempre han estado conmigo a pesar de todo.

**A mis profesores** que exigieron de mí tanto como podía dar y aún un poco más.

**Al Dr. Luis Carlos Altamirano Robles**, su paciencia inestimable, su guía y amistad han sido importantes en muchos aspectos de mi vida profesional y personal.

**A Esther** por su amor, comprensión y su ayuda en la revisión del manuscrito ;).

Para todos aquellos a quienes me está faltando mencionar, les agradezco tanto permitirme andar el mismo camino aunque haya sido sólo un instante.



## ÍNDICE GENERAL

---

1	INTRODUCCIÓN	1
1.1	Apariencia de un objeto	2
1.2	Modelos basados en apariencia	2
1.3	Sistema de adquisición de modelos	4
2	PRELIMINARES MATEMÁTICOS	7
2.1	Apariencia de un objeto	7
2.2	Creación de modelos utilizando la transformada de Hotelling	7
2.3	Medidas de similaridad	9
2.3.1	Distancia euclidea	10
2.3.2	Distancia geodesica	11
2.4	Grados de libertad	11
2.5	Parche lineal maximal	12
2.5.1	Medida de distancia entre ejemplares	12
3	PROBLEMÁTICA A RESOLVER	13
4	PROPUESTA DE SOLUCIÓN	17
4.1	Cálculo de los parches lineales maximales	17
5	EXPERIMENTOS	19
5.1	Verificación de los modelos simplificados	23
6	CONCLUSIONES Y TRABAJO FUTURO	25

## ÍNDICE DE FIGURAS

---

Figura 1	Izquierda: imagen de un objeto; derecha: gráfica de curvas de nivel <i>manifolds</i> para el mismo objeto utilizando los valores de los tres primeros eigenvectores y dos grados de libertad: $\theta_1$ y $\theta_2$ [Nay96]. 2
Figura 2	Modelo de apariencia de un objeto 3
Figura 3	Modelo de apariencia con dos grados de libertad: ángulo de rotación y elevación de la cámara 4
Figura 4	Componentes principales de un sistema simple de captura de modelos 5
Figura 5	Izquierda. Imagen <i>cruda</i> Derecha. Imagen normalizada en escala y brillantez 5
Figura 6	Eigenimágenes obtenidas de COIL-20 y ejemplo de una imagen reconstruida 8
Figura 7	ejemplos de manifolds extraídos de COIL-20 9
Figura 8	Ejemplos de bases de modelos 13
Figura 9	Secuencias de imágenes utilizadas en aplicaciones de vigilancia y seguimiento 14
Figura 10	Ejemplos de modelos de apariencia de gestos de la mano 14
Figura 11	Ejemplos de las imágenes utilizadas en los experimentos 19
Figura 12	Algunos resultados para modelos sencillos. El punto marcado con un asterisco es el ejemplar correspondiente a cada MLP. 21
Figura 13	Algunos resultados para modelos con modificaciones 23

## ÍNDICE DE CUADROS

---

Cuadro 1	Resultados de la aplicación del algoritmo 1 a la biblioteca COIL-20 20
Cuadro 2	Resultados de la aplicación del algoritmo 1 a la biblioteca COIL-20. En este caso se agregaron 4 modelos con modificaciones. 22

## INTRODUCCIÓN

---

La visión es un fenómeno sensorial que se produce en el ojo como respuesta a los estímulos luminosos del ambiente. La luz incide en la retina donde células especializadas, llamadas bastones y conos, la transforman en impulsos eléctricos que son transmitidos al cerebro por medio del nervio óptico. Esta señal es procesada luego por la corteza visual. Ya en los últimos años de 1700, Emmanuel Kant estableció que la percepción en general es una representación de la realidad objetiva externa[[kano6](#)], donde la vista tiene un papel preponderante. Así, lo que experimentamos como *ver* es sólo una parte de la percepción humana, una parte de un modelo del mundo real captado por el sistema *cerebro-ojo*, esto último ha sido confirmado empíricamente por la psicología cognitiva y las neurociencias. En nuestra vida cotidiana es fácil pasar por alto que estamos continuamente bombardeados por cantidades enormes de estímulos visuales. Para hacer frente a esta avalancha de información, el cerebro las organiza en experiencias visuales que son almacenadas en la memoria y se encuentran asociadas con otras funciones como la emoción y el lenguaje. De esta manera, el cerebro puede encargarse de la tarea de clasificar e identificar los objetos que percibe visualmente.

En los inicios del reconocimiento de objetos mediante imágenes digitales, se utilizó ampliamente el paradigma geométrico, dada la amplia disponibilidad de algoritmos para la detección de bordes y esquinas. A pesar de arrojar buenos resultados, se pone de manifiesto que utilizar únicamente la forma para reconocer objetos tiene sus limitantes, por ejemplo, si tenemos dos cubos iguales en forma y tamaño pero uno es rojo y el otro azul. Obviamente cuando se le presenten estas instancias a un sistema de reconocimiento que tome sólo en cuenta las características geométricas, éste nos dirá que ambos son el mismo objeto pues tienen la misma forma.

Esto sucede porque se pierde mucha información inherente al objeto. En general ambos métodos tienen sus ventajas y sus desventajas pero parecen complementarse. Así, tener un sistema clasificador de objetos más natural requiere de:

- Maneras adecuadas de diferenciar entre objetos modelados a través de conjuntos de imágenes digitales (medidas de similitud/disimilitud).
- Algoritmos de agrupamiento capaces de poner modelos de objetos en clases en lugar de clasificar sólo puntos como los algoritmos usuales.

## 1.1 APARIENCIA DE UN OBJETO

La apariencia de un objeto está dada por la combinación de su forma, sus propiedades de reflectancia, su posición ante la cámara en la escena y la iluminación presente en ésta [Nay96]. Es importante notar que mientras las dos primeras son inherentes al objeto, la posición del elemento frente al sensor y la iluminación pueden variar de una escena a otra [Nay96]. Así mismo, es claro que esta definición implica que la apariencia de un objeto en un momento determinado, (o al menos, de parte de ese objeto), puede deducirse a partir de una imagen digital en la que se capture ese elemento en una escena específica, (con una luz determinada y un cierto ángulo de la cámara). De igual manera, podrá emplearse entonces un conjunto de imágenes para generar un modelo del objeto en diversas condiciones.

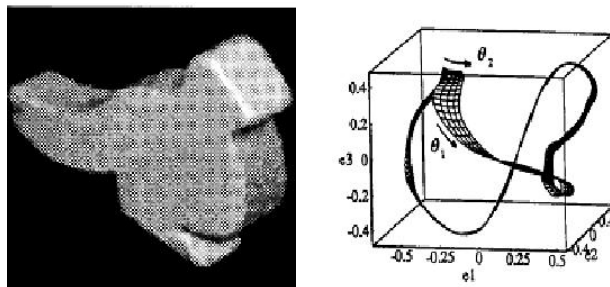


Figura 1: Izquierda: imagen de un objeto; derecha: gráfica de curvas de nivel *manifolds* para el mismo objeto utilizando los valores de los tres primeros eigenvectores y dos grados de libertad:  $\theta_1$  y  $\theta_2$  [Nay96].

En la figura se puede ver la representación gráfica de la apariencia del objeto en el espacio de apariencia para la vista que contiene esta imagen digital.

## 1.2 MODELOS BASADOS EN APARIENCIA

Uno de los problemas centrales dentro de la visión computarizada, es la construcción de modelos que puedan ser usados por la computadora, para poder llevar a cabo la comparación de éstos con patrones que son extraídos de una imagen, y así poder realizar la tarea de reconocimiento. Tradicionalmente, estos modelos se han construido a través de las propiedades geométricas, dando lugar a los llamados modelos geométricos; así mismo, se ha hecho uso de descriptores invariantes para representar a los objetos. Este tipo de representaciones han sido ampliamente investigadas, sin embargo, el reconocimiento de objetos a través de estos enfoques presenta algunos problemas que hasta la fecha son motivo de investigación: Uno de ellos consiste en la pérdida de información que se lleva a cabo al extraer la información geométrica. Un ejemplo es la pérdida de la información contenida dentro del contorno del objeto, la cual es importante en muchas tareas de

reconocimiento.

Debido a estos inconvenientes, se propuso a mediados de los años 90 del siglo pasado, una representación de objetos basándose en su apariencia, que preserve además de las características geométricas, la textura y el color del objeto.



Figura 2: Modelo de apariencia de un objeto

Típicamente, este tipo de enfoque trabaja de la siguiente manera: se toma una cierta cantidad de imágenes del objeto del que se quiere construir el modelo, una para cada una de las posibles formas en que se piensa que el objeto podrá aparecer en una escena, y combinado con las diversas condiciones de iluminación que incidan sobre él. Enseguida se almacenan estas imágenes en algún formato elegido (gif, tiff, bmp, etc.), para su posterior uso; estas imágenes por sí mismas pueden formar el modelo del objeto, o bien, se les pueden extraer algunas características que se piensa describen al objeto de manera precisa, y entonces éstas constituirán al modelo. Finalmente, el proceso de reconocimiento propiamente dicho, consiste en comparar el modelo con una nueva imagen que probablemente contenga al objeto. Por supuesto, este último proceso requiere algún procesamiento extra sobre la nueva imagen, tal como la segmentación del objeto de interés en la imagen, interpolación entre imágenes, etc.



Figura 3: Modelo de apariencia con dos grados de libertad: ángulo de rotación y elevación de la cámara

Aunque en este enfoque se tienen algunos problemas con estos modelos, se ha demostrado en algunos estudios tanto en humanos como en animales, que los sistemas biológicos trabajan con representaciones basadas en vistas. De aquí que es razonable reproducir esta forma de visión en sistemas automáticos.

### 1.3 SISTEMA DE ADQUISICIÓN DE MODELOS

Para la adquisición de modelos de objetos se requiere obtener un conjunto de imágenes digitales(vistas) de cada objeto en particular para luego obtener de ellas el modelo correspondiente a ese conjunto de objetos o *modelbase*.

El sistema consta básicamente de[Reyo6]:

- Un tornamesa controlado por un motor a pasos.
- Una fuente de iluminación constante.
- Una cámara digital.

Además, este sistema debe situarse en un entorno que asegura completa oscuridad, con el fin de mantener constantes las características de iluminación, ya que los modelos de apariencia son muy sensibles a los cambios de éstas.

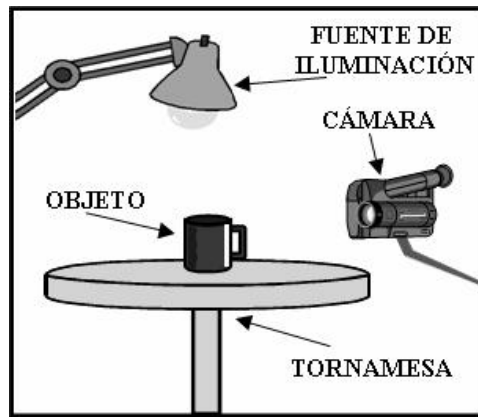


Figura 4: Componentes principales de un sistema simple de captura de modelos

Es necesario introducir algunas condiciones de ideales para la captura de estos modelos, a saber:

1. Cada imagen contendrá sólo un objeto
2. Todas las imágenes tendrán la misma resolución y profundidad en bits.
3. El objeto a modelar se presentará libre de oclusiones.
4. las imágenes están normalizadas en escala y brillantez[Roboza]

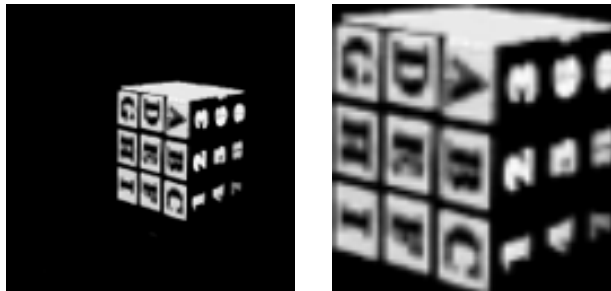


Figura 5: Izquierda. Imagen *cruda* Derecha. Imagen normalizada en escala y brillantez

Se ha determinado empíricamente que para tener un modelo *completo* sólo se requieren de 36 a 72 imágenes del objeto. A partir de este modelo se puede obtener un modelo no uniforme[Roboza] que mantiene la información necesaria del objeto al mismo tiempo que se reduce el número de imágenes necesarias para la generación del modelo sin afectar de manera significativa el porcentaje de reconocimiento del sistema.



## PRELIMINARES MATEMÁTICOS

---

### 2.1 APARIENCIA DE UN OBJETO

Formalmente podemos definir la apariencia de un objeto como una función.

$$\sigma : [0, 2\pi] \rightarrow \mathbb{R}^{n,m}$$

que está parametrizada por el ángulo de rotación del objeto frente a la cámara. Esta función es llamada también la trayectoria definida por la apariencia del objeto en el espacio de matrices  $\mathbb{R}_{n,m}$ . Cuando la trayectoria es parametrizada por más de una variable, el azimut de la cámara y el ángulo de rotación. La apariencia de un objeto se define formalmente como una superficie parametrizada por  $D$ , donde  $D$  es algún dominio en  $\mathbb{R}^n$ . Esta trayectoria o superficie es también es llamada *manifold*. Así, la apariencia de un objeto puede ser tratada de manera abstracta como un objeto matemático, en este trabajo se considera al *manifold* como el modelo de apariencia de un objeto.

### 2.2 CREACIÓN DE MODELOS UTILIZANDO LA TRANSFORMADA DE HOTELLING

Dada la cantidad de imágenes que se utilizan para modelar un único objeto y a que en la práctica es común que se tengan grandes cantidades de objetos diferentes, se pone de manifiesto la necesidad de encontrar una representación comprimida de todas las imágenes necesarias. En este trabajo se utiliza una técnica frecuentemente utilizada en el reconocimiento de patrones: la transformada de Hotelling o, como es más conocida, Análisis de Componentes Principales [Nay96, Cluo9].

El procedimiento a seguir es el siguiente: [Nay96, Coo04]

1. Cada imagen se empila por columnas, dando como resultado un vector columna  $i_j$  de  $m * n$  filas. Donde  $m$  y  $n$  son las dimensiones de la imagen original.
2. Se normaliza cada una de las imágenes obtenidas en el paso anterior dividiendo entre la norma correspondiente:

$$\hat{i}_j = \frac{i_j}{\|i_j\|}$$

3. Se obtiene el vector promedio de todas las imágenes

$$i_m = \frac{\sum_{j=1}^r \hat{i}_j}{r}$$

4. Se resta el promedio a cada imagen
5. Se obtiene la matriz  $P$ , donde cada columna es uno de los vectores antes calculados.  $P$  es de dimensiones  $(m \times n)$  filas por  $r$  columnas
6. Se obtiene la matriz de covarianza implícita, que es de dimensiones  $r \times r$ , mucho menor a la matriz de covarianza:

$$\bar{Q} = P^T * P$$

7. Se obtienen los eigenvalores y eigenvectores de  $\bar{Q}$ , de los cuales se obtienen los correspondientes a la matriz de covarianza, mediante:

$$\lambda_i = \bar{\lambda}_i \quad (1)$$

$$e_i = \bar{\lambda}_i^{-1/2} P \bar{e}_i \quad (2)$$

Después de este proceso se obtiene un conjunto de vectores propios o *eigenimágenes*, estos vectores y valores propios forman un *eigenespacio* de pocas dimensiones. Las dimensión del eigenespacio corresponden al número de componentes principales que se desee utilizar. En la práctica son suficientes 20 como máximo. Se debe tomar en cuenta que este procedimiento debe incluir a todas las imágenes de todos los objetos. Al resolver el sistema de ecuaciones:

$$eA = P \quad (3)$$

Se obtiene la matriz  $A$  que contiene los coeficientes necesarios para reconstruir cualquier imagen de cualquier modelo. Este esquema de compresión muestra pérdidas de información insignificantes para el problema que nos atañe. En la figura 7 se muestra el resultado de la extracción de manifolds para dos objetos y en la figura 6 se pueden observar algunas de las eigenimágenes obtenidas mediante este método.

### 2.3 MEDIDAS DE SIMILARIDAD

Como ya se dijo en la sección anterior, la matriz  $A$  contiene los *manifolds* de los objetos en la modelbase. Estos definen una trayectoria parametrizada en  $\mathbb{R}^k$  donde  $k$  es el número de componentes principales seleccionados. El problema de agrupar modelos de objetos basados en apariencia debe tratar con esta representación compacta. Es indispensable encontrar la medida o conjunto de medidas de similitud que sean necesarias para realizar una clasificación de objetos parecida a la que un ser humano haría. En la literatura lo común es encontrar medidas de similitud entre objetos representados como un vector (columna o fila) con  $d$  características, almacenados en bases de datos de  $N$  objetos (dataset) [Cluog]. La fuerza de la relación entre dos puntos de un *dataset* es indicada por un *coeficiente de similitud*,

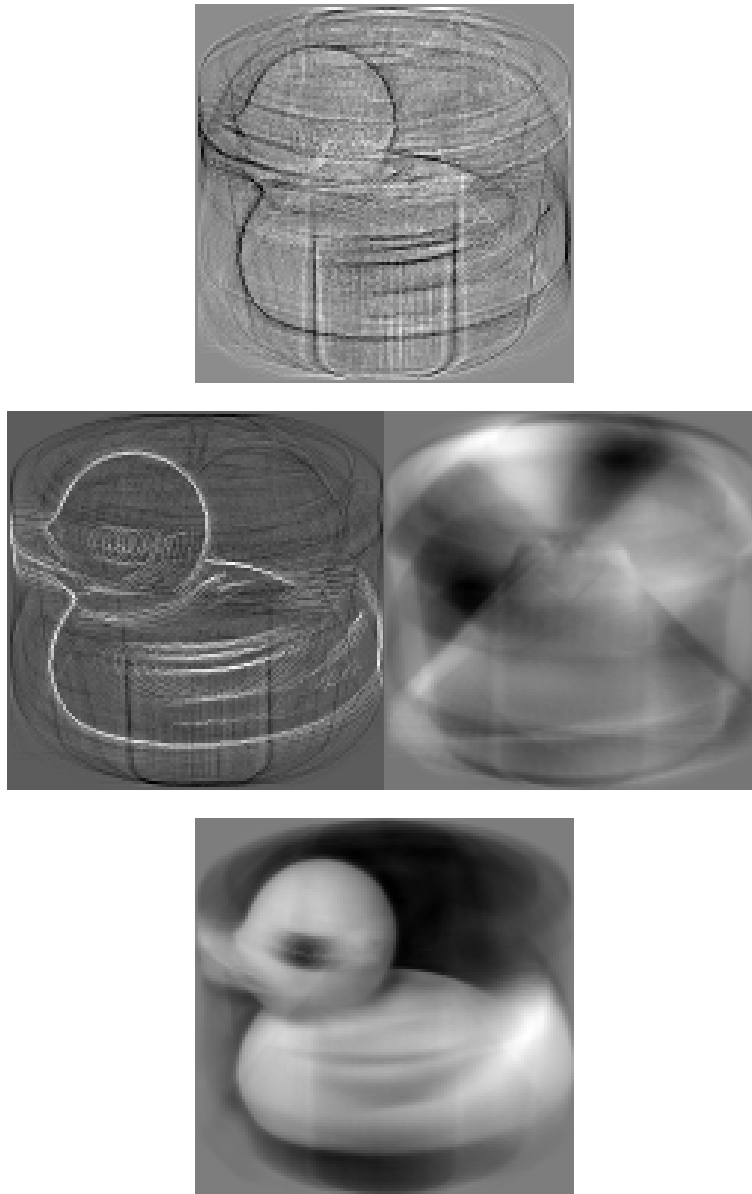


Figura 6: Eigenimágenes obtenidas de COIL-20 y ejemplo de una imagen reconstruida

entre más grande es este coeficiente, más relacionados están dichos puntos. Para que una función definida en los atributos de un objeto sea considerada una métrica debe cumplir las siguientes características:

1. Simetría
2. Reflexividad
3. Positividad
4. Cumplir con la desigualdad triangular.

Sólo cuando una función de similaridad no cumple con la desigualdad del triángulo, es llamada una semimétrica[Clu09]. Existen muchas otras estructuras de similaridad sobre un dataset  $D$ [GG07] como puede ser la definida por un árbol o una partición de  $D$ . Estas medi-

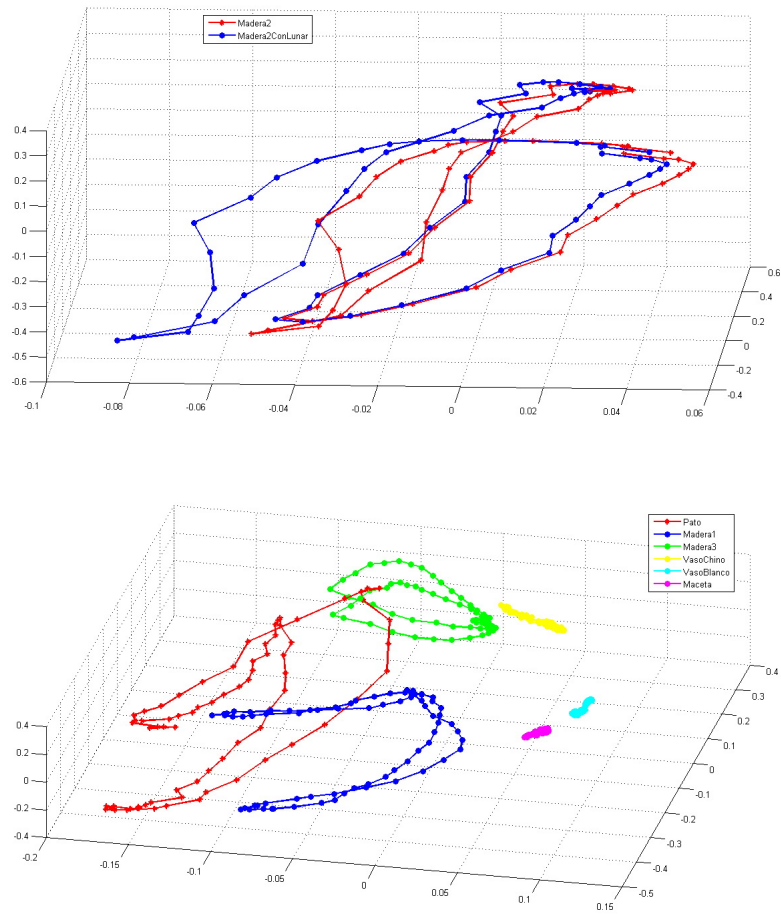


Figura 7: ejemplos de manifolds extraídos de COIL-20

das de similitud son utilizadas extensivamente como parte de las técnicas corrientes de agrupamiento. Por mencionar las más importantes tenemos:

- La distancia euclídeana
- La distancia Manhattan
- La distancia de Mahalanobis
- La distancia geodésica

### 2.3.1 Distancia euclídeana

La distancia euclídeana es probablemente la medida de distancia más utilizada con datos numéricos. Si tenemos  $x$  y  $y$ , dos vectores en un espacio de dimensión  $d$ , la distancia euclídeana entre  $x$  y  $y$  se define como:

$$D(x, y) = \sqrt{\sum_{j=1}^d (x_j - y_j)^2}$$

Su uso en algoritmos de agrupamiento tiende a formar clusters hiperesféricos invariantes a translaciones y rotaciones en el espacio de características[GG07]. En este trabajo se utilizó la distancia euclídeana porque es fácil de calcular y se ajusta a los requerimientos del problema.

### 2.3.2 Distancia geodesica

Esta distancia está definida en teoría de grafos como el número de aristas en el camino mínimo entre dos vértices, donde si éstos no están conectados la distancia geodésica es convencionalmente definida como *infinita*[JB03]. Para calcular la distancia geodésica primero debe construirse un grafo simétrico de vecindad conectando el punto(nodo) $x_i$  con el punto  $x_j$  si éste es uno de los  $k$ -vecinos más cercanos del primero. Las aristas de este grafo están pesadas con la distancia euclídeana entre los puntos vecinos. Luego, las distancias geodesicas son calculadas como la longitud del camino mínimo sobre las aristas utilizando alguno de los algoritmos disponibles como el algoritmo de Dijkstra o el de Floyd. Formalmente[Clu09], sea  $D(x_i, x_j)$  la distancia euclídeana entre  $x_i$  y  $x_j$ . La distancia geodésica  $D_G(x_i, x_j)$  se inicializa:

$$D_G(x_i, x_j) = \begin{cases} D(x_i, x_j) & \text{si } x_i \text{ está conectado a } x_j \\ \infty & \text{en cualquier otro caso} \end{cases}$$

Luego, para cada punto  $x_k$ , donde  $k = 1, 2, \dots, N$  y  $N$  es el número de puntos en el dataset, se calcula :

$$D_G(x_i, x_j) = \min(D_G(x_i, x_j), D_G(x_i, x_k) + D_G(x_k, x_j))$$

Lo que da como resultado la matriz de distancias geodésicas.

## 2.4 GRADOS DE LIBERTAD

Cada modelo de apariencia está parametrizado por un conjunto de variables que dependen de la tarea de visión particular que se esté realizando. Para el caso del reconocimiento y modelado de objetos, estos parámetros pueden ser: ángulo de giro de la tornamesa en el sistema de adquisición, la posición de la cámara respecto al eje vertical o azimut o la intensidad de la iluminación. Dichas variables son denominadas los *Grados de Libertad* o DOFs (*Degrees of Freedom*) de la tarea de visión [Nay96] y se denotan de la siguiente manera:

$$q = [q_1, q_2, \dots, q_s]$$

Donde  $s$  corresponde al número de grados de libertad. Para cualquier vector  $q$ , el sensor capturaré entonces un conjunto de  $r$  imágenes:

$$i = [i_1, i_2, \dots, i_r]$$

Así, para el conjunto continuo de valores del conjunto  $q$ , se obtiene un dominio de imágenes que corresponden a un rango de apariencias del objeto en cuestión. Dicho rango es conocido como *el área de trabajo visual* [Nay96]

## 2.5 PARCHE LINEAL MAXIMAL

El concepto de parche lineal maximal (Maximal Linear Patch) o MLP está definido como un conjunto de puntos que definen un subespacio lineal maximal donde la perturbación lineal de estos puntos está dada naturalmente como la desviación entre la distancia euclidiana y la distancia geodésica. La idea es partir el conjunto de datos en varios MLPs que contengan conjuntos de imágenes muy parecidas entre sí. Estos parches se calculan partiendo de un punto aleatorio del dataset hasta que la condición de linealidad se pierda. Este procedimiento garantiza la linealidad local de los modelos y controla adaptativamente el número de modelos calculados. Formalmente [RW34], sea  $X = x_1, x_2, \dots, x_N$  un conjunto de datos, en nuestro caso son puntos que conforman manifolds, donde  $x_i \in \mathbb{R}^D$  y  $N$  es el número de puntos en  $X$ . Se busca partir  $X$  en una colección disjunta de MLPs  $C_i$  tales que:

$$\begin{aligned} X &= \bigcup_{i=1}^m C_i, \\ C_i \cup C_j &= \emptyset (i \neq j, i, j = 1, 2, \dots, m) \\ C_i &= x_1^{(i)}, x_2^{(i)}, \dots, x_{N_i}^{(i)} \left( \sum_{i=1}^m N_i = N \right) \end{aligned}$$

donde  $m$  es el número de parches y  $N_i$  es el número de puntos en el parche. Para realizar el particionamiento es necesario calcular la matriz de distancias euclidianas  $D_E$  y la matriz de distancias geodésicas  $D_G$ . El procedimiento para calcular ésta última es descrito en 2.3.2 y el algoritmo utilizado en el cálculo se muestra en 1.

### 2.5.1 Medida de distancia entre ejemplares

Una vez obtenido el conjunto de los subespacios lineales de los objetos de la biblioteca, es necesario discriminar entre los modelos y un conjunto de imágenes de prueba para lograr el reconocimiento de objetos. Para este objetivo se utiliza a medida de similaridad entre ejemplares, que es definida como la correlación de dos ejemplares  $e_i$  y  $e_j$  mediante:

$$d_E(e_i, e_j) = \|e_i\| \|e_j\| / e_i^T e_j \quad (4)$$

Cuando se comparan dos conjuntos de imágenes, la distancia entre ejemplares nos muestra la similaridad entre la apariencia de los objetos.

## PROBLEMÁTICA A RESOLVER

---

Las técnicas basadas en apariencia implican un muestreo uniforme y discreto de imágenes que no toman en cuenta la complejidad del objeto así que el número de imágenes necesarias para la representación de un objeto se supone como una constante independiente [Robo2a, Robo2b] en el caso de escenas con parámetros controlados. Por otro lado en la práctica se tiene que las condiciones en las que se capturan los modelos no son ideales y cambian constantemente. Esto provoca que las tareas de identificación en situaciones reales se enfrenten a tres problemas principales:

- El espacio de búsqueda de imágenes para la estimación de pose es grande.
- El espacio de almacenamiento en disco es prohibitivo.
- Los manifolds resultantes de la proyección a un espacio de menor dimensionalidad tienden a traslaparse.

En condiciones ideales, un modelo de apariencia *completo* requiere, según [ea99], de seis grados de libertad para considerar las variantes de la posición del objeto y dos para considerar la dirección de la iluminación. Este modelo utiliza imágenes de  $128 * 128$  píxeles que da como resultado vectores de dimensión 16384. Además del problema de almacenamiento de los modelos, es necesario tratar el problema del reconocimiento de objetos en una imagen recién presentada al sistema. La primera aproximación al reconocimiento es el comparar cada imagen de los modelos almacenados con la nueva instancia calculando la diferencia entre las imágenes, lo que es obviamente ineficiente. También se puede utilizar otras medidas como la suma de diferencias cuadradas [Ber09] lo que no resuelve los problemas antes mencionados.

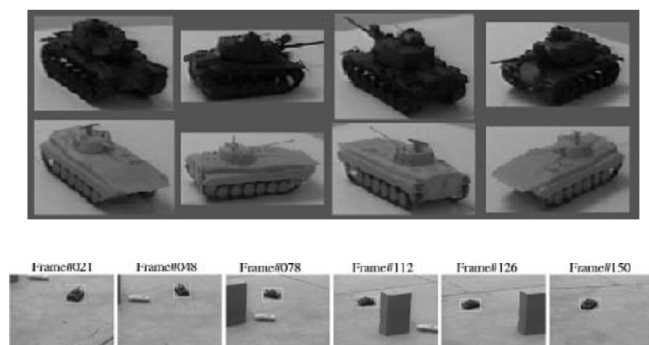


Figura 8: Ejemplos de bases de modelos

Uno de los problemas prácticos encontrados en los cuales se utiliza el paradigma de apariencia es, por ejemplo, el uso de estas técnicas en grandes bases de datos de imágenes infrarrojas para aplicaciones militares de identificación de vehículos en misiones de reconocimiento. Estas bases de datos contienen hasta 17000 imágenes pertenecientes a 10 clases diferentes donde los objetos aparecen en gran cantidad de posiciones frente al sensor, lo que provoca traslapes en los manifolds después de aplicar la transformada de Hotelling, lo que hace el reconocimiento más difícil[BLo1]. En casos como estos se tiene que recurrir a los métodos geométricos que tienen el grave inconveniente de depender del punto de vista en el que se toman las imágenes.

Este tipo de problemas se presentan también en los sistemas de vigilancia y seguimiento donde es necesario identificar un objeto como perteneciente a un conjunto de hipótesis o identificarlo en una secuencia que contiene varios otros objetos, en este caso se tiene que utilizar la apariencia del objeto en conjunto con un esquema geométrico.

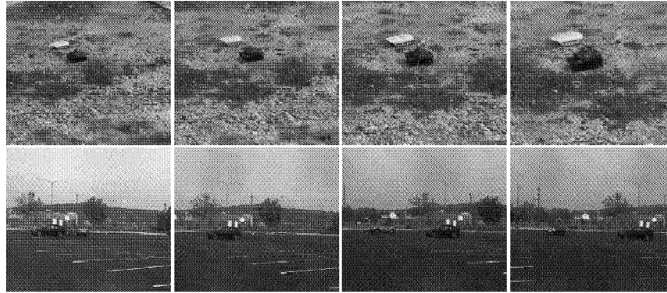


Figura 9: Secuencias de imágenes utilizadas en aplicaciones de vigilancia y seguimiento

Otra aplicación práctica es el reconocimiento de gestos y signos de la mano para traducción del lenguaje de signos para sordomudos o en la computación ubicua [JM01] donde los problemas antes mencionados de espacio de almacenamiento, traslape de manifolds y tamaño del espacio de búsqueda aparecen como graves desventajas.

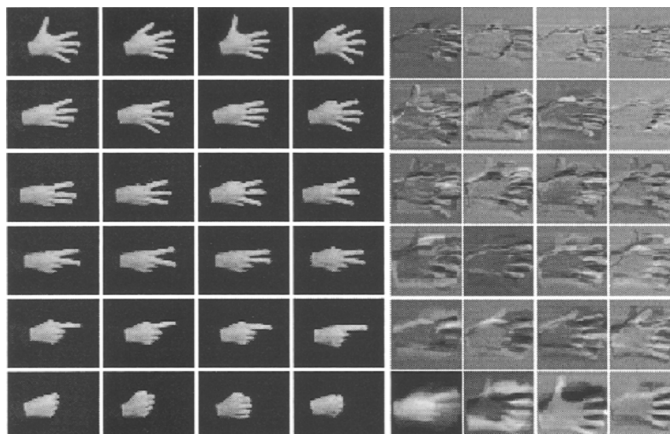


Figura 10: Ejemplos de modelos de apariencia de gestos de la mano

Para lograr el reconocimiento de objetos en cualquier posible circunstancia es necesario contar con modelos desde todos los puntos de vista posibles. Aunque esto último es imposible, se pueden obtener modelos de los objetos en condiciones que se consideren típicas para el problema que se quiere resolver incluyendo situaciones en las cuales el objeto de interés puede estar ocluido parcialmente o puede aparecer en compañía de otros como en la figura 8.

Sin embargo los problema de espacio de almacenamiento, traslape de manifolds y del tamaño del espacio de búsqueda permanecen. Este problema se ha intentado resolver mediante diversas técnicas como el muestreo no uniforme [Roboza] y la representación paramétrica de modelos [Nay96] o FisherFaces [PNBK97].

Por las situaciones antes mencionadas, en este trabajo de tesis se tiene por objetivo reducir el número de imágenes requeridas para representar el modelo de un objeto tomando en cuenta situaciones donde estas imágenes contienen alguna alteración que ocluya parcialmente al objeto principalmente aunque bajo la misma hipótesis se puede extender a modelos que contengan múltiples vistas.



## PROPUESTA DE SOLUCIÓN

---

Para mejorar el desempeño de los sistemas de visión basados en modelos de apariencia de objetos es indispensable reducir el número de imágenes necesarias para representar un objeto bajo este paradigma. Se han propuesto varios métodos para lograr este objetivo que trabajan directamente con el conjunto de imágenes como [Roboza, Ber09] y aprovechan características del objeto como la simetría [Ber09], el hecho de que las imágenes cercanas pueden aproximarse mediante una interpolación lineal y la cuasi-invarianza en la iluminación [Ale09].

En este trabajo de tesis se utilizarán los coeficientes obtenidos con la ecuación 3 que son los necesarios para reconstruir cualquier imagen del modelo. Se observó que la complejidad geométrica de los manifolds corresponde a la los objetos a los que corresponden. Se busca aprovechar esta propiedad suponiendo que los puntos de un manifold que representan imágenes muy similares yacen en el mismo subespacio lineal. Entonces el problema de reducir el número de imágenes de un modelo se reduce a identificar estos subespacios lineales y encontrar una imagen que represente a éstos. Para tal efecto se utilizó el algoritmo 1 que separa el manifold en varios segmentos lineales. Estos segmentos lineales representan grupos de imágenes semejantes entre sí que pueden ser remplazadas por un *ejemplar*. Este ejemplar es calculado como el promedio de los puntos que conforman el MLP. Con esto se asegura que los ejemplares representan conjuntos de imágenes similares que servirán para simplificar el modelo del objeto reduciendo el número de imágenes que lo componen.

Este método utiliza el cociente de la distancia euclídeana sobre la distancia geodésica como criterio de linealidad y sólo necesita ajustar los parámetros de umbral para éste criterio y el número de vecinos más cercanos que se utilizarán en el cálculo de las distancias geodésicas. En la siguiente sección se detalla el algoritmo utilizado para encontrar los subespacios lineales y los correspondientes ejemplares.

### 4.1 CÁLCULO DE LOS PARCHES LINEALES MAXIMALES

El cálculo de los MLPs se utiliza como paso intermedio para encontrar la distancia entre dos manifolds. Esta distancia se ha empleado en el reconocimiento de rostros mediante conjuntos de imágenes [RW34] y se explica en el apartado 2.5.1. Como ya se mencionó anteriormente, la idea es partir el conjunto de imágenes en varios parches que se encuentran embebidos en un subespacio lineal. A continuación se lista

el algoritmo utilizado para este efecto.

---

**Algorithm 1** Construcción de MLPs
 

---

```

i = 1, Ci = ∅, XT = ∅, XT = ∅, XR = X
while XT ≠ ∅ do
  Seleccionar aleatoriamente un punto en XR como semilla x1(i),
  Ci = {x1(i)}, XR = XR - {x1(i)}
  for ∀xmi ∈ Ci do
    Identificar cada uno de los k vecinos más cercanos xc como
    candidato
    if xc ∈ XR (DG(xc, xn(i))/DE(xc, xn(i))) < θ then
      Ci = Ci ∪ {xc}
      XR = XR - {xc}
    end if
  end for
  XT = ∪j=1i Cj, XR = X - XT; i ← i + 1, Ci = ∅
end while

```

---

La principal ventaja de este algoritmo es que sólo debe utilizarse una vez para cada manifold, con lo que se alcanza una gran eficiencia. Por otro lado, este método tiende a formar MLPs más grandes al principio y más pequeños al final lo que no parece afectar el rendimiento general del algoritmo[RW34].

Los ejemplares obtenidos se utilizarán para el reconocimiento de objetos comparando un conjunto de imágenes de prueba con éstos últimos y seleccionando el que resulte más cercano.

## EXPERIMENTOS

Para probar que el algoritmo propuesto efectivamente reduzca el número de imágenes necesarias para cada objeto se utilizaron dos modelos: uno *sencillo* que considera únicamente las imágenes de la biblioteca de modelos COIL-20 y ésta misma biblioteca con la adición de imágenes modificadas como se muestra en 11. La idea básica de estas modificaciones es que el algoritmo puede identificar las imágenes similares aún con estos cambios y así nos permitirá reducir modelos con varios grados de libertad.

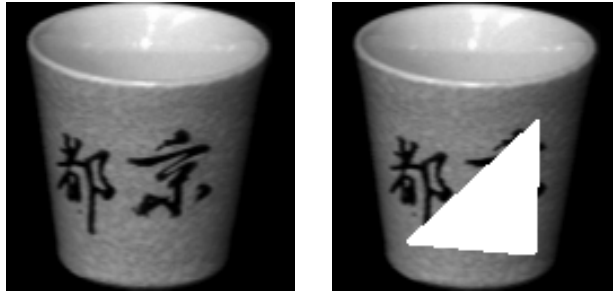


Figura 11: Ejemplos de las imágenes utilizadas en los experimentos

Los resultados obtenidos cuando se procesaron los modelos originales se muestran en la tabla 1. En general se consiguió una reducción importante del tamaño del modelo utilizando como parámetros 18 vecinos más cercanos y un criterio de linealidad de 1.01. Además, se utilizaron los modelos resultantes de utilizar la transformada de Hotelling con 36 componentes principales.

Objeto	Ejemplares Calculados	Porcentaje
1	11	84,72
2	12	83,33
3	13	81,94
4	11	84,72
5	14	80,55
6	11	84,72
7	12	83,33
8	13	81,94
9	11	84,72
10	11	84,72
11	11	84,72
12	14	80,55
13	12	83,33
14	15	79,16
15	11	84,72
16	13	81,94
17	13	81,94
18	14	80,55
19	13	81,94
20	13	81,94

Cuadro 1: Resultados de la aplicación del algoritmo1 a la biblioteca COIL-20

Como se puede observar, el porcentaje de reducción es alto comparado con los obtenidos por otros métodos como el expuesto en[Roboza]. También se puede ver que independientemente de la complejidad del objeto, se obtienen más o menos los mismos ejemplares, esto se atribuye a cuestiones tales como la simetría que presentan los objetos y a la medida de distancia utilizada, en este caso la euclideana. A continuación se muestra una figura en la que se aprecian los MLPs obtenidos para los objetos 5 y 6. El punto rojo indica la posición del ejemplar de cada uno de los MLPs.

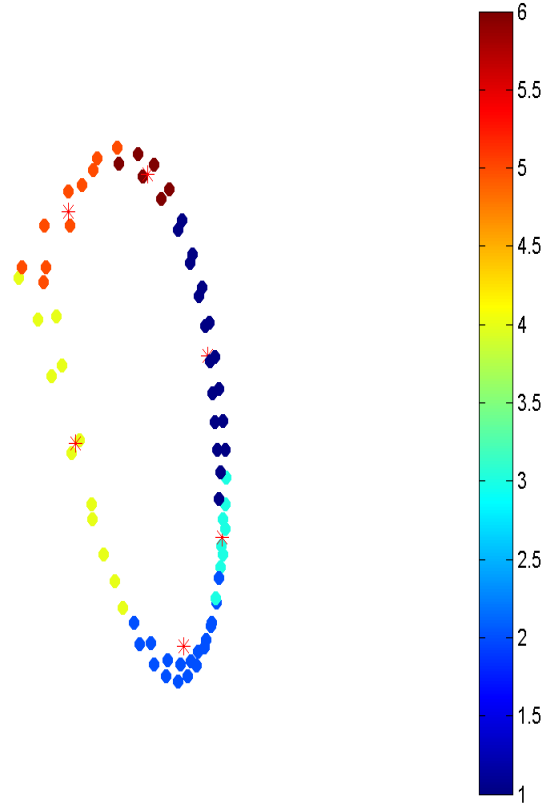


Figura 12: Algunos resultados para modelos sencillos. El punto marcado con un asterisco es el ejemplar correspondiente a cada MLP.

El mismo algoritmo con los mismos parámetros se probó aplicándolo a los modelos extendidos presentados en 11. En este caso se toman las modificaciones como diferentes DOF. Los resultados obtenidos se presentan en la tabla 2.

Objeto	Ejemplares Calculados	Porcentaje reducido
1	67	81,39
2	63	82,50
3	62	82,78
4	61	83,06
5	62	82,78
6	60	83,33
7	60	83,33
8	57	84,17
9	63	82,50
10	63	82,50
11	60	83,33
12	65	81,94
13	59	83,61
14	68	81,11
15	56	84,44
16	62	82,78
17	65	81,94
18	65	81,94
19	59	83,61
20	58	83,89

Cuadro 2: Resultados de la aplicación del algoritmo 1 a la biblioteca COIL-20. En este caso se agregaron 4 modelos con modificaciones.

Como puede verse, sucede lo mismo que para los modelos sencillos en lo que respecta al porcentaje de compresión.

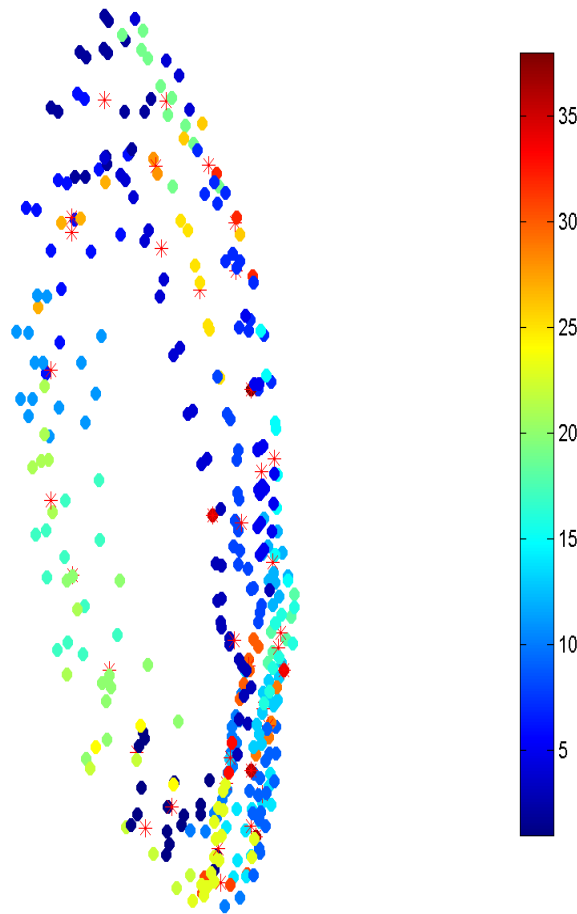


Figura 13: Algunos resultados para modelos con modificaciones

En la figura anterior 13 se pueden ver los manifolds correspondientes a los modelos procesados para el objeto 9 de la biblioteca COIL-20.

## 5.1 VERIFICACIÓN DE LOS MODELOS SIMPLIFICADOS

Para comprobar la efectividad de este procedimiento de simplificación de modelos se realizó una verificación del reconocimiento de objetos. Esta se realizó comparando aleatoriamente una imagen por cada modelo de objeto, con todos los ejemplares obtenidos. La comparación se realizó mediante la distancia euclídeana. Esta comprobación se realizó 500 veces y el porcentaje de reconocimiento presentado aquí es el promedio de los porcentajes de cada ejecución. El resultado obtenido es el porcentaje de aciertos que el programa obtuvo y que fue del 73.5% para los modelos simples de COIL-20, esto nos indica que tiene un grado de efectividad medio que puede ser mejorado evitando que el algoritmo favorezca a los subespacios que se calculan primero.



## CONCLUSIONES Y TRABAJO FUTURO

---

Actualmente, en las tareas de visión computacional el enfoque geométrico se usa extensivamente, aceptando la pérdida de información relevante del objeto. Este hecho es justificado por la posibilidad de extraer características únicas y representativas de los objetos como en el caso del reconocimiento de rostros, donde es posible definir una medida antropométrica única para cada individuo. No obstante, en otras tareas de visión puede no ser posible tomar ventaja de este tipo de particularidades y es aquí donde el enfoque basado en apariencia demuestra su valor, por ejemplo en el caso de los sistemas de visión activa.

Es importante aclarar que los modelos de apariencia y geométricos se complementan mutuamente y que es posible utilizarlos al mismo tiempo. En ambos casos es importante conseguir la simplificación de los modelos. Los resultados obtenidos con la técnica propuesta ponen de manifiesto que es posible simplificar, en gran medida, modelos con diferentes clases de modificaciones así como los modelos *simples*. También cabe mencionar que el algoritmo sólo tiene que ajustar dos parámetros únicamente: el criterio de linealidad y el número de vecinos para el cálculo de las distancias geodésicas. El cálculo de MLPs es realizado en un sólo paso. Esto ayuda a reducir el tiempo necesario para procesar los modelos aunque presenta el inconveniente de favorecer a los parches calculados primero.

**LIMITACIONES** El método propuesto no está exento de problemas, los que se han detectado son:

- El algoritmo calcula subespacios grandes primero.
- Las medidas usuales de distancia (euclídeana, por ejemplo) arrojan resultados malos.
- El número de parches lineales y su calidad para el reconocimiento se ven afectados directamente por la magnitud del criterio de linealidad y por el número de vecinos que se utilizan para el cálculo de la distancia geodésica.
- El hecho de que algunos ejemplares contengan más información que otros provoca que el esquema utilizado para la verificación se confunda y reconozca de manera errónea a los objetos.

- Ante las mismas entradas y los mismos parámetros cada ejecución independiente del algoritmo de simplificación arroja resultados ligeramente distintos.

La solución de estos problemas forma parte del trabajo que se realizará en un futuro próximo. También es interesante probar éste método en combinación con el algoritmo kmeans para estimar de mejor manera los ejemplares y así mejorar los resultados del reconocimiento. Por otro lado, se puede usar el número de ejemplares calculados para estimar el número  $k$  de grupos que se necesita como entrada. Además, la obtención de los parches lineales maximales es un paso intermedio en el cómputo de la distancia Manifold-Manifold[RW34]. Esta medida de similaridad será usada para el desarrollo de un algoritmo de agrupamiento para modelos de objetos basado en el ya mencionado kmeans ya que cumple con el requisito indispensable de mostrar la similaridad entre dos conjuntos de puntos. Este algoritmo se usará para generar una jerarquía de objetos que haga más eficientes las tareas de reconocimiento, identificación y adición de nuevos modelos. Este trabajo forma parte de un proyecto de maestría que actualmente se encuentra en desarrollo.

## BIBLIOGRAFÍA

---

- [Ale09] González Becerra Diego Alexander. Quasi-invarianza en iluminación para inferencia en apariencia., 2009.
- [Ber09] Barrientos Flores Berenice. Inferencia en apariencia para el reconocimiento de objetos tridimensionales. Master's thesis, FCC-BUAP, 2009.
- [BL01] Rama Chellapa Baoxin Li. Model-based temporal object verification using video. *IEEE transactions on image processing*, 10:897–908, 2001.
- [Clu09] *Clustering*. IEEE Press, 2009.
- [Coo04] Taylor C.J Cootes, T.F. Statistical model of appearance for computer vision. Technical report, Imaging Science Biomedical Engineering, University of Manchester, March 2004.
- [ea99] Nayar et al. An experimental comparison of appearance and geometric based recognition. *NA*, NA:NA, 1999.
- [GG07] Jianhong Wu Guojun Gan, Chaoqun Ma. *Data Clustering: Theory, Algorithms, and Applications*. ASA-SIAM Series on Statistics and Applied Probability, SIAM, 2007.
- [JB03] E. Guitter J. Boutier, P. Di Francesco. Geodesic distance in planar graphs. *Nuclear Physics B*, 663:535–567, 2003.
- [JM01] James L. Crowley Jerome Martin. An appearance based approach to gesture recognition. *GRAVIR-IMAG*, na:340–348, 2001.
- [kan06] *Crítica de la razón pura*. Tecnos, 2006.
- [Nay96] Nene Nayar, Murase. Parametric appearance representation. *Nayar, S.K., Poggio, T., (Eds.) Early Visual Learning, Oxford University Press. New York*, 1:131–160, 1996.
- [PNBK97] J. P. Hespanha P. N. Belhumeur and D. J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *PAMI*, 19:711–720, 1997.
- [Rey06] Isaid Reyes. Sistema para la adquisición de imágenes, modelado y reconocimiento de objetos empleando apariencia. Tesis de Licenciatura, FCC-BUAP, 2006.
- [Robo2a] Luis Carlos Altamirano Robles. *Metodologías para la reducción del número de imágenes requeridas para el reconocimiento de objetos articulados*. PhD thesis, IPN-CIC, 2002.

- [Robo2b] Luis Carlos Altamirano Robles. *Metodologías para la reducción del número de imágenes requeridas para el reconocimiento de objetos articulados*. PhD thesis, IPN-CIC, 2002.
- [RW34] Xilin Chen Wen Gao Ruiping Wang, Shiguang Shan. Manifold-manifold distance with application to face recognition based on image set. *a, b:b*, 1234.