

DEPARTAMENT D' ÒPTICA

CORRELACIONES INVARIANTES DE OBJETOS
TRIDIMENSIONALES

JOSÉ JAVIER VALLÉS VILAR

UNIVERSITAT DE VALÈNCIA
Servei de Publicacions
2009

Aquesta Tesi Doctoral va ser presentada a València el dia 25 de maig de 2009 davant un tribunal format per:

- Dr. Santiago Vallmitjana Rico
- Dr. Henri H. Arsenault
- Dra. María Sagrario Millán García-Varela
- Dr. David Mas Candela
- Dr. Ignacio Moreno Soriano

Va ser dirigida per:

Dr. Carlos Ferreira García

Dr. Javier García Monreal

Dra. Pascuala García Martínez

©Copyright: Servei de Publicacions
José Javier Vallés Vilar

Dipòsit legal: V-1036-2010

I.S.B.N.: 978-84-370-7563-1

Edita: Universitat de València

Servei de Publicacions

C/ Arts Gràfiques, 13 baix

46010 València

Spain

Telèfon:(0034)963864115

Departament d'Òptica

Facultat de Física



VNIVERSITAT
DE VALÈNCIA

Correlaciones invariantes de objetos tridimensionales

Tesis Doctoral

José Javier Vallés Vilar

Valencia, Mayo de 2009

D. Carlos Ferreira García
Catedrático de Óptica de la Universitat de València

D. Javier García Monreal
Catedrático de Óptica de la Universitat de València, y

Pascuala García Martínez
Profesora Titular de Óptica de la Universitat de València

CERTIFICAN

que la presente memoria, titulada “Correlaciones invariantes de objetos tridimensionales”, resume el trabajo de investigación realizado bajo nuestra dirección por José Javier Vallés Vilar, y constituye su Tesis Doctoral para optar al grado de Doctor en Física.

Y para que así conste, en cumplimiento de la legislación vigente, firman el presente certificado en Valencia, a 22 de Mayo de 2009.

Firmado:

Dr. Carlos Ferreira García

Dr. Javier García Monreal

Dra. Pascuala García Martínez

A mis padres

A Mari

Arturo: ¿Cuál es la cualidad más valiosa en un caballero?, ¿Valor, compasión, lealtad, humildad, ...? Responde tú, Merlin.

Merlin: Ah, ¿la mejor? Veréis, todas se funden cual metales para templar una buena espada.

Arturo: Deja la poesía, quiero una respuesta directa. ¿Cuál es?

Merlin: Os contestaré. La verdad. Sólo eso, así es. Por encima de todo, la verdad. Cuando un hombre miente mata una parte del mundo. Deberíais saberlo.

(*Excalibur*, John Boorman)

塞翁失马, *Y el anciano de la frontera perdió su caballo.*

(Tetrasilabo chino)

Agradecimientos

Antes de nada quiero agradecer a mis padres por darme la vida y apoyarme durante estos largos años de trabajo, y a Mari su cariño y comprensión.

A mis directores, Javier, Pas y Carlos, sin ellos este trabajo habría sido imposible. Y a Henri por la misma razón y además acogerme en Canadá.

A Adolfo, Kino, Joan, David, Vicente y Stanis, compañeros de “cubil” a lo largo de estos años. Y a Gabor, otro “compañero” de trabajo fiel.

A todos mis compañeros de estudios, que, a estas alturas, son tantos que no sabría ni por donde empezar.

Al areópago por sus múltiples enseñanzas a lo largo de los últimos años.

Y a Dama Roja, hogar de doctores y maestros del pelaje más variado.

Este trabajo ha recibido el apoyo económico del Ministerio de Educación y Ciencia, bajo el proyecto FIS2007-60626, de la Dirección General de Investigación i Transferència Tecnològica, proyecto IIARC0/2004/217, y la Agencia Valenciana de Ciencia y Tecnología (AVCT), proyecto GRUPOS03/117. Así mismo agradezco al Ministerio de Educación y Ciencia la concesión de la beca de Formación de Profesorado Universitario.

Índice

1 - Introducción general y contenidos	9
1.1 Introducción	9
1.2 La correlación aplicada al reconocimiento de objetos	10
1.3 Reconocimiento invariante de objetos	14
1.4 Reconocimiento de objetos tridimensionales	16
1.5 Objetivos y esquema general	19
2 - Reconocimiento de objetos tridimensionales con invariancia a cambios de iluminación	21
2.1 Introducción	21
2.2 Modelos de iluminación 3D	26
2.2.1 Analogía entre el modelo de iluminación elegido y los espacios vectoriales	32
2.3 Espacios vectoriales e imágenes	34
2.4 Métodos vectoriales de reconocimiento de objetos	36
2.4.1 El filtraje LACIF	42
2.4.2 El filtraje LADC	45
2.5 Implementación y resultados	48
2.5.1 Transformación de Karhunen-Loeve	56
2.5.2 Comportamiento del filtraje LADC frente a ruido y discriminación	62
2.6 Conclusiones	78
3 - Estimación de rotaciones de objetos tridimensionales	71
3.1 Introducción	71
3.2 Codificación de objetos 3D	77
3.2.1 Imágenes de rango	77
3.2.2 Transformada de Fourier de fase (PhFT)	78
3.2.3 3DOOM	86
3.3 Correlaciones y análisis en $SO(3)$	91
3.3.1 Correlación en $SO(3)$	91
3.3.2 Análisis de Fourier en $SO(3)$	94
3.4 Implementación y resultados	104
3.4.1 Cálculo de PHFTs y composición del 3DOOM	105
3.4.2 Desarrollo de funciones en S^2	108
3.4.3 Cálculo de correlaciones en $SO(3)$	110
3.4.4 Resultados de correlaciones en S^2	114

3.5 Conclusiones	119
4 - Reconocimiento de objetos tridimensionales con invariancia a cambios de escala	121
4.1 Introducción	121
4.2 Cambios de escala y PhFT	126
4.3 Métodos vectoriales para el reconocimiento invariante a escala	130
4.3.1 LACIF aplicado a invariancia a escala	130
4.3.2 Transformación de Karhunen-Loeve aplicada a la invariancia de escala	134
4.4 Implementación y resultados	136
4.5 Conclusiones	145
5 - Reconocimiento de objetos tridimensionales con invariancia a cambios de escala y rotaciones	147
5.1 Introducción	147
5.2 LACIF y 3DOOM	150
5.3 Implementación y resultados	153
5.3.1 Resultados con el 3DOOM parcial	154
5.3.2 Resultados con el 3DOOM completo	161
5.4 Conclusiones	166
6 - Conclusiones	167
Bibliografía	171
Apéndices	
Apéndice I - Abreviaturas usadas en la memoria	191
Apéndice II - Objetos 3D usados	193
Anexo - Recopilación de Publicaciones	195

Capítulo 1

Introducción general y contenidos

1.1 Introducción

La detección de objetos tridimensionales (3D) puede considerarse, hasta cierto punto, como una extensión al reconocimiento clásico en el caso bidimensional. Son múltiples las aplicaciones prácticas de este campo, como por ejemplo, las militares, las relacionadas con el tráfico, médicas, etc. Se pueden consultar referencias como [Rosen-98a, Rosen-98B], [Poon-99], [Campbell-01], [Javidi-02], [Khazdan-03], [Makadia-03], entre otras muchas, para tener una pequeña visión de la amplitud del tema de reconocimiento de objetos tridimensionales.

Además, el reconocimiento de objetos 3D no sólo tiene aplicaciones en los campos que se podrían considerar usuales, como pueden ser los anteriores, sino también en ámbitos como la geodesia [Blais-02], la química y la cristalografía [Blanco-97], la astronomía [Wandelt-01], y la biología [Petrich-00, Javidi-06], entre otros.

Una de las mayores dificultades que se presenta en el reconocimiento de objetos, ya sea bidimensional o tridimensional, es la diversidad de alteraciones que puede presentar el objeto que queremos reconocer. En general, una modificación del objeto dificulta o imposibilita su detección. Así, se hace necesario el desarrollo de técnicas específicas para reconocer un mismo objeto de manera invariante a una alteración concreta o un conjunto de ellas, tales como cambios de orientación, de escala, de proyección, etc.

En este trabajo planteamos diversos métodos para superar las limitaciones producidas por los diferentes cambios del objeto al reconocimiento tridimensional. Se estudian los casos concretos de rotaciones del objeto, cambios de escala, cambios de iluminación y la aplicación simultánea de cambios de escala y rotación.

Todos los métodos que hemos desarrollado tienen en común el uso de correlaciones en su implementación. La correlación es la herramienta básica del reconocimiento bidimensional, tanto por sus propiedades como por su facilidad de implementación, ya sea ésta digital mediante FFT¹ o analógica por medios ópticos. Además, al poderse interpretar este trabajo como una continuación del reconocimiento de objetos 2D realizado por el grupo al que pertenece el doctorando, resulta lógico el tomar en consideración técnicas análogas.

Por todo esto se explican brevemente las propiedades básicas de la correlación y su utilidad en el reconocimiento de objetos.

1.2 La correlación aplicada al reconocimiento de objetos

En esencia, el reconocimiento de objetos consiste en determinar si un objeto, conocido de antemano, se encuentra en un cierto entorno, viniendo definidos ambos por imágenes de los mismos.

Expresado de forma matemática para el caso concreto de objetos de dos dimensiones, \mathbb{R}^2 , el reconocimiento se basa en lo siguiente: Dada una función $f(x, y)$, el objeto, y otra función $g(x, y)$ de la que queremos saber

¹ Fast Fourier Transform, transformada rápida de Fourier en inglés.

si contiene al objeto, hemos de realizar una operación tal que su salida contenga información sobre lo similares que son las funciones $f(x, y)$ y $g(x, y)$.

Aunque existen diversas operaciones que sirven para este fin, una de las más sencillas es la correlación, comúnmente notada por $*$ y que se define como sigue para funciones en \mathbb{R}^2 :

$$f(x, y) * g(x, y) = corr(x, y) \equiv \int_{\mathbb{R}^2} f^*(x' - x, y' - y) \cdot g(x', y') dx' dy' \quad (1.1)$$

donde en este ámbito y por razones de implementación, la expresión (1.1) suele calcularse en términos de la transformada de Fourier:

$$f(x, y) * g(x, y) = F^{-1} \{ F^* \{ f(x, y) \} \cdot F \{ g(x, y) \} \} \quad (1.2)$$

siendo $F(u, v) \{ f(x, y) \}$ la transformada de Fourier definida de la forma usual:

$$F(u, v) \{ f(x, y) \} = \int_{\mathbb{R}^2} e^{-i2\pi(ux+vy)} f(x, y) dx dy \quad (1.3)$$

La Ec. (1.1) significa lo siguiente: Para cada punto del resultado, el valor de la correlación es la integral del producto de la primera función conjugada y centrada en el punto en el que evaluamos la correlación, con la segunda función.

Si aplicamos la correlación a un caso fuertemente simplificado, pongamos como ejemplo que las dos funciones $f(x,y)$ y $g(x,y)$ son discretas y sólo pueden tomar como valor 0 ó 1, se puede ver gráficamente como se calcula la correlación en las siguientes figuras:

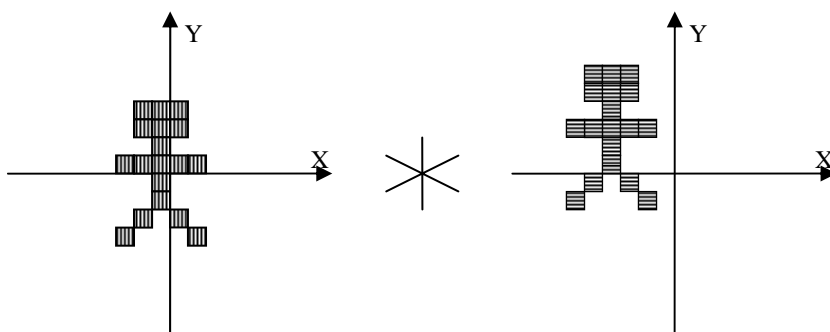
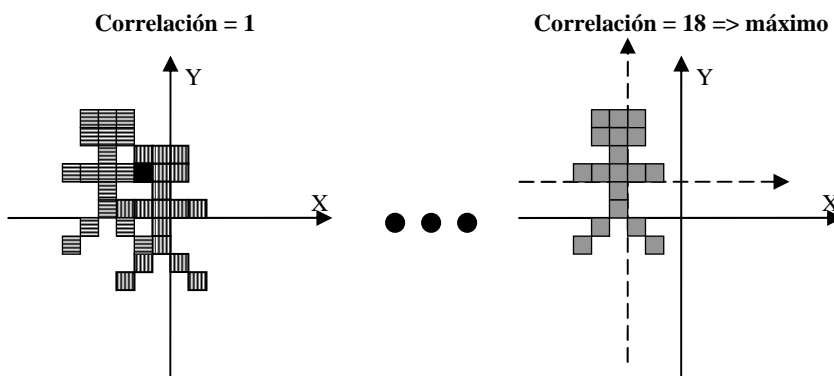


Figura 1.1 Funciones entre las que hacemos la correlación



**Figura 1.2 Cálculo de la correlación para el punto (0,0) a la izquierda
y para (-3,2) a la derecha**

Para cada uno de los puntos de la correlación, su valor no es más que el área que tienen en común la primera función, centrada en el punto de la salida, y la segunda sin modificar. Así, donde la correlación tenga un máximo eso indicará que el solapamiento entre el objeto y la imagen es máximo y ello

indica que el objeto $f(x, y)$ está contenido en $g(x, y)$ y centrado en esa posición.

Otra razón de la importancia de la correlación viene de su relación con el error cuadrático medio, véase [Duda-73]. Tomándose $g(x, y)$ como función de entrada, se define el MSE¹ respecto de una referencia $f(x, y)$ de la siguiente forma:

$$MSE(f; g) = \int_{\mathbb{R}^2} |f(x, y) - g(x, y)|^2 dx dy \quad (1.4)$$

Partiendo de la definición de la Ec. (1.4) puede demostrarse la siguiente propiedad [Duda-73]:

$$MSE(f; g) = \int_{\mathbb{R}^2} |f(x, y)|^2 dx dy + \int_{\mathbb{R}^2} |g(x, y)|^2 dx dy - 2f(x, y) * g(x, y) \quad (1.5)$$

La Ec. (1.5) muestra como la correlación y el error cuadrático medio están ligadas de forma inversa; si la correlación se hace máxima, el error se hace mínimo y viceversa. Esta propiedad estadística, unida al hecho de ser un criterio de similitud entre funciones, así como por la facilidad de implementación digital mediante FFT, [Walker-96], u óptica, [VanderLugt-64, Weaver-66], hacen de la correlación la herramienta básica en el reconocimiento de objetos.

¹ La notación MSE viene del inglés Mean Squared Error, error cuadrático medio.

1.3 Reconocimiento invariante de objetos

En el apartado anterior nos hemos ceñido al caso en que los objetos bidimensionales no han sido alterados en forma alguna. Pero el tratar con objetos que han sido modificados respecto a su forma original plantea dificultades que han de ser examinadas en detalle.

Consideremos un objeto, ya sea 2D o 3D. Si se altera de alguna forma, ya sea rotándolo, cambiándolo de escala, etc., y calculamos la correlación con una escena donde esté una versión sin alterar del mismo, ya no es posible alcanzar el máximo solapamiento entre éste y la imagen. Si se altera el objeto ya no “encajará” en la escena de la misma forma, por tanto la correlación daría un resultado inferior al esperado y fallaríamos al detectar el objeto.

Debido a esto, se hacen necesarios métodos de reconocimiento invariante a las alteraciones del objeto, es decir que den la misma respuesta frente a las diversas versiones del objeto, constituyendo la búsqueda de estos métodos una línea básica dentro del reconocimiento de formas.

Considerándolas de forma individual, la alteración más fácil de abordar en reconocimiento mediante correlaciones es la traslación del objeto, ya que la propia correlación es invariante a traslaciones de forma directa. Un desplazamiento del objeto causa que la respuesta de la correlación se desplace, pero no altera su valor.

Desgraciadamente el resto de alteraciones no son triviales. Cronológicamente los primeros intentos de tener sistemas de reconocimiento invariantes fueron realizados por Casasent, llegando a desarrollar un método de reconocimiento invariante a cambios de escala y rotaciones, [Casasent-76a] y [Casasent-76b], basado en aplicar una transformación logarítmica a las coordenadas del

objeto. Este método tiene el inconveniente de que se pierde la invariancia a traslaciones. Trabajos posteriores de Mendlovic *et al*, Rosen *et al* y Li *et al*, [Mendlovic-88, Rosen-89, Li-03], desarrollaron ideas similares, aplicando transformaciones de coordenadas sobre los objetos. En concreto desarrollos en términos de series de funciones, los armónicos radiales de Mellin y armónicos logarítmicos radiales, respectivamente, y tomando como nueva función objeto un solo término del desarrollo del objeto original en la nueva base.

Esta idea es la que se había aplicado anteriormente al caso de la invariancia a rotaciones. Hsu *et al*, [Hsu-82a], aplicaron un desarrollo de la función objeto en armónicos circulares, utilizando un filtro adaptado a un solo término del desarrollo para llevar a cabo la detección con invariancia a rotaciones, [Hsu-82a, Hsu-82b].

Los planteamientos anteriores tienen el inconveniente de limitarse a usar sólo una fracción de la información global, reduciendo así su capacidad de discriminación. Debido a estos inconvenientes, se desarrollaron posteriormente métodos que permiten elegir con mayor precisión los parámetros de una componente del desarrollo en serie, [Moya-97, Hsu-82b, Garcia-Martinez-95], y otros métodos que permiten usar simultáneamente varios términos del desarrollo mediante técnicas de multiplexado, [Zalevsky-97, Esteve-Taboada-00].

Otra alteración del objeto que consideraremos en este trabajo será el cambio de iluminación en la escena. En el caso de objetos 3D, la invariancia a iluminaciones, principalmente ha sido abordada como un caso particular del reconocimiento de caras, [Tian-99, Zhou-05], de gran importancia práctica y enorme actualidad. En dos dimensiones, una variación basada en la adición de un factor multiplicativo constante o aditivo global es análoga a un cambio

en la intensidad de la fuente de iluminación y en este caso se puede lograr invariancia, tanto con filtros normalizados [Dickey-91], como con el LACIF¹ [Lefebvre-02], que se explicará posteriormente.

Más tarde, el LACIF también fue aplicado con éxito al reconocimiento bajo transformaciones afines en imágenes 2D, basadas en sumar funciones gradientes o combinaciones de gradientes sobre el objeto a reconocer. Esta alteración se puede interpretar tanto como un efecto similar a cambiar la posición de la fuente de iluminación en 3D, véase [Roy-04], como otras alteraciones más complejas y no relacionadas con cambios de iluminación en dos dimensiones, como las rotaciones a lo largo de un eje perpendicular al eje de visión y la adición de patrones de camuflaje, [Lefebvre-03, Arsenault-05].

El LACIF será de gran importancia en este trabajo y se describirá en profundidad en el capítulo 2. De momento adelantaremos que se basa en considerar las imágenes como vectores y al conjunto de posibles imágenes alteradas como un subespacio vectorial, siendo la pertenencia al subespacio el criterio de reconocimiento.

1.4 Reconocimiento de objetos tridimensionales

Aparte de las alteraciones a las que se pueda someter el objeto, el propio reconocimiento de objetos 3D plantea una serie de problemas cuya naturaleza es diferente a los que nos encontramos en el caso bidimensional.

¹ Locally Adaptive Contrast Invariant Filter, Filtro invariante a contraste localmente adaptativo en inglés

En primer lugar, el hecho de tener una dimensión espacial adicional amplía la necesidad de cálculos e implica la obtención y almacenamiento de una mayor cantidad de información. Una posibilidad es usar un planteamiento directo en el reconocimiento 3D, es decir, aplicar una ecuación análoga a la correlación (1.1) pero en 3D. Esto fue llevado a cabo por medios ópticos por Rosen, [Rosen-98a, Rosen-98b, Li-02], siendo necesario conocer la escena $g(x, y, z)$ completa, son múltiples las dificultades que esto conlleva y aumenta de forma muy notable los cálculos a realizar.

Si a lo anterior añadimos el considerar algún tipo de alteración al objeto, tenemos que el aumento de complejidad en el caso de reconocimiento 3D es superior al experimentado en el 2D. Por ejemplo, sin ningún tipo de invariancias, el caso de reconocimiento 2D es un problema con dos grados de libertad mientras que el 3D tiene tres grados de libertad. Además considerando rotaciones en el objeto para identificarlo en dos dimensiones se han de concretar tres grados de libertad, dos coordenadas espaciales más la rotación, mientras que el mismo caso de rotaciones aplicado a 3D asciende a seis grados de libertad, tres coordenadas espaciales más un mínimo de tres ángulos, o parámetros, que caractericen la rotación. Es decir, para una alteración equivalente, la complejidad aumenta en mayor medida en el caso tridimensional.

Vistas estas dificultades, se pueden agrupar los métodos de reconocimiento 3D según la manera en la que se aborde el problema. Partiendo de lo más general, son posibles tanto planteamientos digitales, muchos de ellos basados en redes neuronales, véase [Tadeusz-98, Takatsuka-98], como planteamientos ópticos o de otra naturaleza. De entre estas opciones, en este trabajo nos limitaremos a un planteamiento del problema basado en la óptica, aun cuando la implementación sea digital.

Dentro del enfoque óptico, existe abundante bibliografía, en la que cada método de reconocimiento tiene sus particularidades concretas, aunque se puedan extraer ciertas generalidades. Dadas las dificultades prácticas que entraña el conocer toda la escena $g(x, y, z)$ se puede usar como criterio de clasificación la información que se usa, toda o una fracción, codificada o no, o cómo se ha obtenido esta información.

Así, existen métodos de reconocimiento 3D basados en registros holográficos de la escena, como los desarrollados por Javidi *et al* y Tajahuerce *et al*, [Javidi-00, Tajahuerce-01]. En ellos se captura, en forma de holograma digital, la distribución de amplitud compleja de un objeto iluminado. Como este holograma contiene información del objeto 3D a reconocer visto desde diferentes perspectivas, se puede usar para reconocerlo y dar una estimación de sus posibles cambios de orientación mediante técnicas de correlación.

Por otra parte, la información más fácilmente disponible de un entorno 3D puede ser una proyección 2D, es decir una imagen convencional, o un conjunto de éstas. Es por ello que existen métodos de reconocimiento de objetos 3D que parten de un conocimiento de la escena obtenido a partir de diversas proyecciones 2D de la misma y correlacionando éstas con la referencia, como los seguidos por Matoba *et al* y Pu *et al*, [Matoba-01, Pu-99]. O a partir de diversos cortes planos de la escena según un eje [Bamler-82]. En el capítulo 2 expondremos un método de reconocimiento 3D invariante a iluminaciones que toma como señal de entrada una proyección 2D generada por ordenador de la escena 3D.

Una tercera opción consiste en codificar la información del objeto 3D, o parte de ella, en una distribución más manejable y usar esta información como

punto de partida. Este planteamiento será el que se use en los capítulos 3 y siguientes, en los cuales desarrollaremos métodos de reconocimiento que toman como entrada una imagen de rango modificada, o un conjunto de ellas. Las imágenes de rango constituyen una forma estándar de representar objetos tridimensionales, véase [Takeda-83] y [Rioux-84] entre otros, basada, en esencia, en asignar un tono de gris a una imagen 2D en función de la distancia entre el objeto 3D y un plano de referencia dado.

Una vez obtenidas estas imágenes de rango se procede a su codificación matemática. Este proceso, como se verá con más detalle en capítulos siguientes, transforma ciertas alteraciones geométricas del objeto 3D original en otras mucho más manejables en el espacio de las imágenes de rango codificadas, permitiéndonos el desarrollar métodos de reconocimiento más sencillos.

1.5 Objetivos y esquema general

El principal objetivo de este trabajo es desarrollar métodos de reconocimiento que permitan reconocer objetos 3D sometidos a diversas alteraciones, o conjunto de ellas, estando estructurado este trabajo en capítulos, en cada uno de los cuales desarrollaremos un método de reconocimiento invariante frente a una alteración particular.

Concretamente, comenzaremos abordando el caso de reconocimiento frente a cambios de iluminación, en el capítulo 2. Continuaremos en los capítulos 3, 4 y 5, con diversas alteraciones geométricas, como cambios en la rotación y la escala.

Esta distinción entre los capítulos no es casual sino que responde a la metodología usada. En el capítulo 2, [Vallés-06], veremos cómo el reconocimiento invariante frente a una alteración, en principio compleja, los cambios de iluminación, equivale a determinar la pertenencia de un vector a un determinado subespacio vectorial.

En el resto de capítulos, véase [García-03], [Vallés-07] y [Vallés-08], usaremos una codificación que nos permita reducir ciertas modificaciones del objeto 3D a problemas similares al tratado en el segundo capítulo. En el capítulo 3 se tratará en detalle la codificación que proponemos, viéndose como la aplicación de dicha codificación a las imágenes de rango, una forma estándar de representar objetos 3D, [Chang-97, Laurin-95], posee unas propiedades que permiten su aplicación directa al reconocimiento y estimación de rotaciones de objetos tridimensionales [García-03].

En el cuarto capítulo veremos cómo una de las propiedades de dicha codificación nos permite reducir la complejidad de la variación de escala a una mera multiplicación de la imagen de rango por una constante [Vallés-07].

Y, para finalizar, en el quinto y último capítulo trataremos, con una combinación de los métodos desarrollados en los capítulos 3 y 4, el caso en el que se someta a un objeto simultáneamente a cambios de escala y rotación que, nuevamente, se puede reducir a un caso más sencillo matemáticamente mediante un proceso de codificación [Vallés-08].

Capítulo 2

Reconocimiento de objetos tridimensionales con invariancia a cambios de iluminación

2.1 Introducción

Tanto en el caso del reconocimiento de objetos con invariancia a rotaciones, como en el de cambios de escala, los problemas a tratar se definen de la misma forma tanto en 2D como en 3D. Un cambio de escala del objeto, o una rotación denotan la misma operación, independientemente del número de dimensiones que se maneje.

En cambio, cuando se habla de un cambio de iluminación en dos dimensiones, nos podemos referir de forma genérica tanto a una variación en el brillo, en el contraste, una alteración sencilla de la forma $a \cdot f(x, y)$ siendo a una constante, o incluso otras más complejas. Por contra, en el caso 3D un cambio de iluminación está asociado a un cambio en las fuentes de luz de una escena, ya sea en intensidad o posición, y esto puede afectar significativamente a la apariencia del objeto haciendo más complejo el reconocimiento.

Centrándonos en el caso bidimensional y la invariancia a alteraciones del tipo $a \cdot f(x, y)$ podemos destacar las siguientes aportaciones, [Arsenault-85], [Dickey-91], [Zhang-99], [García-Martínez-00] y [Arsenault-00].

En particular, el método sugerido por Arsenault y Delisle, [Arsenault-85], está basado en la descomposición en armónicos circulares. Al descomponerse

el objeto a reconocer en dicha base y normalizando una componente de dicho desarrollo, se puede obtener reconocimiento invariante a la alteración $a \cdot f(x, y)$.

Los filtros normalizados propuestos por Dickey y Romero, [Dickey-91], serán explicados en detalle más adelante. A modo de avance, se puede decir que consisten en una versión normalizada del filtro adaptado clásico al dividir éste por el total de la energía del plano de correlación. Con ello se elimina la influencia de la constante a desconocida y se obtiene la invariancia deseada.

Con un planteamiento diferente a los anteriores, Zhang *et al.*, [Zhang-99], propusieron un método basado en una implementación óptica, mediante un JTC¹ de la correlación morfológica desarrollada por García-Martínez *et al.*, [García-Martínez-98].

García-Martínez *et al.*, [García-Martínez-00], demostraron que era posible obtener invariancia a esa misma transformación mediante la aplicación de una correlación no lineal por rodajas definida SONG² [García-Martínez-99].

Por su parte, Arsenault y Lefebvre, [Arsenault-00], propusieron un método basado en una transformación homomórfica sobre la escena de entrada. Dicha transformación consiste, básicamente, en tomar un logaritmo de la escena, con lo que la alteración $a \cdot f(x, y)$ pasa a convertirse en otra del tipo $\ln a + \ln[f(x, y)]$, de más sencillo manejo.

¹ Joint Transform Correlator, correlador de transformada conjunta, en inglés.

² Sliced Orthogonal Nonlinear Generalized correlation, correlación no lineal por rodajas generalizada, en inglés.

También hay diversas opciones, [Chalasincka-Macukow-93], [Lefebvre-02], [Roy-04] y [Arsenault-05], en el reconocimiento de objetos 2D invariante con respecto a variaciones de las formas $a \cdot f(x, y) + b$, $a \cdot f(x, y) + b \cdot g(x, y)$, donde b es otra constante y $g(x, y)$ otra función determinada, u otro tipo de transformaciones más complejas.

Así, Chalasincka-Macukow *et al.*, [Chalasincka-Macukow-93], demostraron que era posible obtener invariancia para el caso de transformaciones lineales de intensidad de la imagen, mediante el uso de correlaciones sólo de fase. Por contra, este método funciona bien sólo si el objeto está segmentado de antemano.

El hecho de necesitar una segmentación previa al reconocimiento es una limitación de la que carece el LACIF desarrollado por Lefebvre *et al.*, [Lefebvre-02], [Lefebvre-03], [Roy-04] y [Arsenault-05]. El LACIF, Locally Adaptive Contrast-Invariant Filter, es un filtraje que permite obtener una respuesta final invariante tanto a transformaciones de la forma $a \cdot f(x, y) + b$ como a otras más complejas y ha sido aplicado con éxito al reconocimiento frente a variaciones de iluminación 2D, rotaciones del plano de la imagen o adición de un patrón de camuflaje al objeto original. Como veremos más adelante, existen modelos de iluminación que se basan en descomponer una vista de una imagen 2D de un objeto 3D, iluminado de una cierta manera, en una combinación de otras imágenes. Este hecho lleva a poder considerar al conjunto de todas las posibles imágenes iluminadas de un objeto como un subespacio vectorial. Dada su importancia en nuestro trabajo, el LACIF será descrito en detalle más adelante en este mismo capítulo.

En cuanto al caso tridimensional de reconocimiento de objetos con invariancia a iluminaciones, también existen múltiples líneas de investigación. Por una parte, son interesantes los trabajos para obtener un reconocimiento invariante de texturas 3D, [Healy-95] y [Yang-05].

Así, el método propuesto por Healy y Wang, [Healy-95], consiste en calcular una matriz con las correlaciones entre las diferentes intensidades que posee el objeto, en cada una de las frecuencias del espectro de la luz con que está iluminado, y es invariante frente a un factor multiplicativo de la intensidad de la imagen.

Otro método de reconocimiento de texturas en objetos 3D con invariancia a iluminaciones, fue propuesto por Yang y Al-Rawi, [Yang-05], y se basa en el cálculo de los momentos y de las normas asociadas a cada banda de color. A partir de esto se puede obtener un reconocimiento invariante de texturas frente a cambios de iluminación del tipo $a \cdot f(x, y) + b$.

Con un planteamiento más digital del reconocimiento de objetos 3D con invariancia a cambios de iluminación existen multitud de trabajos basados en el reconocimiento de caras. Este es un campo de gran actualidad y utilidad práctica y como tal recibe gran interés y muchos recursos investigadores sobre todo durante los últimos años, [Gross-04], [Ishiyama-07], [Li-04] y [Zhang-03].

El trabajo de Gross *et al.*, [Gross-04], propone un método invariante tanto a la iluminación de la cara a reconocer, como a la postura de la persona. Para ello considera el conjunto de poses como un vector y obtiene la clasificación aplicando técnicas de autoestados al conjunto de estas imágenes en diferentes poses e iluminaciones, un concepto de gran interés.

Otra opción es la sugerida por Ishiyama *et al.*, [Ishiyama-07], en la que el reconocimiento también es invariante a iluminaciones y permite también discriminar con qué pose se encuentra el objeto. Un enfoque puramente basado en redes neuronales también es posible, como el expuesto por Li *et al.*, [Li-04], en la que el reconocimiento se basa en usar un conjunto de entrenamiento que contenga las caras a reconocer, con diferentes iluminaciones *típicas*, lo que permite un reconocimiento invariante a esta alteración.

Otro planteamiento más sobre el reconocimiento de caras es el propuesto por Zhang y Samaras, [Zhang-03], que presenta un cierto parecido al que se desarrolla en este trabajo. En él el reconocimiento se lleva a cabo usando la descomposición de las imágenes de la cara en una base de armónicos; conociendo esta base, se puede calcular la verosimilitud de que una escena dada, con cualquier iluminación, contenga la cara a reconocer.

Es importante mencionar, que es aplicado en la literatura el hecho de considerar las versiones con diferente iluminación de un objeto 3D como un *casi* subespacio vectorial [Hallinan-94], [Belhumeur-96], [Nayar-96] y [Basri-03]. En estos modelos, el objeto iluminado se define como una suma de diferentes imágenes, con lo que la analogía entre subespacio vectorial e iluminación es inmediata.

Es decir, en algunos casos el estudio de la invariancia a iluminaciones puede ser equivalente a detectar la pertenencia del objeto a un determinado subespacio vectorial. Esta será la idea que proponemos en este capítulo. Así, empezaremos definiendo el modelo que utilizaremos y demostrando la analogía entre esos modelos de iluminación y los espacios vectoriales. Una vez comprobado que el conjunto de imágenes que definen nuestro modelo de

iluminación tiene estructura de espacio vectorial, definiremos la operación de reconocimiento de objetos 3D invariante a iluminaciones como aquella basada en determinar la pertenencia de un cierto vector a un cierto subespacio. Explicaremos cuáles son los métodos matemáticos de reconocimiento con que trabajaremos, como el LACIF y nuestra propuesta, el LADC¹. Estudiaremos ambos métodos desde un punto de vista de espacios vectoriales y compararemos los resultados que se obtienen con ambos métodos. Para finalizar, estudiaremos cómo mejorar el resultado ajustando el tiempo de cálculo de forma óptima a la complejidad del problema que queramos considerar.

En su mayor parte este capítulo está basado en el artículo [Vallés-06].

2.2 Modelos de iluminación 3D

El primer aspecto a tratar en esta parte de la memoria será cómo modelar la iluminación en objetos 3D. En la literatura existen diferentes modelos de iluminación, [Foley-96] y [Hear-04], siendo las principales diferencias entre ellos los fenómenos físicos que se tienen en cuenta en la descripción. En cualquier caso, la problemática de todos estos métodos es la siguiente: dada una escena con objetos en ella y una distribución de luz en el espacio, determinar cuál es la imagen que verá un observador que esté situado en un punto determinado. Y así, por un lado tendremos que considerar tanto las iluminaciones direccionales como las no direccionales y por otra parte estudiaremos la naturaleza de la superficie exterior del objeto a iluminar.

¹ Local Angular Distance Correlation, del inglés correlación por distancia angular local.

En el modelo de iluminación que hemos planteado, comenzaremos estudiando la influencia de la llamada iluminación ambiental definida como el resultado de las múltiples reflexiones de esa luz ambiente que se producen en el entorno de un objeto. Así, estas reflexiones producen un fondo de luz no direccional y uniforme en todo el espacio, dentro del cuál está inmerso el objeto.

Definimos una distribución de iluminación ambiental 3D como:

$$I_1(x, y, z) = I_a \cdot k_a(x, y, z) \quad (2.1)$$

donde I_a es una constante que representa la intensidad de la iluminación ambiental y $k_a(x, y, z)$ el coeficiente de reflexión ambiental, variable según el punto de espacio que consideremos.

Además de la iluminación ambiental vamos a considerar la iluminación difusa, que posee un comportamiento diferente [Foley-96, Hear-04]. Por simplicidad supondremos que la fuente de luz está situada a gran distancia, de tal manera que los rayos que emite son paralelos. Esta suposición no representa ninguna limitación práctica, ya que cualquier iluminación más compleja se puede descomponer en la contribución de múltiples iluminaciones planas. Además supondremos que la superficie del objeto 3D no posee brillo y por tanto la reflexión de la luz vendrá dada por la ley de Lambert para una iluminación difusa.

La siguiente figura representa la posición de la fuente así como los vectores que intervienen en la ley lambertiana tomando como referencia la superficie 3D para el caso de iluminación difusa.

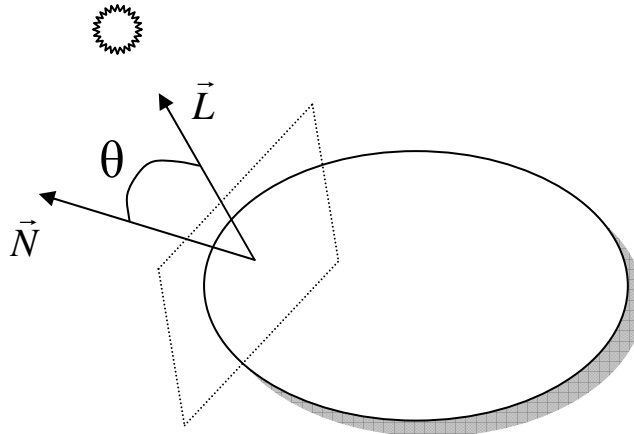


Figura 2.1 Modelo de iluminación difusa

Donde \vec{L} es el vector de iluminación, \vec{N} el vector normal a la superficie en el punto considerado y θ el ángulo formado por ambos vectores. Nótese que estamos siguiendo una convención de signos tal que \vec{L} va de la superficie a la fuente y no al revés.

En este modelo de iluminación difusa, se cumple que la luz es reflejada por la superficie de manera adireccional, es decir no importa el ángulo de observación. Además se verifica también que la componente de la luz difusa reflejada por dicha superficie es proporcional al coseno del ángulo de incidencia. Con todo esto, la expresión matemática para la iluminación difusa viene dada por la siguiente expresión.

$$I_2(x, y, z) = I_p k_l(x, y, z) (\vec{L} \cdot \vec{N}(x, y, z)) \quad (2.2)$$

donde I_p es la intensidad de la fuente direccional y $k_l(x, y, z)$ la reflectividad difusa del objeto iluminado.

Si ahora consideramos la actuación simultánea de la iluminación ambiental Ec. (2.1), y la difusa, Ec. (2.2) obtenemos:

$$I(x, y, z) = I_a \cdot k_a(x, y, z) + I_p k_l(x, y, z) (\vec{L} \cdot \vec{N}(x, y, z)) \quad (2.3)$$

Se hace notar que \vec{L} en la Ec. (2.3) es un vector constante, mientras que el vector $\vec{N}(x, y, z)$ es una función en \mathbb{R}^3 , esto se debe a que \vec{L} es el vector que representa la dirección de la iluminación, por lo que su valor es uniforme en todo el espacio, en cambio $\vec{N}(x, y, z)$ da la orientación del vector normal a la superficie del objeto iluminado en cada punto del espacio siendo, por tanto, una función de la posición en 3 dimensiones.

La ecuación (2.3) nos da la distribución de intensidad para todo el espacio tridimensional. Así, la imagen representada por la Ec. (2.3) es una imagen tridimensional. Con el fin de simplificar el cálculo y sin perder el carácter tridimensional del objeto, vamos a considerar una determinada vista o perspectiva del objeto tridimensional desde un punto de observación dado. En la siguiente figura mostramos el aspecto que tiene un objeto 3D antes de ser iluminado y después de aplicar la iluminación, habiendo fijado una vista concreta:

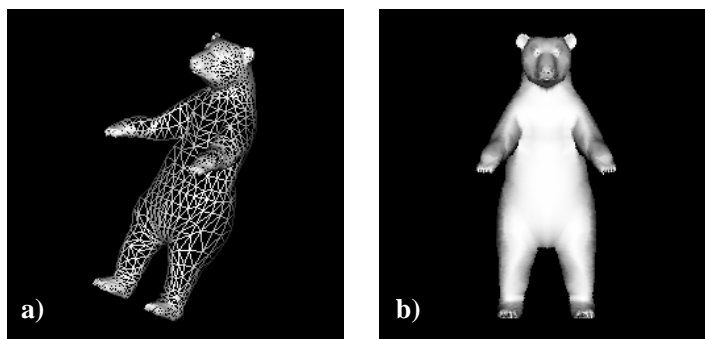


Figura 2.2 En a) se muestra un objeto 3D, en b) una vista concreta del mismo objeto iluminado de forma frontal

Es importante aclarar que aunque en este trabajo se usen imágenes 2D del objeto, similares a la imagen b) de la figura anterior, estas imágenes son una vista concreta del objeto 3D completo que se representa en a), que es el objeto que realmente se quiere reconocer con el método propuesto.

También es importante el hecho de que el objeto antes de ser iluminado, está representado por una malla de facetas y vértices, tal y como ilustra la Figura 2.2 a). Dicha malla viene descrita por un par de matrices, uno que contiene las coordenadas de los vértices y otro que indica cuales son las aristas que forman la malla. Todos los objetos usados a lo largo de este trabajo se representan de la misma forma y fueron recopilados de diferentes webs públicas.

Asimismo, nótese qué, como resultado de la ecuación (2.3), se pueden obtener valores de intensidad negativos en aquellos puntos en los que se cumpla que $(\vec{L} \cdot \vec{N}) \leq 0$. En sentido geométrico, dichos puntos se corresponden con una situación en donde la fuente debería de estar dentro del objeto u ocluida por parte de él y, por tanto, no pueden recibir luz de la fuente direccional, debiendo asignar su contribución a cero. Los puntos que cumplen esta condición corresponden a las sombras en la imagen. De esta forma, si consideramos únicamente los valores positivos del coseno y una particular vista del objeto tridimensional, podemos escribir el modelo de iluminación considerado en esta memoria como:

$$I(x, y) = I_a \cdot k_a(x, y) + \begin{cases} I_a k_f(x, y) (L_x N_x(x, y) + L_y N_y(x, y) + L_z N_z(x, y)) & \text{si } (\vec{L} \cdot \vec{N}) \geq 0 \\ 0 & \text{en el resto} \end{cases} \quad (2.4)$$

donde $\{N_i(x, y)\}_{i=x,y,z}$ y $\{L_i\}_{i=x,y,z}$ son las componentes cartesianas de \vec{N} y \vec{L} , respectivamente. Nótese que los L_i son escalares, mientras que los $N_i(x, y)$ son funciones 2D que representan las normales a la superficie del objeto 3D, aun siendo ambos componentes cartesianas de vectores.

La Ec. (2.4) será el modelo de iluminación que usaremos en lo que resta de memoria. Existen otros modelos más complejos que no sólo tienen en cuenta superficies mates sino otras de tipo metálico o brillante. Estos modelos se llaman de iluminación especular y su contribución a la imagen final se puede apreciar en la siguiente figura, en la que mostramos las diferentes contribuciones a la iluminación de un objeto 3D:

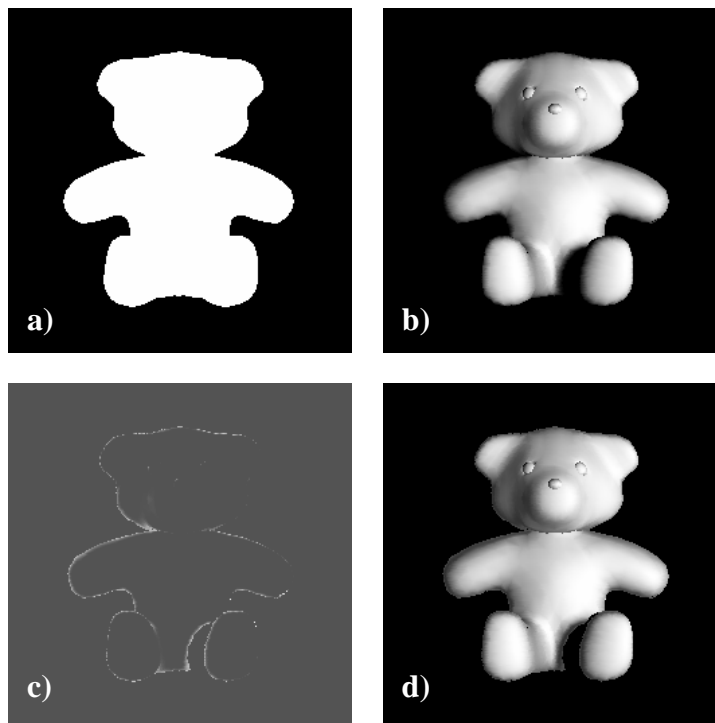


Figura 2.3 En a) se muestra un objeto 3D sometido a iluminación ambiental, en b) y c) se tiene el mismo objeto pero sólo con iluminación difusa y especular respectivamente y d) muestra el resultado conjunto de los tres tipos de iluminación

Es evidente de la figura que la inmensa mayoría de la información sobre el objeto 3D nos la aportan las iluminaciones ambiental, que da su silueta, y difusa, que da su forma. La iluminación especular añade poca información a las otras dos anteriores, siendo esta la razón por la que se ha omitido a la hora de implementar nuestro método. Por ello, baste con decir que se basa en considerar términos adicionales que contribuyen a la iluminación mediante potencias de la forma $[\vec{L} \cdot \vec{N}(x, y, z)]^\alpha$, [Foley-96], [Hear-04] y [Hecht-01]. Con estas potencias la luz reflejada por la superficie especular no se distribuye de igual forma para todos los ángulos, como pasaba con la iluminación difusa, sino que se concentra en determinadas orientaciones, tal y como se ilustra en la Figura 2.3 c).

2.2.1 Analogía entre el modelo de iluminación elegido y los espacios vectoriales

Una imagen cualquiera, puede ser expresada en términos de una base vectorial cuyos elementos de esa base son imágenes. Esto puede expresarse con la siguiente ecuación:

$$f(x, y) = c_0 \cdot v_0(x, y) + c_1 \cdot v_1(x, y) + c_2 \cdot v_2(x, y) + c_3 \cdot v_3(x, y) + \dots \quad (2.5)$$

donde $f(x, y)$ es una imagen cualquiera bidimensional, los $\{c_i\}_{i=0,1,\dots}$ son coeficientes arbitrarios y determinados por la iluminación concreta que apliquemos y los $\{v_i\}_{i=0,1,\dots}$ son los elementos de la base cuyas propiedades matemáticas serán discutidas en un apartado posterior. Nótese que el número de elementos de la base dependerá de la base elegida.

La ecuación (2.4) está formada por cuatro imágenes y en cada una de ellas contribuyen factores que dependen de las características del material así como de la posición en la que se encuentre la fuente que ilumine el objeto. Se puede observar una analogía entre el resultado del modelo de iluminación elegido y la expresión (2.5). De esta forma, identificando los términos de la Ec. (2.5) con los de la Ec. (2.4), si suponemos que las reflectividades $k_a(x, y)$ y $k_l(x, y)$ son constantes para todo el objeto, lo cual constituye una suposición no demasiado restrictiva, tenemos que:

$$k_a(x, y) \propto v_0(x, y) = \begin{cases} 1 & \text{si } f(x, y) \neq 0 \\ 0 & \text{en el resto} \end{cases} \quad (2.6.1)$$

$$N_i(x, y) \propto \{v_i(x, y)\}_{i=1,2,3} \quad (2.6.2)$$

Así, la reflectividad ambiental es proporcional a $v_0(x, y)$ y este elemento de la base coincide con la silueta del objeto, es decir, la luz ambiental sólo se refleja en aquellas zonas de la imagen donde tenemos el objeto. La silueta del objeto se define como la ecuación (2.6.1). El resto de componentes cartesianas del vector normal a la superficie son proporcionales a los elementos de la base $v_1(x, y)$, $v_2(x, y)$ y $v_3(x, y)$.

Insistimos, la ecuación (2.4), se reescribirá en la forma de la ecuación (2.5) para resaltar esta analogía. La ecuación (2.5) es la expresión de una combinación lineal entre los diferentes elementos de la base $\{v_i(x, y)\}$, basta con dar al conjunto de imágenes estructura de espacio vectorial para las posibles versiones de un objeto iluminado sean un subespacio de éste. A ello dedicaremos el siguiente apartado, indicando las propiedades matemáticas de esos espacios vectoriales y las bases que los definen.

2.3 Espacios vectoriales e imágenes

Si consideramos dos imágenes bidimensionales caracterizadas por las funciones $f(x, y)$ y $g(x, y)$, el conjunto de todas las posibles funciones definidas en \mathbb{R}^2 tiene estructura de espacio vectorial si se define el producto escalar como sigue:

$$\langle f | g \rangle \equiv (f * g)(0, 0) = \int_{\mathbb{R}^2} f^*(x, y) g(x, y) dx dy \quad (2.7)$$

Nótese que el producto escalar entre dos funciones no es más que la correlación entre las mismas en el origen. Esta métrica es comúnmente utilizada en todo proceso de reconocimiento de imágenes. Además, es fácil comprobar que la ecuación (2.7) cumple toda la axiomática requerida para un producto escalar, [de Burgos-94], es decir la definición (2.7) cumple las siguientes propiedades:

$$\langle f | g \rangle = \langle g | f \rangle^* \quad (2.8.1)$$

$$\langle \alpha f_1 + \beta f_2 | g \rangle = \alpha^* \langle f_1 | g \rangle + \beta^* \langle f_2 | g \rangle \quad (2.8.2)$$

$$\langle f | f \rangle \geq 0 \quad (2.8.3)$$

$$\langle f | f \rangle = 0 \text{ sii } |f\rangle = 0 \quad (2.8.4)$$

Hacemos notar que con estas propiedades el conjunto de imágenes pasa a ser, en sentido estricto, un espacio de Hilbert, ya que el producto escalar es hermítico y no está solamente definido para funciones reales, sino también complejas y es de dimensión continua. Pero por sencillez en la redacción, y

abusando del lenguaje, en lo sucesivo hablaremos de espacios vectoriales y no de espacios de Hilbert.

El hecho de que el producto escalar sea igual a la correlación entre las dos funciones evaluadas en el origen es de gran importancia. Por un lado nos permite conectar con las ramas más clásicas del reconocimiento de objetos y por el otro, nos posibilita a usar una herramienta matemática muy poderosa: la FFT, transformada rápida de Fourier.

Tengamos en cuenta que la formulación anterior sólo es válida si trabajamos con imágenes analógicas, es decir, con distribuciones continuas de intensidad. Normalmente las imágenes con las que trabajaremos son digitales (discretas) y así el cálculo del producto escalar se lleva a cabo mediante FFT.

Con todo lo anterior, las imágenes pasan a comportarse como un espacio vectorial y todo el aparato matemático asociado a éstos pasa a estar a nuestra disposición, pudiendo obtenerse resultados de gran interés en nuestro trabajo. Dado que una imagen es equivalente a un vector y el conjunto de todas las imágenes de un mismo objeto, pero iluminado de diferente manera, tiene la estructura de un subespacio vectorial, el resultado que hemos obtenido es que hemos equiparado la detección con invariancia a iluminaciones de un objeto 3D a determinar la pertenencia de un vector a un subespacio determinado, formado por todas las imágenes que corresponden al objeto bajo todas las posibles iluminaciones.

2.4 Métodos vectoriales de reconocimiento de objetos

Ya hemos dicho previamente que el reconocimiento 3D con invariancia a iluminaciones encuentra su análogo 2D, en cierta medida, en las variaciones de contraste e intensidad en imágenes bidimensionales. Históricamente, uno de los primeros tratamientos para este problema lo dieron Dickey y Romero con su propuesta de filtros normalizados, [Dickey-91]. Esta propuesta se basa en usar ciertas propiedades matemáticas para obtener filtros tales que fueran invariantes a una variación en el objeto de la forma $a \cdot f(x, y)$, es decir, una variación de contraste. En dicho artículo, Dickey y Romero demostraban que era posible obtener un filtro equivalente al adaptado y normalizado, que cumplía con las condiciones exigidas. La expresión de este filtro es la siguiente:

$$FN_{g,f}(x,y) \equiv \frac{|(g * f)(x,y)|^2}{(|g|^2 * v_0)(x,y) \cdot \int_{\mathbb{R}^2} |f|^2 dx dy} \quad (2.9)$$

donde FN significa filtro normalizado, $v_0(x,y)$ sigue la notación anteriormente introducida y es la silueta, o soporte, del objeto $f(x,y)$ que queremos reconocer y $g(x,y)$ representa la escena de entrada.

La ecuación (2.9) se puede interpretar fácilmente en forma vectorial como veremos a continuación. Si la evaluamos en el origen tenemos de forma trivial que:

$$FN_{g,f}(0,0) \equiv \frac{|\langle f | g \rangle|^2}{\|g\|_{\text{soporte}}^2 \cdot \|f\|^2} \quad (2.10)$$

donde $\|f\|^2 = \int_{\mathbb{R}^2} |f|^2 dx dy = \langle f|f \rangle$ es la norma al cuadrado del vector $|f\rangle$ y

$\|g\|_{\text{Soporte}}^2$ está definido de la siguiente manera:

$$\|g\|_{\text{Soporte}}^2 = \int_{\mathbb{R}^2} |g(x, y)|^2 v_o(x, y) dx dy = \int_{v_o \neq 0} |g(x, y)|^2 dx dy \quad (2.11)$$

Es decir, $\|g\|_{\text{Soporte}}^2$ nos da la norma del vector $|g\rangle$ dentro de la región de definición, o silueta, del objeto. O en términos de intensidad, si $\|g\|^2$ nos indica cuál es la intensidad asociada a toda la distribución de amplitud $|g\rangle$, $\|g\|_{\text{Soporte}}^2$ proporciona la fracción de la intensidad confinada en la región definida por el soporte del objeto a detectar, $|v_o\rangle$.

Si en lugar de evaluar (2.9) en el origen de coordenadas, lo hacemos en un punto cualquiera $(x - x_0, y - y_0)$, tendremos expresiones del tipo:

$$\begin{aligned} (f * g)(x_0, y_0) &= \int_{\mathbb{R}^2} f^*(x - x_0, y - y_0) g(x, y) dx dy = \\ &= \langle f(x - x_0, y - y_0) | g \rangle \end{aligned} \quad (2.12)$$

La ecuación anterior es la expresión del producto escalar entre $|g\rangle$ y $|f\rangle$ centrada en el punto (x_0, y_0) . Es decir, en términos de espacios vectoriales la correlación es la operación que calcula el producto escalar entre dos vectores $|f\rangle$ y $|g\rangle$, uno de los cuales lo vamos centrando en cada punto del espacio.

Con esto, el resultado al aplicar la correlación ya no es un escalar evaluado en un solo punto sino una imagen bidimensional.

El resultado anterior es una consecuencia de las propiedades de linealidad e invariancia a traslaciones de la correlación, [Goodman-96], y será usado a lo largo de todo este trabajo. En lo que sigue, cada vez que calculemos la correlación entre dos imágenes, lo que estamos haciendo es calcular el producto escalar entre un vector y otro vector centrado en un determinado punto. Es decir, estamos calculando el producto escalar referido a un cierto sistema evaluado dentro de la región de definición del objeto. Términos similares a la Ec. (2.12) surgen de la relación entre las distintas operaciones de reconocimiento que definiremos en este capítulo y otras operaciones de filtraje en las que nos hemos basado para comparar los resultados.

De esta forma, el filtro normalizado lo que hace es calcular para cada punto del espacio el cociente dado por la ecuación (2.10). Una de las aplicaciones más inmediatas del filtro normalizado definido por Dickey *et al* [Dickey-91] es la detección con invariancia a intensidad o iluminación en imágenes 2D. Esto se cumplirá siempre que definamos un cambio de intensidad o iluminación como una multiplicación del objeto de referencia por una constante. Notemos que en esta situación ningún filtraje lineal, como el filtro adaptado, por ejemplo, podría detectar correctamente dos imágenes iguales pero con distinta iluminación, ya que la correlación lineal es lineal con la intensidad. Así, si encontramos que la imagen a detectar se puede expresar como, $g(x, y) = a \cdot f(x, y)$, es decir el objeto que buscamos con una variación de iluminación, la salida del filtro normalizado será igual a la unidad:

$$FN_{a,f,f}(0,0) \equiv \frac{|\langle f | a \cdot f \rangle|^2}{\|a \cdot f\|_{\text{Soporte}}^2 \cdot \|f\|^2} = 1 \quad (2.13)$$

En cualquier otro punto de la imagen en el cual no se encuentre una versión cambiada de intensidad de la forma definida con anterioridad, tendremos una respuesta menor que la unidad debido a la desigualdad de Cauchy–Schwarz, [Dickey-91].

Existe otra forma de interpretar el filtro normalizado, basada en espacios vectoriales. De acuerdo con ella, el filtro normalizado trata de detectar si un vector $|g\rangle$ es de la forma $a \cdot |f\rangle$ es decir, si pertenece al subespacio generado por el vector $|f\rangle$, siendo esto consistente con la decisión de diseño de Dickey y Romero de obtener invariancia a una alteración del tipo $a \cdot f(x, y)$.

Pero existen casos más complicados que sólo pueden ser modelados por subespacios vectoriales con una base más amplia, de más de un solo vector, lo que impide el uso del filtro normalizado en tales circunstancias y hace necesario el buscar otras alternativas. Por ejemplo, podemos definir el conjunto de todas las versiones de un mismo objeto $f(x, y)$ en términos de una base formada por dos elementos en vez de uno sólo:

$$f'(x, y) = a \cdot v_0(x, y) + b \cdot f(x, y) \quad (2.14)$$

con $v_0(x, y)$ la ventana de definición de $f(x, y)$. Según la expresión anterior, en $f'(x, y)$ contemplamos cambios de iluminación de tipo multiplicativo, por la constante b , así como de tipo aditivo, representado por la constante a .

De esta forma, si las constantes a y b pueden ser cualesquiera, la expresión (2.14) define un subespacio generado por los vectores $\{v_0(x, y), f(x, y)\}$.

Este caso se correspondería con problemas como una modificación de la iluminación de la imagen, [Lefebvre-02], en la que la transformación es de la forma $f'(x, y) = a + b \cdot f(x, y)$, o la adición de un patrón de camuflaje al objeto a reconocer [Arsenault-05].

Podemos definir casos que incluyan más elementos de la base, definiendo así subespacios de dimensión más alta. De esta forma el reconocimiento de una imagen que pertenezca a ese subespacio vendrá determinada por la detección de un vector $f(x, y)$ que está generado por un conjunto de imágenes $\{v_i(x, y)\}$ de más de dos elementos. Un posible problema que vendría descrito de esta forma sería la transformación afín de la intensidad del objeto, [Roy-04]. Una forma matemática de expresar lo mencionado con anterioridad es utilizando la siguiente ecuación:

$$f(x, y) = c_0 \cdot v_0(x, y) + c_1 \cdot v_1(x, y) + c_2 \cdot v_2(x, y) \dots \quad (2.15)$$

siendo $\{c_i\}$ las componentes del vector $|f\rangle$ en la base $\{v_i(x, y)\}$.

Dado que la Ec. (2.15) sólo describe una combinación lineal de vectores, el primer paso para definir una base matemática consiste en ortonormalizarla, ya que es necesario para poder operar con ella y poder aplicar las propiedades básicas del álgebra vectorial. Un método general utilizado para ortonormalizar bases es el método de Gramm-Schmidt, [Mathworld-06]. De

esta forma, cambiaremos la base definida por el conjunto $\{v_i(x, y)\}$ por otro conjunto cuyos elementos de la base sean ahora ortonormales. De esta forma:

$$f(x, y) = d_0 \cdot \hat{v}_0(x, y) + d_1 \cdot \hat{v}_1(x, y) + d_2 \cdot \hat{v}_2(x, y) \dots \quad (2.16)$$

donde $\{\hat{v}_i(x, y)\}$ es la base ortonormal del subespacio y $\{d_i\}$ las nuevas componentes de $f(x, y)$ en dicha base.

Nótese que esta ortonormalización es necesaria para cálculos posteriores, ya que en la mayoría de casos prácticos los $\{v_i(x, y)\}$ originales no son ortogonales. Por ejemplo, en el caso de la invariancia a contraste o iluminación definida según la ecuación (2.15), no se cumple que $f(x, y) \perp v_0(x, y)$.

Una vez ortonormalizado el conjunto generador del subespacio ya es posible operar con él de forma estándar. Ahora sólo falta encontrar una manera de detectar vectores de la forma (2.15), que en nuestro caso comprobaremos que va a ser equivalente a detectar objetos 3D con invariancia a iluminación.

Para el caso bidimensional, Arsenault *et al.*, [Lefebvre-02], [Lefebvre-03], [Roy-04] y [Arsenault-05], propusieron un método, el filtraje definido con las siglas LACIF que, aplicado a cambios de iluminaciones definidas con distintos modelos, sirvió para reconocer objetos 2D con invariancia a todos estos tipos de transformaciones de intensidad. Por su interés y la aplicación que le daremos a lo largo de esta memoria vamos a explicar con detalle este método.

2.4.1 El filtraje LACIF

En su forma original el LACIF, que significa filtro invariante adaptativo localmente a cambios de contraste (Locally Adaptive Contrast-Invariant Filter), fue desarrollado para obtener una respuesta invariante a cambios de contraste o iluminación de la forma definida en la ecuación (2.14) para imágenes 2D y se definió de la forma siguiente:

$$LACIF_{gf}(x, y) = \frac{(g * f_0)^2(x, y)}{\sqrt{N}(g^2 * \hat{v}_0)(x, y) - (g * \hat{v}_0)^2(x, y)} \quad (2.17)$$

donde $f(x, y)$ es el objeto que queremos reconocer, $g(x, y)$ es una escena arbitraria de entrada, N es el número de píxeles dentro del área de soporte, $\hat{v}_0(x, y)$ y $f_0(x, y)$ es la versión de media cero del objeto dentro del soporte en cuestión:

$$f_0(x, y) = f(x, y) - \mu_f v_0(x, y) \quad (2.18)$$

Con μ_f la media de $f(x, y)$ dentro del soporte.

Ahora el área de soporte de la nueva base, es la versión normalizada del soporte real del objeto $v_0(x, y)$. Es decir:

$$\hat{v}_0(x, y) = \frac{v_0(x, y)}{\sqrt{\langle v_0 | v_0 \rangle}} = \frac{v_0(x, y)}{\sqrt{N}} \quad (2.19)$$

Y se define también la versión normalizada de $f_0(x, y)$ de la siguiente forma:

$$\hat{f}_0(x, y) = \frac{f_0(x, y)}{\sqrt{\langle f_0 | f_0 \rangle}} \quad (2.20)$$

El filtraje LACIF permite discriminar si una escena dada pertenece o no al subespacio formado por la base $\{v_0(x, y), f(x, y)\}$; o por su equivalente ortonormalizado, el conjunto $\{\hat{v}_0(x, y), \hat{f}_0(x, y)\}$, que es el resultado de ortonormalizar el conjunto anterior, $\{v_0(x, y), f(x, y)\}$, por el método de Gramm-Schmidt.

El filtro descrito por la ecuación (2.17), en el caso de localizar en $g(x, y)$ algún fragmento de $f(x, y)$ que sea una combinación lineal de la base $\{v_0(x, y), f(x, y)\}$, ofrecerá un pico de respuesta igual a la unidad y una respuesta inferior en caso contrario.

Dado nuestro interés en identificar las operaciones de detección que vamos estudiando con los espacios vectoriales, diremos que, el LACIF es equivalente a proyectar la escena $g(x, y)$ en el subespacio ortogonal a $v_0(x, y)$, esto es, proyectar la versión $g_0(x, y)$, definida de forma análoga a la ec.(2.18) como $g(x, y) - \mu_f v_0(x, y)$, con media cero de la escena en la región de soporte y calcular el coseno del ángulo α entre el vector escena y el vector objeto a detectar, como ilustramos en la siguiente figura [Lefebvre-02]:

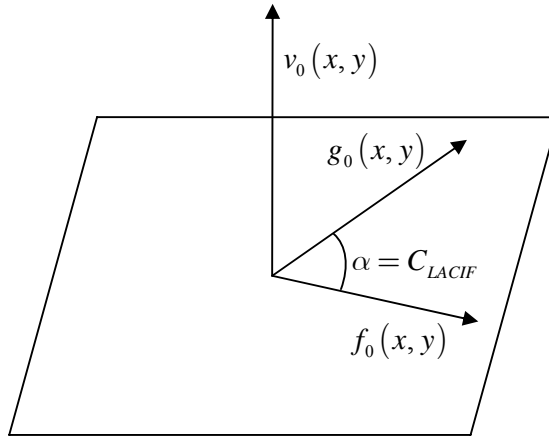


Figura 2.4 Interpretación vectorial del LACIF

Nótese que los elementos de la base $\{v_0(x, y), f_0(x, y)\}$ representados en la figura anterior son ortogonales, ya que el producto escalar entre un objeto de media cero dentro de un determinado soporte y ese soporte es cero. Lo que no se cumple en general es que esos vectores sean ortonormales. Con el fin de aclarar la interpretación vectorial y geométrica del filtraje LACIF hemos representado los vectores de la base ortogonales. Además, el filtraje LACIF se puede generalizar para detectar vectores dentro de un subespacio de dimensión más alta. En concreto, si la base es $\{\hat{v}_i(x, y)\}_{i=0,1,\dots}$, con $\hat{v}_0(x, y)$ la versión normalizada del soporte, el LACIF se puede expresar del modo siguiente:

$$LACIF_{g, \hat{v}_i}(x, y) = \frac{(g * \hat{v}_1)^2(x, y)}{\sqrt{N} (g^2 * \hat{v}_0)(x, y) - \sum_{i=1} (g * \hat{v}_i)^2(x, y)} \quad (2.21)$$

La expresión anterior ha sido utilizada por Roy *et al.* para reconocer objetos 2D con invariancia a transformaciones afines de intensidad y por Arsenault *et*

al. para el reconocimiento con invariancia a la adición de patrones de camuflaje en [Roy-04] y [Arsenault-05], respectivamente.

Sin embargo, el principal inconveniente del filtraje LACIF es que no usa toda la información relativa a la base como criterio de pertenencia al subespacio, sino que sólo usa una parte de ella. Este hecho hace que la capacidad de discriminación del sistema sea excesivamente restrictiva. Con el fin de solucionar este inconveniente planteamos un nuevo tipo de filtraje.

2.4.2 El filtraje LADC

En este capítulo proponemos una alternativa más robusta, [Vallés-06], para el reconocimiento invariante a iluminaciones basada en la misma idea de detección como la pertenencia de un vector, o imagen dada, a un subespacio definido por una base determinada, pero intentando usar el máximo posible de información asociada al subespacio.

Este nuevo método de reconocimiento está basado en usar como criterio de pertenencia a un subespacio el cálculo del coseno del ángulo entre la escena de entrada y todo el subespacio generado por los vectores base, a diferencia del método anterior basado en usar como criterio de pertenencia a un subespacio el coseno del ángulo entre una proyección en el subespacio de media cero y la escena de entrada. Es decir, donde el filtraje LACIF sólo usaba la parte referida al espacio de media cero, se pretende usar la totalidad de la información de la base. Visto de forma gráfica y limitando la figura a una base de dos dimensiones por simplicidad de interpretación:

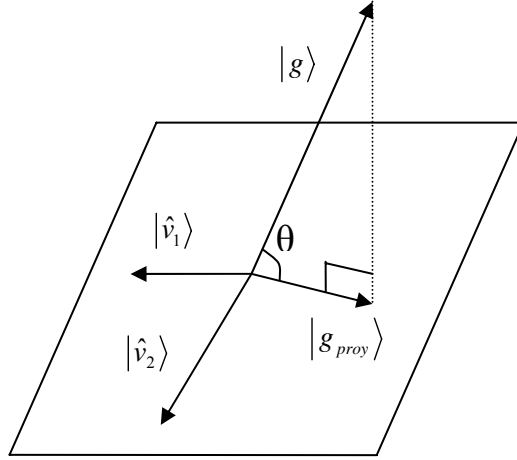


Figura 2.5 Interpretación geométrica del nuevo método propuesto

Hemos denominado a la operación de filtraje anterior como LADC¹, es decir correlación basada en una distancia angular calculada localmente. La definición consistirá en calcular el coseno del ángulo entre una escena $|g\rangle$ y un subespacio generado por $\{\hat{v}_i(x, y)\}_{i=0,1,\dots}$, con $\hat{v}_0(x, y)$ la ventana normalizada, como:

$$\cos^2(\theta) = \frac{\|g_{proj}\|^2}{\|g\|^2} \quad (2.22)$$

La ecuación anterior es una expresión totalmente general y válida para cualquier espacio vectorial; además $|g_{proj}\rangle$ es la proyección del vector $|g\rangle$ en el subespacio vectorial definido por $\{\hat{v}_i(x, y)\}_{i=0,1,\dots}$ que se calcula de la siguiente forma [Mathworld-02a]:

¹ Local Angular Distance Correlation, correlación por distancia angular local, en inglés.

$$|g_{proj}\rangle \equiv \sum_i \langle \hat{v}_i | g \rangle | \hat{v}_i \rangle \quad (2.23)$$

De la misma forma que para el filtraje LACIF y el filtro normalizado de Dickey, la operación definida en la ecuación (2.22) asegura que el valor de la respuesta esté definido entre 0 y 1. Si el vector a estudiar pertenece al subespacio tenemos que $|g_{proj}\rangle = |g\rangle$ y por tanto el coseno es igual a la unidad, siendo inferior en cualquier otro caso.

Estudiando la ecuación (2.22), vemos que necesitamos calcular la norma del vector proyección. Para expresar dicha norma en términos de correlaciones haremos uso de las mismas expresiones que en el apartado anterior. Por ello el numerador se calculará como:

$$\|g_{proj}(x, y)\|^2 = \sum_i |\langle \hat{v}_i | g \rangle|^2 = \sum_i |\hat{v}_i * g|^2(x, y) \quad (2.24)$$

Y de forma análoga el denominador:

$$\|g(x, y)\|^2 = \int_{\mathbb{R}^2} v_0(x, y) g^2(x, y) = \sqrt{N} (\hat{v}_0 * g^2)(x, y) \quad (2.25)$$

Nótese que las normas definidas anteriormente están evaluadas en el interior del contorno que define el objeto que queremos detectar. Además, N es el área en píxeles del soporte $v_0(x, y)$. Esto ya fue utilizado al calcular el filtro normalizado y en el filtraje LACIF consideramos sólo el modulo local de la escena, es decir, la integral de la intensidad en el entorno de (x, y) dentro del contorno.

Juntando (2.24) y (2.25) queda la expresión final del LADC como:

$$LADC_{g, \hat{v}_i}(x, y) \equiv [\cos^2 \theta]_{g, \hat{v}_i}(x, y) = \frac{\sum_i (g * \hat{v}_i)^2(x, y)}{\sqrt{N} (g^2 * \hat{v}_0)(y, y)} \quad (2.26)$$

La ecuación anterior presenta ciertas similitudes con el filtraje LACIF, pero también claras diferencias. El parecido más claro está en el uso de correlaciones en la implementación de ambos métodos, esto nos permite pensar tanto en una posible implementación óptica o en el uso de FFTs en una simulación digital.

Por otro lado y, como ya hemos mencionado anteriormente, la diferencia más clara está en el criterio de pertenencia al subespacio, [Vallés-06]. En el caso del filtraje LACIF este criterio viene determinado por el ángulo entre el vector objeto y su proyección sobre el espacio de media cero, con lo que sólo consideramos la proyección en una fracción del soporte. En el caso del LADC usamos la proyección de la escena de entrada con todo el subespacio. Esto hará a nuestro método más robusto para el reconocimiento 3D invariante a iluminaciones tal y como verificaremos en la presentación de los resultados obtenidos comparando ambos métodos.

2.5 Implementación y resultados

En el apartado 2.2 hemos visto que el problema del reconocimiento de objetos 3D con invariancia a iluminaciones equivale, obviando la influencia de las sombras, a determinar la pertenencia de un cierto vector a un subespacio determinado. En el apartado 2.3, hemos definido las expresiones

matemáticas necesarias para convertir el conjunto de imágenes que definen un objeto de referencia con cambios en la iluminación en un espacio vectorial y finalmente en el apartado 2.4 hemos revisado otros métodos de reconocimiento basados en este planteamiento vectorial e introducido nuevas propuestas para abordar el tema.

Ahora vamos a presentar cómo se han implementado ambos métodos y vamos a comparar los resultados obtenidos con el LADC y con el filtraje LACIF.

La implementación de ambos métodos ha sido llevada a cabo digitalmente con simulaciones realizadas con el software Matlab, versión 7.1 r14. Hemos escogido una implementación digital por su rapidez y bajo coste frente a otras opciones basadas en implementaciones ópticas para calcular las correlaciones.

En un primer paso vamos a buscar cuál es la base vectorial asociada a un objeto a reconocer determinado. Utilizaremos el objeto 3D “oso” mostrado de la figura siguiente como referencia:

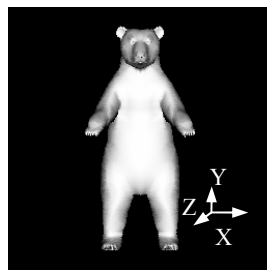


Figura 2.6 Objeto “oso” a reconocer iluminado de forma frontal y convención de ejes usada en este capítulo

Nótese que en la Figura 2.6 no sólo hemos mostrado el objeto 3D que pretendemos reconocer, además hemos fijado con qué punto de vista

trabajaremos. Por simplicidad, la orientación desde la cual observamos al objeto será la frontal.

Otro punto a tener en cuenta es averiguar cuál es la base del espacio vectorial. Para ello compararemos el modelo de iluminación propuesto en la Ec. (2.4) que genera todas las versiones iluminadas del objeto con los elementos de la base que presentamos en las ecuaciones (2.6.1) y (2.6.2).

Así, retomando la ecuación (2.4), los L_i serán las componentes del vector de iluminación que usemos en cada momento, la reflectividad $k_l(x, y)$ del objeto se considera constante, por lo que la base estará formada por el soporte del objeto $\nu_0(x, y)$ y las componentes del vector normal a la superficie, $N_i(x, y)$. Iluminando el objeto mostrado en la Figura 2.6 resultará en una combinación lineal de los $N_i(x, y)$ y el soporte $\nu_0(x, y)$. Esas imágenes que definen la base vectorial utilizada las mostramos en la siguiente figura:

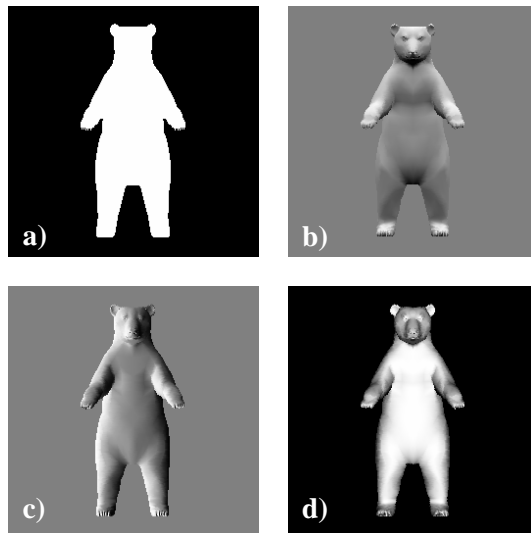


Figura 2.7 Base vectorial $\{\nu_0, N_i\}_{i=x,y,z}$, en a) se muestra el soporte del objeto y en b), c) y d), las componentes de las normales, N_x , N_y y N_z , respectivamente.

En la siguiente figura presentamos tres versiones del objeto a reconocer cuando es iluminado de diferentes formas:



Figura 2.8 Imagen formada por la composición de tres versiones del objeto sometido a las iluminaciones $[0.2, 0.9, 0.8]$, $[0.75, 0.2, 0.5]$ y $[0, -0.1, 1.0]$

Si aplicamos el LADC y el filtraje LACIF a la imagen anterior obtenemos el siguiente resultado. Con el fin de facilitar la visualización del valor del pico de correlación, presentamos un perfil horizontal de los planos de salida:

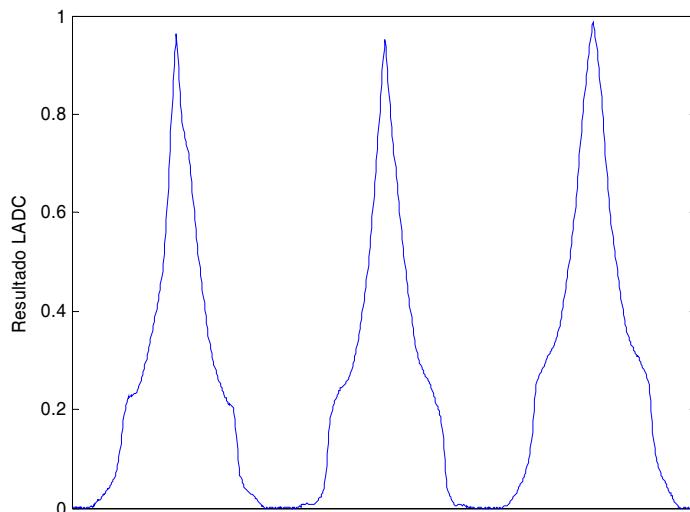


Figura 2.9 Perfil en el eje horizontal del resultado obtenido con el LADC

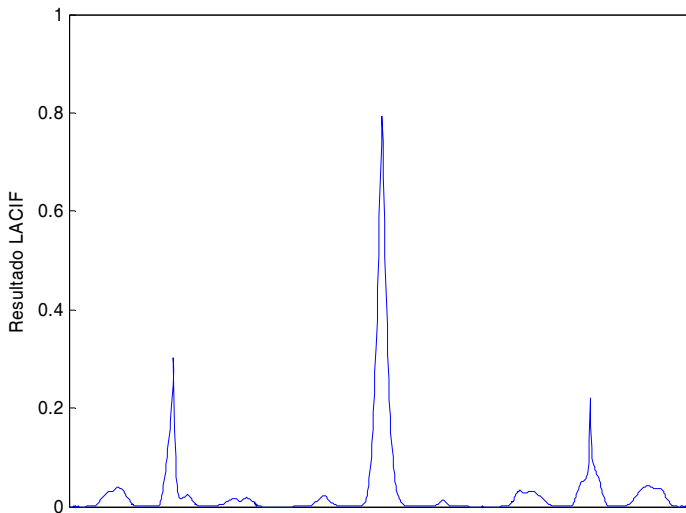


Figura 2.10 Perfil en el eje horizontal del resultado obtenido con el filtraje LACIF

Las dos figuras anteriores muestran que los resultados utilizando el LADC son netamente superiores ya que es capaz de detectar las tres versiones del objeto iluminado con unos picos de correlación próximos a la unidad. Por otra parte el resultado con el filtraje LACIF fracasa en el reconocimiento con invariancia a iluminaciones. Únicamente el objeto central es detectado por encima del 0.5 de valor de correlación. Hacemos notar que ambos métodos detectan los objetos independientemente de su posición en el espacio, es decir, son invariantes frente a traslaciones. Esto es así debido a las propiedades de linealidad de las correlaciones en las que se expresan los dos métodos de reconocimiento.

Explicaremos a continuación por qué el filtraje LACIF falla al detectar objetos que son iluminados de forma diferente. La razón la encontramos en que la base $\{v_0, N_i\}_{i=X,Y,Z}$ no está completa ya que falta la influencia de las sombras y, por tanto, no genera al conjunto de todas las posibles versiones iluminadas del objeto. Gráficamente se puede comprobar si comparamos una

imagen dada y su proyección en la base determinada por todas las iluminaciones en la siguiente figura:

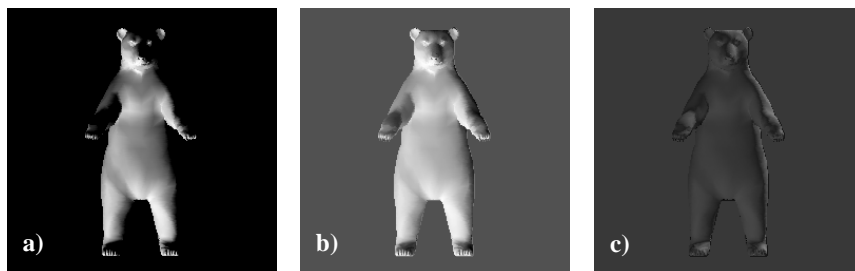


Figura 2.11 En a) mostramos un objeto, en b) tenemos la proyección de la imagen a) en la base $\{v_0, N_i\}_{i=x,y,z}$, sin contar con la influencia de las sombras y en c) mostramos la diferencia entre a) y b)

Como se ve en la Figura 2.11, la base $\{v_0, N_i\}_{i=x,y,z}$ es incapaz de reproducir de manera completa los objetos iluminados, es una buena aproximación, pero incompleta. El hecho de que el LADC presente mejores resultados que el filtraje LACIF se debe a su robustez. Para explicarlo, recordemos que en la definición del LADC se proyecta la escena de entrada sobre el subespacio, mientras que el filtraje LACIF sólo hace una proyección de la escena de entrada sobre el soporte de media cero. De esta forma el LADC usa una proporción mayor de la información de la base, lo que lo hace más tolerante que el filtraje LACIF a entornos donde falta información.

Para ilustrar hasta qué punto es robusto el LADC vamos a obtener un conjunto de resultados que consistirán en ir variando la iluminación del objeto 3D progresivamente e iremos calculando el valor de pico de correlación obtenido por los dos métodos que estamos comparando. Para generar todas las posibles iluminaciones, tomaremos un objeto 3D en una orientación dada e iremos desplazando una fuente de iluminación frente a él. La siguiente figura ilustra cómo hemos caracterizado la posición de la fuente.

Vemos que la posición de la fuente está determinada por las coordenadas esféricas θ y φ .

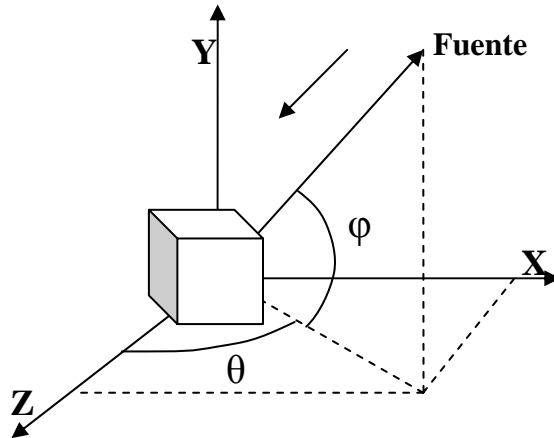
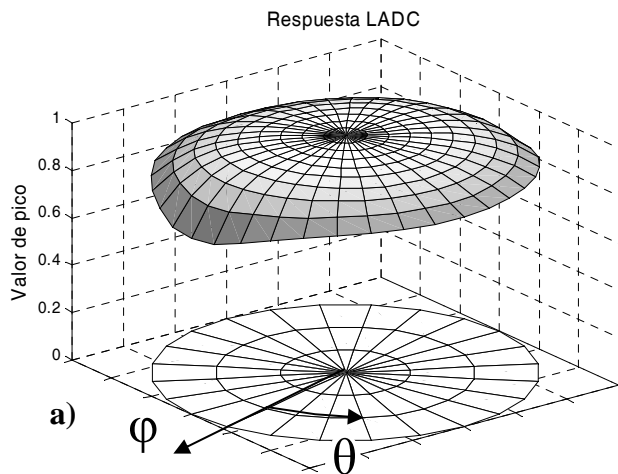
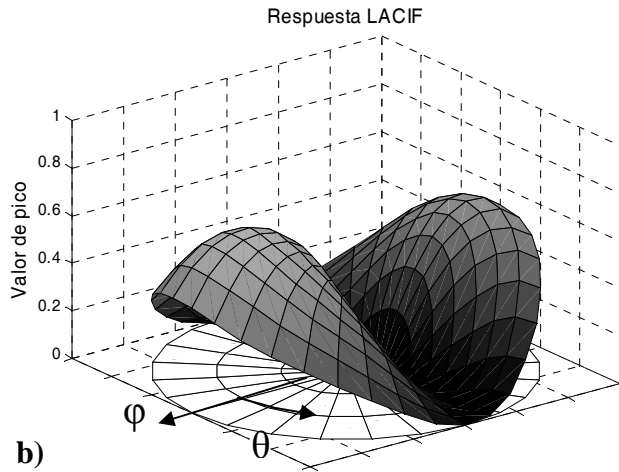


Figura 2.12 Posición de la fuente en función de las coordenadas esféricas.

En la siguiente figura comparamos los resultados obtenidos de aplicar el LADC y el filtraje LACIF al reconocimiento del objeto correcto, para un muestreo completo caracterizado según el criterio de ángulos de la Figura 2.12, con θ variando de 0 a 2π y φ entre 0 y $\pi/2$:





**Figura 2.13 En a) mostramos los resultados del LADC,
en b) los correspondientes al LACIF**

Los resultados del LADC permanecen prácticamente constantes y próximos a la unidad, variando entre un valor mínimo de 0.74 y un máximo de 1.00 con una media de 0.96, indicándonos que reconocemos el objeto con independencia de la posición de la fuente, es decir con invariancia a cualquier iluminación. En cambio el filtraje LACIF ofrece unos resultados mucho más pobres, ofreciendo una media de 0.26 con un valor mínimo de 0.00 y un máximo de 0.68, fallando la detección para la mayoría de iluminaciones. Otra explicación a tal falla puede deberse también a la elección de la base que define el objeto. Como hemos dicho anteriormente la elección de la base es crucial ya que todas estas técnicas de detección están basadas en calcular el ángulo que forma un determinado vector a un subespacio definido por una base determinada o calcular el ángulo que forma una proyección de ese vector dado sobre un determinado elemento de la base. En el siguiente apartado estudiaremos la influencia de otras posibles bases, diferentes a las definidas por el propio modelo de iluminación.

2.5.1 Transformación de Karhunen-Loeve

El espacio formado por funciones continuas en \mathbb{R}^2 tiene como base canónica el conjunto $\{\delta(x-x', y-y'), \forall x', y' \in \mathbb{R}^2\}$, de dimensión infinita y continua.

Como en la práctica trabajaremos con imágenes digitales de $N_1 \times N_2$ píxeles, la dimensión del espacio vectorial en un caso típico donde $N_1 = N_2 = 256$ será $256 \times 256 = 65536$. Es decir, el espacio vectorial completo de todas las posibles imágenes es de dimensionalidad altísima y difícilmente manejable, por lo que se hace necesario restringir nuestro problema de alguna manera.

En nuestro caso, según el modelo de iluminación expresado por la ecuación (2.4) dedujimos que el conjunto de cuatro elementos $\{v_0, N_i\}_{i=x,y,z}$ ha de ser razonablemente parecido a la base real del subespacio con el que queremos trabajar. Debido a ello, en primera instancia, se consigue reducir un problema cuya dimensionalidad teórica es 65536 a otro de 4. Esto es una primera aproximación ya que esta reducción sólo es exacta si no tenemos en cuenta la influencia de las sombras en el modelo. Al tener en cuenta la influencia de las sombras nos encontramos con una situación en la que la dimensión teóricamente es muy alta, pero en la práctica se comporta como un subespacio de baja dimensión. Es decir, tenemos un subespacio vectorial en el que no todos los elementos de su base son igualmente importantes. Este es el punto de partida de una técnica estándar del análisis de datos, la transformación de componentes principales o KLT¹. Antes de comenzar con la descripción de la KLT, conviene aclarar que no es un método que reduzca realmente la dimensionalidad de un problema, ya que tal cosa es imposible, la dimensionalidad de un problema es intrínseca al mismo. Su objetivo es, dado un problema de dimensionalidad alta, escoger aquellas componentes

¹ Transformación de Karhunen-Loeve, en inglés.

vectoriales de mayor importancia, consiguiendo tratar, de forma aproximada, el problema como uno de dimensión más reducida, con las consiguientes ventajas.

Por tanto, la KLT es una transformación matemática tal que, dado un conjunto de vectores $|a_i\rangle$, con $i=1,\dots,n$, nos permite encontrar sus componentes vectoriales y la importancia relativa de éstas. Para ello se comienza construyendo la matriz de dispersión asociada al conjunto $|a_i\rangle$:

$$\psi_{i,j} = \sum_i |a_i - \bar{a}\rangle \langle a_i - \bar{a}| \quad (2.27)$$

donde $|\bar{a}\rangle = \frac{\sum_i |a_i\rangle}{n}$ es el vector promedio de los $|a_i\rangle$.

Una vez hallada esta matriz, la base vectorial del espacio definido por los $|a_i\rangle$ es el conjunto de los autovectores de $\psi_{i,j}$, viniendo expresada su importancia relativa por el autovalor asociado a cada uno, [Karhunen-47] y [Loeve-78].

La KLT es una técnica estándar que ya había sido usada con anterioridad en el reconocimiento de objetos, muy particularmente en el problema de reconocimiento de caras. Existe una línea de trabajo dentro de ese campo, en la cual se usa el concepto de *eigenface*, o autocara. Las eigenfaces son el producto de aplicar la KLT a un conjunto de caras, dentro de las cuales se pretende obtener un reconocimiento con respuesta invariante a cambios de iluminación. Varios ejemplos sobre esta posibilidad lo constituyen las propuestas de Belhumeur *et al.*, [Belhumeur-97], que consigue invariancia a iluminaciones y cambios de pose, de Turk *et al.* y Cendrillon, [Turk-91a],

[Turk-91b], y [Cendrillon-00], con unos métodos similares al de Belhumeur pero enfocado a aplicaciones en tiempo real. Al aplicar la KLT a nuestro método de reconocimiento con invariancia a iluminaciones, el beneficio que obtenemos es que nos va a permitir encontrar de forma óptima la base del subespacio de imágenes iluminadas.

Es de particular interés el mencionar la propuesta de Fortuna y Capson, [Fortuna-04], en la que se obtiene un reconocimiento de caras con invariancia a iluminaciones basado en el uso del Análisis de Componentes Principales, ICA¹, una técnica propia del campo de redes neuronales, [Hyvärinen-01], que posee cierto parecido con la KLT.

Una vez calculada la transformación KLT, esperamos que los primeros elementos sean similares a los $\{v_0, N_i\}_{i=X,Y,Z}$ ya que según nuestro modelo de iluminación, cualquier versión del objeto iluminado se obtiene como combinación lineal de los $\{v_0, N_i\}_{i=X,Y,Z}$, exceptuando la influencia de las sombras. Esto equivale a decir que el conjunto de todas las versiones del objeto iluminado de diferentes formas es un espacio vectorial de alta dimensionalidad, pero cuyas componentes principales son los vectores $\{v_0, N_i\}_{i=X,Y,Z}$. Además, los resultados obtenidos por el LADC usando esta base corroboran tal hipótesis. Es por todo ello esperable que los primeros autovectores obtenidos por la KLT y los $\{v_0, N_i\}_{i=X,Y,Z}$ tengan un gran parecido, como de hecho ocurre y se muestra en la figura siguiente:

¹ Independent Component Analysis, del inglés análisis de componentes independientes.

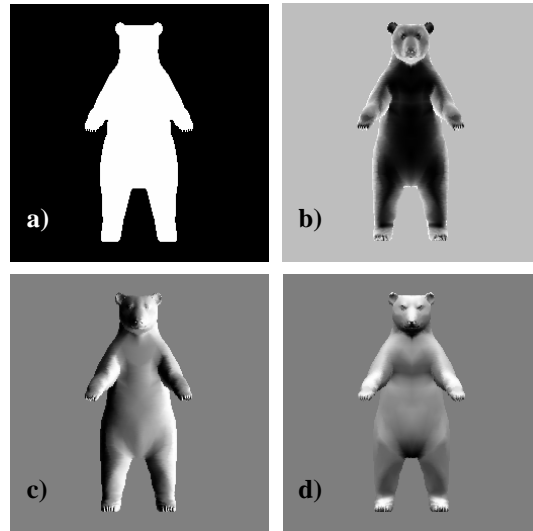


Figura 2.14 En a), b) y c), mostramos los cuatro primeros autovectores dados por KLT, ordenados según su autovalor

Nótese el parecido entre esta figura y la Figura 2.7. El primer autovector es el soporte del objeto y el resto son muy similares a los N_i salvo en el orden de los elementos y en el hecho de que la imagen b) de la figura anterior es la d) en la Figura 2.7 multiplicado por un número negativo. Éste es un detalle matemático que no afecta a la base.

Tal y como esperábamos, la suposición inicial de que la base del problema es prácticamente igual a $\{v_0, N_i\}_{i=X,Y,Z}$, se ha visto demostrada matemáticamente mediante la aplicación de la KLT. Dada su importancia, insistimos en este concepto. Dentro de nuestro planteamiento matemático, el problema vendrá determinado por una cierta base real, pero que no nos es posible conocerla, dada la gran dimensionalidad del problema.

Por otro lado, cuando usamos el filtraje LACIF o LADC, necesitamos definir una base para operar con ellos. Esta base no es, en principio la real del problema, es una base aproximada y cuanto mejor se escoja, en el sentido de

semejanza con la base real, mejores resultados se obtendrá con el LACIF o LADC. Tanto con el modelo de iluminación definido, como con la KLT, se tiene que una buena aproximación a la base real del problema la constituye el conjunto $\{v_0, N_i\}_{i=X,Y,Z}$.

Además, la KLT no sólo nos ha permitido corroborar el resultado anterior, es decir que los $\{v_0, N_i\}_{i=X,Y,Z}$ son una buena base aproximada. Esta transformación nos asegura, que la base que mejores resultados dará al aplicar el LADC, será la formada por los autovectores de $\psi_{i,j}$; la KLT nos permite seleccionar la base para operar con el LADC de forma óptima. Si aplicamos el LADC con la nueva base de autovectores, los resultados son muy similares a los mostrados en la Figura 2.13 aunque ligeramente superiores, estando entre un valor mínimo de 0.75 y un máximo de 1.00 con una media de 0.97, como se muestra en la siguiente figura:

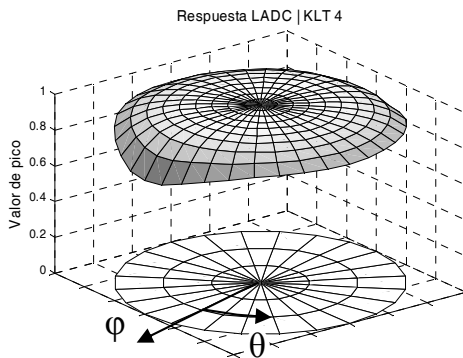


Figura 2.15 Resultados del LADC, con la base formada por los cuatro primeros autovectores de la KLT

Estos resultados son mejorables mediante la adición de sucesivos autovectores de $\psi_{i,j}$ a la base, a costa de un aumento en el tiempo de cálculo.

Si observamos en la siguiente figura la evolución de los autovalores para los

cien primeros autovectores vemos que éstos son muy inferiores a partir del cuarto elemento, de acuerdo con lo esperado con el modelo de iluminación, y pasan a ser prácticamente despreciables a partir del décimo autovector.

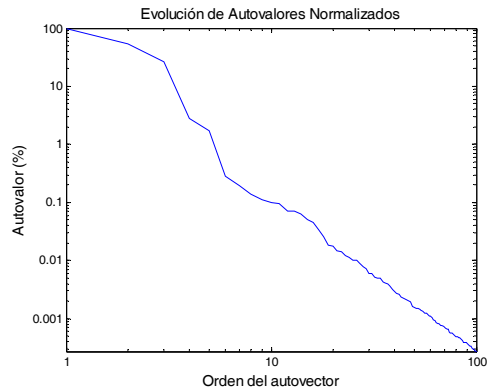


Figura 2.16 Evolución de los autovalores asociados a los primeros autovectores

Con esto la contribución a la base vectorial es casi nula a partir del décimo elemento. De hecho si se muestran los resultados obtenidos con el LADC usando la base formada por diez elementos obtenemos lo siguiente:

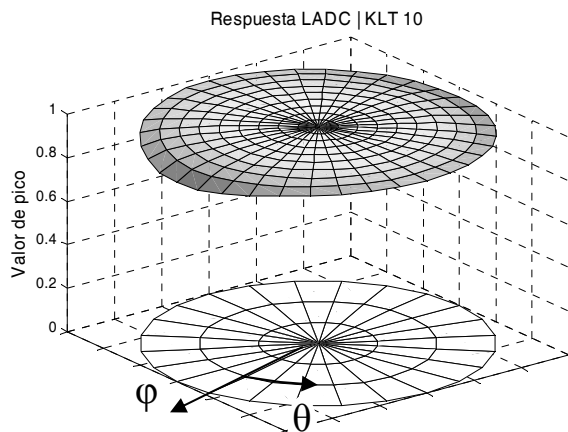


Figura 2.17 Resultados del LADC, con la base formada por los diez primeros autovectores de la KLT

Vemos que el valor del pico de correlación para el LADC es prácticamente plana y próximo a la unidad, con valor mínimo de 0.91 y una media de 0.99. La ligera bajada en la respuesta se debe a que la contribución del resto de elementos de la base es muy pequeña pero no idénticamente nula y sería corregible si considerásemos más elementos.

2.5.2 Comportamiento del filtraje LADC frente a ruido y discriminación

Otro factor de importancia al considerar un método de reconocimiento de objetos es su comportamiento frente al ruido. Se entiende como ruido aquella parte de una señal que es debida a fluctuaciones aleatorias y que no aporta información útil [Réfrégier-04] y [Van Etten-05].

Existen diferentes tipos de ruidos, según sea el criterio con el que se clasifique, éste es un tema extenso y aunque tratarlo en detalle está fuera del ámbito de este trabajo, conviene aclarar algunos puntos básicos. Así, se puede tener un ruido solapado o de fondo no solapado, según se superponga o no el ruido a la señal original. También se habla de ruidos aditivos o multiplicativos, según se tenga que el ruido se agregue como una suma o multiplicación a la señal original. Por último, es de gran importancia averiguar qué ley de distribución aleatoria sigue el ruido, [Réfrégier-04] y [Van Etten-05].

En nuestro caso, estudiaremos la influencia sobre la imagen de entrada de un ruido gaussiano de fondo no solapado, por ser este tipo de ruido bastante habitual en procesado de imágenes. Al ser de fondo no solapado, el ruido estará repartido por la zona exterior al soporte del objeto y seguirá una ley de distribución gaussiana. Vamos a evaluar la capacidad de discriminación del LADC en la presencia de éste tipo de ruido. En lo que sigue, los cálculos han

sido realizados usando la base generada por los 4 primeros autovectores de la KLT, prácticamente equivalente a $\{v_0, N_i\}_{i=X,Y,Z}$.

En la Figura 2.18 mostramos una escena compuesta por el objeto a reconocer en presencia del ruido gaussiano de fondo no solapado:

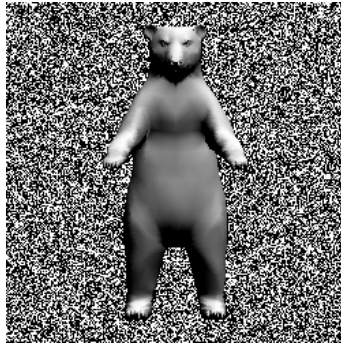


Figura 2.18 Objeto en presencia de ruido gaussiano (SNR = 10) de fondo no solapado

Para definir la calidad del proceso de reconocimiento en presencia de ruido se pueden emplear diferentes parámetros de calidad [Kumar-90], de los que vamos a utilizar la PCE¹ y SNR².

La PCE es un parámetro de calidad que, como su propio nombre indica, mide el cociente entre la energía del pico del resultado, respecto a la energía del plano en su totalidad, [Kumar-90]. Supongamos que tenemos una imagen $f'(x, y)$, definida como sigue:

$$f'(x, y) \equiv f(x, y) + R(x, y) \quad (2.28)$$

¹ Peak to Correlation Energy, relación entre la energía del pico y la total de la energía.

² Signal to Noise Ratio, en inglés.

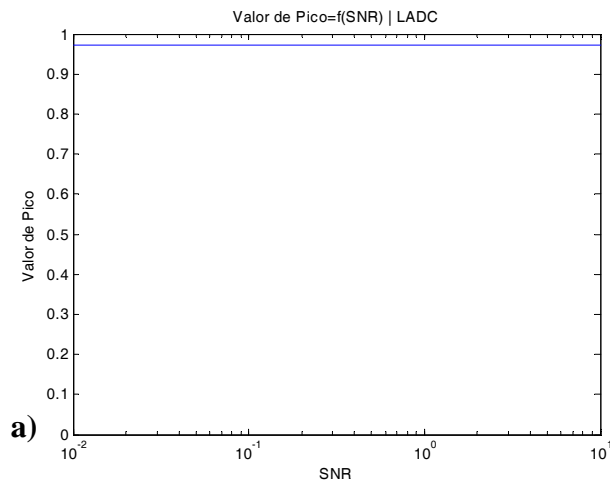
donde $f(x, y)$ es una imagen antes de añadir el ruido y $R(x, y)$ es la distribución de ruido considerada, en este caso gaussiano no solapado.

Entonces la PCE se calcula de la forma siguiente:

$$PCE = \frac{[LADC_{f, \hat{v}_i}(0, 0)]^2}{\int_{\mathbb{R}^2} [LADC_{f, \hat{v}_i}(x, y)]^2 dx dy} \quad (2.29)$$

Este parámetro proporciona información acerca de la concentración de intensidad en el plano de salida. También caracteriza la forma del pico de salida; tanto la altura, como la estrechez vienen determinadas por la PCE. Un valor alto de este parámetro se traduce en un pico muy puntiagudo.

En la Figura 2.19 mostramos cuál es la evolución del valor de pico y de la PCE, del valor del pico de correlación del LADC al variar la SNR, cuando queremos reconocer el objeto de la Figura 2.6:



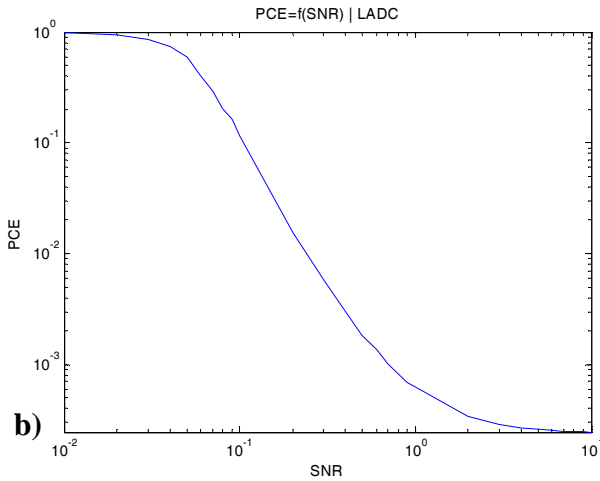


Figura 2.19 En a) mostramos la evolución del pico de correlación obtenido con el LADC frente a una variación de ruido, en b) mostramos la variación de la PCE

De la Figura 2.19 a) se observa que el valor de pico es invariante frente a la adición de este tipo de ruido. Si recordamos que en la ecuación (2.25) hemos dicho que el LADC sólo calcula la intensidad de la escena en una zona igual al soporte del objeto, es lógico que el valor del pico de correlación no varíe por la adición de ruido de fondo no solapado.

La evolución de la PCE con el aumento del ruido utilizando el LADC es más difícil de explicar. Tenemos que $R(x, y)$ la distribución de ruido gaussiano, de media cero y desviación σ , que no se solapa con la imagen, se puede escribir como sigue:

$$R(x, y) = \begin{cases} 0 & \text{si } f(x, y) \neq 0 \\ \text{Ruido Gaussiano} & \end{cases} \quad (2.30)$$

Si calculamos la respuesta del LADC para $f'(x, y)$ y calculamos explícitamente la PCE correspondiente en función de σ , obtenemos lo siguiente:

$$PCE(\sigma) = \frac{[LADC_{f', \hat{v}_i}(0, 0)]^2}{\int_{\mathbb{R}^2} [LADC_{f', \hat{v}_i}(x, y)]^2 dx dy} \quad (2.31)$$

Como el ruido $R(x, y)$ es no superpuesto a la señal podemos descomponer el denominador en dos términos:

$$PCE(\sigma) = \frac{[LADC_{f, \hat{v}_i}(0, 0)]^2}{[LADC_{f, \hat{v}_i}(0, 0)]^2 + \int_{(x, y) \neq (0, 0)} [LADC_{f', \hat{v}_i}(x, y)]^2 dx dy} \quad (2.32)$$

en donde el segundo término del denominador es igual a:

$$LADC_{f', \hat{v}_i}(x, y) = \frac{\sum_i \left[\int_{\mathbb{R}^2} \hat{v}_i^*(x' - x, y' - y) [f(x, y) + R(x, y)] dx' dy' \right]^2}{\sqrt{N} \int_{\mathbb{R}^2} \hat{v}_0^*(x' - x, y' - y) [f^2(x', y') + f \cdot R(x', y') + R^2(x', y')] dx' dy'} \quad (2.33)$$

Usando el hecho de que el ruido es gaussiano con un ancho σ , se puede reescribir (2.32) de la forma siguiente:

$$LADC_{f', \hat{v}_i}(x, y) = \frac{K}{\sqrt{N} \int_{\mathbb{R}^2} \hat{v}_0^*(x' - x, y' - y) [R^2(x', y')] dx' dy'} \propto K' \sigma^{-2} \quad (2.34)$$

donde K y K' son constantes.

Con las ecuaciones (2.32) y (2.34) podemos concluir que la PCE varía con el ruido de la forma siguiente:

$$PCE(\sigma) = \frac{[LADC_{f,\hat{v}_i}(0,0)]^2}{[LADC_{f,\hat{v}_i}(0,0)]^2 + O(\sigma^{-4})} \quad (2.35)$$

donde $O(\sigma^{-4})$ denota una dependencia funcional respecto a σ en órdenes potencias inferiores a -4.

Si la intensidad del ruido aumenta, entonces $\sigma \rightarrow \infty$, o, equivalentemente $SNR \rightarrow 0$, con esto la PCE tiende a la unidad. Así, toda la respuesta del LADC pasa a concentrarse en el pico de correlación que, además, no varía con el ruido. Este es un comportamiento sorprendente y muy interesante frente al ruido.

En cuanto a la capacidad de discriminación del LADC cuando en la escena de entrada se encuentran otro tipo de objetos distintos al de referencia se muestra en la siguiente figura. Consiste en una escena con ruido y compuesta por cuatro objetos, dos de ellos correctos y dos falsos. El resultado de aplicar el LADC a la escena de entrada se muestra a continuación:

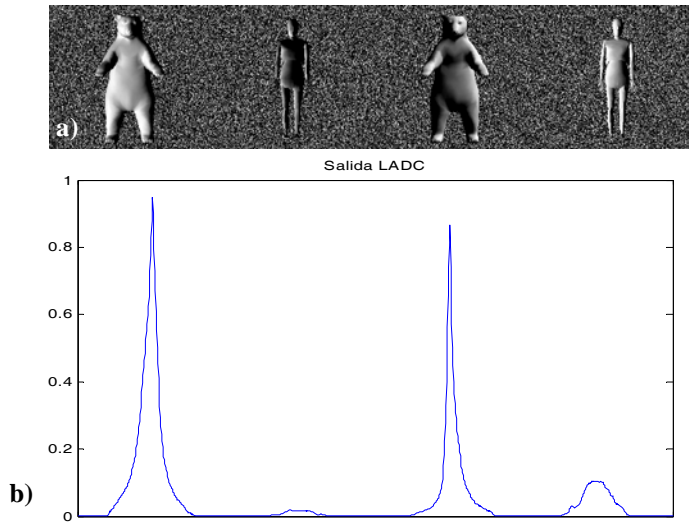


Figura 2.20 En a) tenemos una escena compuesta por varios objetos en presencia de ruido ($SNR = 1$), en b) mostramos el perfil del eje horizontal de la salida del LADC

Con el LADC se obtienen valores muy próximos a la unidad para el objeto a reconocer y otros muy inferiores, casi cercanos a cero, para los objetos falsos. Se comprueba que el LADC es capaz de discriminar el objeto correcto aun en presencia del ruido.

2.6 Conclusiones

En este capítulo hemos visto cómo es posible definir una estructura de espacio vectorial sobre el conjunto de imágenes para modelar cambios en la iluminación. Aplicado al reconocimiento de objetos 3D con invariancia a iluminaciones, este planteamiento vectorial reduce dicho problema, de forma aproximada, a la detección de vectores en un subespacio de dimensionalidad baja.

Aprovechando este hecho, se han descrito dos métodos, el filtraje LADC propuesto en este trabajo y el filtraje LACIF, para resolver este problema. Además se ha utilizado la transformación de Karhunen-Loeve para ajustar el número de elementos de la base, y con ello el tiempo de cálculo, frente al rendimiento del método propuesto, de forma matemáticamente óptima.

Hemos aplicado el LADC a una serie de situaciones en las que hemos demostrado su capacidad de reconocimiento frente a todas las posibles iluminaciones así como pruebas de discriminación y estabilidad frente a ruido.

Para concluir, se menciona que los resultados relativos a las capacidades de reconocimiento y discriminación de objetos del LADC y su comparativa con el filtraje LACIF han sido publicados en [Vallés-06].

En capítulos posteriores, 4 y 5, veremos cómo algunas de estas ideas referentes a la conexión entre subespacios vectoriales e imágenes, volverán a ser usadas en aplicaciones diferentes.

Capítulo 3

Estimación de rotaciones de objetos tridimensionales

3.1 Introducción

En este capítulo desarrollaremos un método que servirá tanto para el reconocimiento tridimensional invariante a rotaciones como para la estimación de la rotación sufrida por un objeto tridimensional. El método está basado en una codificación de la información tridimensional del objeto de entrada utilizando la PhFT¹. La información tridimensional de los objetos viene dada en forma de imágenes de rango.

El procedimiento de cálculo se inicia calculando las PhFTs de las distintas imágenes de rango obtenidas a partir de la rotación sucesiva del objeto tridimensional de entrada. Debido a las propiedades de esta codificación, el conjunto de PhFTs obtenidas definirá un mapa al que hemos llamado 3DOOM², que contiene toda la información geométrica del objeto de partida. Una vez codificado el objeto, se ha definido una operación matemática en el espacio 3D equivalente a la correlación en \mathbb{R}^2 , que permite reconocer el objeto tridimensional y estimar la rotación a la que ha sido sometido.

Este esquema de trabajo es sustancialmente diferente a los métodos clásicos de reconocimiento invariante a rotaciones tanto para el caso bidimensional como en el tridimensional. En dos dimensiones la opción clásica es el

¹ Phase Fourier Transform, transformada de Fourier de fase, en inglés.

² 3 Dimensional Object Orientation Map, en inglés mapa de orientaciones de un objeto 3D.

reconocimiento basado en el desarrollo en armónicos circulares de las funciones bidimensionales a reconocer. Introducidos por Hansen *et al.*, [Hansen-81a, Hansen-81b, Hansen-82], los armónicos circulares se definen como el conjunto de funciones del tipo $\{\exp(im\theta)\}$, con $m \in \mathbb{Z}$ y θ es la variable rotación. Este conjunto de funciones es ortonormal ya que se cumple:

$$\frac{1}{2\pi} \int_0^{2\pi} \exp(im\theta) \exp(in\theta) d\theta = \delta_{m,n} \quad (3.1)$$

donde $\delta_{m,n} \equiv \begin{cases} 1, & \text{si } m = n \\ 0, & \text{si } m \neq n \end{cases}$ es la delta de Kronecker.

Al ser los armónicos circulares un conjunto ortonormal, cualquier función $f(x, y) \in \mathbb{R}^2$ se puede desarrollar en infinitos términos usando como nueva base este conjunto:

$$f(x, y) = f(r, \theta) = \sum_{m=-\infty}^{\infty} f_m(r) \exp(im\theta) \quad (3.2)$$

con:

$$f_m(r) = \frac{1}{2\pi} \int_0^{2\pi} f(r, \theta) \exp(-im\theta) d\theta \quad (3.3)$$

Tal y como mostraron Hsu *et al.*, [Hsu-82a, Hsu-82b], usando este desarrollo es posible conseguir invariancia a rotaciones en el caso 2D; basta con usar un filtro adaptado a uno sólo de los términos del desarrollo $f_m(r) \exp(im\theta)$, llamado comúnmente componente de armónico circular de orden m . Si

consideramos una imagen de entrada igual al objeto de referencia $f(x, y)$ pero girada un cierto ángulo α , tenemos:

$$f'(r, \theta) = f(r, \theta + \alpha) = \sum_{m=-\infty}^{\infty} f_m(r) \exp(im\theta) \exp(im\alpha) \quad (3.4)$$

Podemos calcular la correlación en el origen entre f' y una de las componentes del desarrollo en armónicos circulares como:

$$(f' * [f_m \exp(im\theta)]) = 2\pi \exp(-im\alpha) \int_0^{\infty} |f_m(r)|^2 r dr \quad (3.5)$$

Es decir, si consideramos un filtro adaptado a un término del desarrollo en armónicos circulares, sólo varía la fase siendo invariante la intensidad de la salida, que es lo que habitualmente se puede medir en laboratorio utilizando un detector convencional. De esta forma se consigue invariancia a rotaciones para objetos 2D.

Este método de reconocimiento basado en la correlación se puede interpretar en términos de espacios vectoriales: los armónicos circulares constituyen una base del espacio, tal que una rotación del vector original no cambia la amplitud de las componentes del vector, sólo su fase. Esto equivale a convertir la invariancia a rotaciones en determinar si dos vectores tienen una componente del desarrollo en términos de los armónicos circulares con la misma amplitud.

El método anterior, no obstante, tiene algunos problemas. El primero es que sólo se trabaja con una fracción de la información del objeto $f(r, \theta)$, lo que acarrea una serie de consecuencias. Por la misma razón que un vector no

tiene ninguna componente mayor que su módulo, la energía de la función a reconocer se reparte entre todos los términos del desarrollo [Arsenault-89]. Por ello siempre es inferior la intensidad dada por la ecuación (3.5) que la que se obtendría usando un filtro adaptado clásico [VanderLugt-64].

Otra consecuencia es que este método implica el escoger unos ejes de coordenadas fijos para poder hacer el desarrollo de la ecuación (3.2). Si se escogen otros sistemas de coordenadas, los armónicos circulares están centrados en posiciones distintas produciendo resultados diferentes. Por ello la elección de un centro apropiado para el desarrollo en la base de armónicos circulares es de gran importancia, véase [Hsu-82b, Sheng-87, García-Martínez-95].

Hsu *et al.*, [Hsu-82b], propusieron un método iterativo para definir el centro del desarrollo en la nueva base. En una primera etapa, se toma como centro cualquier punto del espacio, usando este punto se calcula del desarrollo en armónicos circulares del objeto y se opera tal y como se ha descrito anteriormente para el método de los armónicos circulares, obteniéndose un pico de correlación en un punto del espacio que será tomado como nuevo centro del desarrollo. Por su parte, Sheng *et al.*, [Sheng-87], propusieron como criterio para determinar el centro del desarrollo aquel punto que maximizase la energía de cada término del desarrollo. El hecho de que en muchas ocasiones el origen de coordenadas que define este método puede estar muy alejado del centro geométrico, hace que se puedan ocasionar problemas de discriminación entre objetos. Así, se propuso un método basado en calcular simultáneamente un mapa de energía y otro mapa de PCE que permitieron una mejor definición del origen para la expansión en armónicos circulares [García-Martínez-95].

Dado que estos métodos de reconocimiento basados en armónicos circulares funcionan adecuadamente para el caso de objetos 2D, nos surge la posible aplicación de una generalización de los armónicos circulares para el caso tridimensional. Sin embargo, intuimos que la extensión al caso 3D, que es el que nos preocupa en este trabajo, no es en absoluto trivial. Por un lado hace falta una base vectorial de propiedades similares a los armónicos circulares. Por otro lado, en un problema tridimensional, el considerar las rotaciones aumenta la complejidad de cálculo en mayor medida que en el bidimensional. En 2D, identificar un objeto rotado equivale, en esencia, a determinar dos coordenadas espaciales y un ángulo, es decir, un total de tres grados de libertad. Sin embargo, en 3D equivale a calcular tres coordenadas espaciales y tres ángulos es decir, un total de seis parámetros.

Debido a estas razones, las líneas de investigación dedicadas al problema del reconocimiento de objetos 3D con invariancia a rotaciones han seguido otros caminos como, por ejemplo, los diversos métodos basados en la extracción de características [Campbell-03, Boyer-02]. Campbell *et al.*, [Campbell-03], propusieron un método basado en la extracción de características, principalmente de la curvatura a nivel local. Una vez extraídas, se clasifica el objeto de acuerdo con estas características y se compara con el objeto a reconocer. El modelo concreto de clasificación de curvaturas no es único y se puede definir diferentes conjuntos de clases, siendo ésta la principal diferencia del método propuesto por Boyer *et al.* [Boyer-02] con respecto al anterior.

En otra línea de investigación, Javidi *et al.* [Javidi-00, Kim-04], sugieren el uso de hologramas digitales aplicados al reconocimiento de objetos 3D con invariancia a diversas alteraciones, entre ellas la rotación. Ambos métodos se basan en registrar el objeto 3D y la escena dentro de la cual se quiere reconocer éste como hologramas digitales. Pero en el primer caso el

reconocimiento se realiza por un filtraje adaptado entre hologramas, lo que permite estimar la rotación del objeto original [Javidi-00]. En el segundo el reconocimiento invariante a distorsiones y rotaciones se lleva a cabo mediante un filtro compuesto [Kim-04].

Otras propuestas de reconocimiento de objetos 3D con invariancia a rotaciones se basan en usar como información del objeto sólo una vista 2D de él, como la propuesta por Li y Rosen [Li-00], o se enmarcan dentro de la resolución de otros problemas como la reconstrucción volumétrica integral de imágenes 3D [Hong-06].

Por todo ello, en vez de buscar un método análogo a los filtros con armónicos circulares en 2D, u optar por una de las opciones ya existentes en para el caso 3D, buscaremos un método de reconocimiento tridimensional que cumpla con los siguientes requisitos:

- Use la mayor parte de la información del objeto, no sólo una pequeña fracción.
- Mantenga invariancia a traslaciones de forma sencilla.
- No aumente de forma desmedida la complejidad del cálculo frente al caso bidimensional.

La solución que proponemos en este capítulo se basa en dos pasos claramente diferenciados. El primero trata de codificar el objeto 3D a reconocer, y en el segundo se define una operación análoga a la correlación entre la información codificada obteniéndose una respuesta invariante. Comencemos analizando con más detalle el primer paso de este proceso.

3.2 Codificación de objetos 3D

Dentro de esta fase de codificación, el primer paso será conseguir la información de entrada del objeto 3D que queremos reconocer utilizando la información tridimensional proporcionada por su imagen de rango. Posteriormente aplicaremos la PhFT a esa distribución tridimensional.

3.2.1 Imágenes de rango

Una representación estándar de objetos 3D en implementaciones ópticas son las imágenes de rango. Para entender qué son veamos el siguiente esquema en la Figura 3.1:s

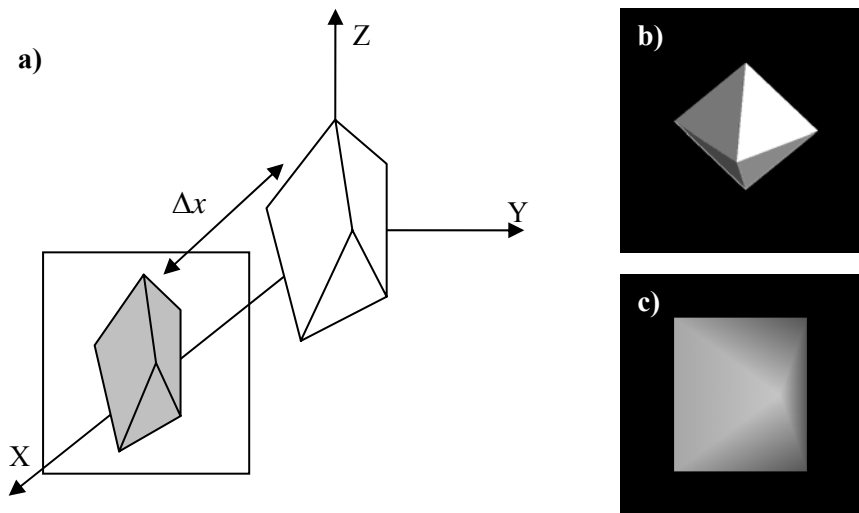


Figura 3.1 Definición de imagen de rango en a), un objeto tridimensional simple y su imagen de rango en b) y c).

Como vemos, una imagen de rango de un objeto tridimensional no es más que una imagen 2D en la que la intensidad, tono de gris, es proporcional a la distancia entre el objeto 3D y un determinado plano de referencia. Ésta es una

técnica bien conocida en la literatura [Chang-97, Laurin-95], para la que existen desarrollados métodos bien conocidos de obtención de dichas imágenes, como la perfilometría por transformada de Fourier, [Takeda-83, Esteve-Taboada-02].

Las imágenes de rango no contienen toda la información del objeto sino que sólo tenemos la información de la parte del objeto que es visible desde el plano de referencia, es decir, sólo vemos la superficie exterior que lo delimita, careciendo de información sobre las zonas internas u ocultas. Sin embargo, es posible tomar un conjunto de imágenes de rango, desde diferentes planos de referencia, a partir de las cuales se puede reconstruir la superficie externa completa de un objeto tridimensional.

3.2.2 Transformada de Fourier de fase (PhFT)

Una de las características más relevantes de las imágenes de rango son las propiedades de su transformada de Fourier de fase, PhFT, que definiremos a continuación. Esta operación matemática constituye la base de la codificación que proponemos en este trabajo y se define de la siguiente forma para una imagen de rango que viene dada por la función $x(y, z)$:

$$PhFT(u, v) \equiv F \left\{ e^{ikx(y, z)} \right\} \quad (3.6)$$

donde k es un factor constante que permite ajustar la pendiente de la fase del objeto. En lo que sigue trabajaremos suponiendo que $k = 1$, sin pérdida de generalidad. También se hace notar que la PhFT está definida en el espacio recíproco (u, v) .

Según la ecuación (3.6), la PhFT consiste en codificar la imagen de rango como una fase y calcular la transformada de Fourier del resultado. Esta codificación tiene múltiples propiedades según la alteración a que sometamos al objeto $x(y, z)$, como se verá a lo largo del trabajo. Veamos cuáles son algunas de estas propiedades de la PhFT.

La primera alteración que vamos a considerar es la traslación del objeto. Se puede decir que la PhFT es invariante a traslaciones del objeto en el sentido de que no cambia de valor cuando se desplaza el objeto. Siendo más precisos, sólo su valor absoluto es invariante. En este trabajo sólo trabajaremos con el valor absoluto de la PhFT, de forma que cuando hablemos de PhFT lo estaremos haciendo de su valor absoluto.

Para comprobar la invariancia a traslaciones de la PhFT supongamos traslaciones respecto a los distintos ejes coordenados. Si hacemos una traslación, de valor Δx , siguiendo un eje normal al plano de referencia de la imagen de rango, lo único que haremos será sumar un factor constante a la imagen de rango, que se reflejará en una multiplicación por un factor de fase constante en la PhFT:

$$PhFT((x(y, z)) + \Delta x) \equiv \left| F \left\{ e^{i(x(y, z) + \Delta x)} \right\} \right| = \left| F \left\{ e^{ix(y, z)} \right\} e^{i\Delta x} \right| = \left| F \left\{ e^{ix(y, z)} \right\} \right| \quad (3.7)$$

Más sencillo aún es el caso de traslación paralela al plano de referencia de la imagen de rango. De forma trivial la imagen de rango es la misma, ya que la transformada de Fourier de un objeto sometido a traslaciones laterales es invariante:

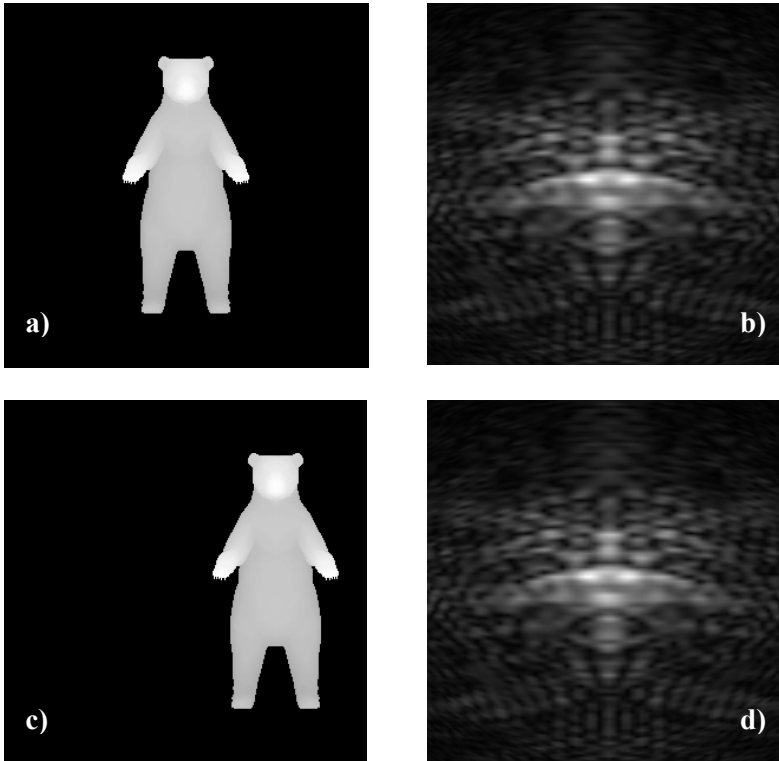


Figura 3.2 En a) y c) mostramos el mismo objeto, antes y después de sufrir una traslación, en b) y d) mostramos el valor absoluto de las PhFTs respectivas.

Nótese como las PhFTs son idénticas. En resumen, al trabajar con la PhFT de una imagen de rango, su norma nos proporciona invariancia a traslaciones de forma automática.

Las otras dos alteraciones geométricas aplicables a un objeto 3D, y que vamos a considerar en este trabajo, son alterar su tamaño, cambio de escala, y someterlo a una rotación arbitraria. El cambio de escala será tratado en el siguiente capítulo, mientras que las rotaciones serán el tema que nos ocupe en el presente.

Para entender la influencia de una rotación del objeto 3D en su PhFT consideremos su imagen de rango, que sólo nos da información sobre la

superficie externa del objeto. Esta superficie será, en general, curva, pero observada de manera local e infinitesimal, se comportará como un plano. Al cambiar de entorno local, observando otra zona del objeto, veremos ahora la superficie del objeto como otro plano, pero con una orientación diferente al anterior. Iterando y llevando este proceso al límite, podemos considerar la superficie del objeto 3D como compuesto por multitud de facetas planas cada una de ellas con su área y orientación respectiva.

Este modelo, aparte de permitirnos interpretar las propiedades de la PhFT frente a cambios de escala y rotaciones, está basado en uno de los métodos estándar para la representación 3D de objetos por ordenador, [OpenGL-05]. Las unidades de procesamiento gráfico, GPU en inglés, de las aceleradoras de video, y el software en general permiten manejar los objetos 3D como una malla, definida por una pareja de arrays, como ya se indicó en el capítulo anterior. En el primer array vienen definidas las posiciones de los vértices por sus coordenadas cartesianas y el segundo contiene grupos de tres vértices que forman una faceta infinitesimal de la malla. En la siguiente figura podemos ver la malla de un objeto 3D:

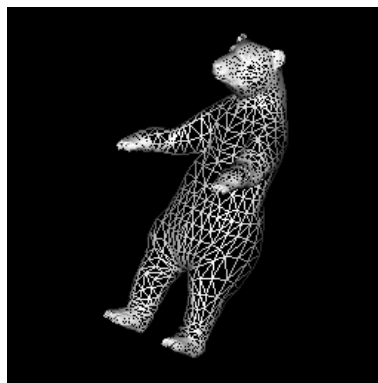


Figura 3.3 Malla de vértices y facetas de un objeto 3D

Una vez obtenida esta malla de facetas y vértices, se le pueden asignar diferentes colores, iluminaciones y texturas a las facetas, rotar el objeto, cambiarlo de escala, etc. Es decir, a nivel de implementación digital, un objeto 3D es considerado como un conjunto de facetas por numerosas aplicaciones. Por ello para analizar las propiedades de la PhFT sobre un objeto 3D alterado es lógico analizar dicha alteración global como composición de las alteraciones sobre cada una de las facetas.

Veamos cuál es la PhFT de una faceta cualquiera antes de realizar ninguna alteración. Como localmente cada faceta es un plano, vamos a calcular la PhFT de un plano infinito, y luego veremos cuál es la influencia de que la faceta no sea un plano infinito sino limitado. Para ello consideremos un plano determinado por una ecuación de la forma $x = Ay + Bz + C$:

$$PhFT(x = Ay + Bz + C) = \left| \int_{\mathbb{R}^2} e^{i2\pi(yu+zv)} e^{i(Ay+Bz+C)} dydz \right| = \quad (3.8)$$

$$\left| e^{iC} \int_{\mathbb{R}^2} \exp\left(i2\pi\left(y\left(u - \frac{A}{2\pi}\right) + z\left(v - \frac{B}{2\pi}\right)\right)\right) dydz \right| = \delta\left(u - \frac{A}{2\pi}\right) \cdot \delta\left(v - \frac{B}{2\pi}\right)$$

La PhFT de un plano infinito no es más que una delta de Dirac. Si en lugar de un plano tenemos una faceta finita de la forma

$$f(y, z) = (Ay + Bz + C) \cdot w(y, z) \quad (3.9)$$

donde el término $Ay + Bz + C$ nos da la orientación de la faceta y $w(y, z)$ es un factor de forma, igual a la unidad en el interior de la faceta y cero en el resto.

Al calcular la PhFT de esta faceta tenemos:

$$\begin{aligned}
 PhFT(u, v) &= FT \left\{ e^{i \cdot f(y, z)} \right\} = \int_{\mathbb{R}^2} e^{i(Ay + Bz + C) \cdot w(y, z)} e^{-i2\pi(uy + vz)} dx dy = \\
 &= \int_{(y, z) \in w} e^{i(Ay + Bz + C) \cdot w(y, z)} e^{-i2\pi(uy + vz)} dx dy + \int_{(y, z) \notin w} e^{-i2\pi(uy + vz)} dx dy = \\
 &= e^{ic} \delta \left(u - \frac{a}{2\pi}, v - \frac{b}{2\pi} \right) \otimes W(u, v) \tag{3.10}
 \end{aligned}$$

Es decir, si a un plano infinito le corresponde un punto en el espacio recíproco de la PhFT, a una faceta finita le corresponde una función más o menos picada centrada en un punto que sólo depende de la orientación de la faceta (A, B) y no de su posición exacta que vendría expresada por C . Este comportamiento es análogo a la transformada de Fourier de una onda plana, e^{ikx} , cuyo resultado es una delta de Dirac, mientras que la de una onda limitada le corresponde una función centrada en la delta, pero con cierto ancho de banda. En la siguiente figura mostramos un objeto sencillo y su PhFT correspondiente:

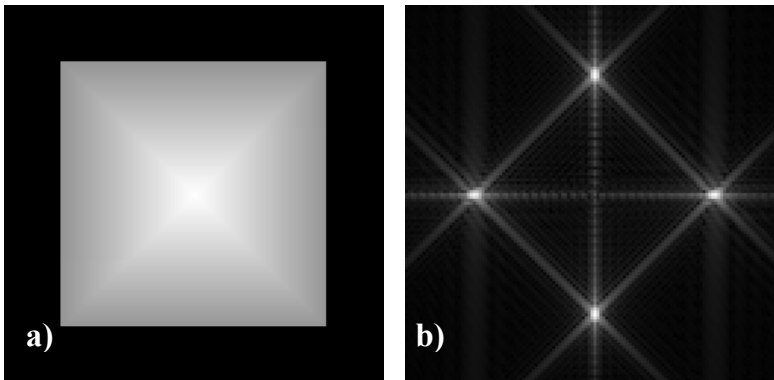


Figura 3.4 Imagen de rango de un objeto sencillo (pirámide) en a) en b) se muestra la PhFT correspondiente al objeto a)

Utilizando estos argumentos, podemos confirmar de nuevo la invariancia a traslaciones de la PhFT. Si hacemos una traslación a una faceta, da igual en que dirección, lo único que le pasará es que cambiaremos el parámetro C , pero no los A, B , con lo que la PhFT de una faceta trasladada es igual a la original, demostrándose una vez más la invariancia a traslaciones de la PhFT.

Consideremos ahora que el objeto 3D es más complejo que el mostrado anteriormente y que lo sometemos a una rotación. Antes de ser rotado, su PhFT es la composición de las PhFTs correspondientes a todas y cada una de las facetas visibles desde el plano de referencia y con la orientación en la que esté el objeto.

Una vez realizada la rotación, cada una de las pequeñas facetas que componían la vista anterior del objeto se ven rotadas a su vez. Si la rotación es suficientemente pequeña, todas las facetas ven alterada su orientación de la misma manera, por esto y recordando la ecuación (3.8) se tiene que a primer orden un giro se corresponde, de manera lineal, con una traslación de la PhFT. Así que en un principio podíamos suponer que lo único que le pasa a la PhFT de un objeto rotado es que sufre una traslación. Pero esto sólo es así en primera aproximación, ya que hemos obviado un par de efectos. El primero es que a medida que rotamos un objeto, la escala con que vemos cada faceta varía ya que la proyección que vemos varía según su orientación. El segundo es que ahora aparecerán partes del objeto antes ocultas y, de la misma manera, partes que antes se veían, ahora dejar de ser visibles. De hecho ambos fenómenos están relacionados, dejamos de ver una faceta perteneciente al objeto cuando ésta está orientada de manera normal al plano de referencia, es decir cuando su sección visible es nula. Conforme nos acercamos a esta situación, el área que es visible se reduce, con lo que la intensidad correspondiente en su PhFT disminuye. Es decir, la PhFT de un objeto rotado es básicamente la misma, pero desplazada una cantidad

proporcional a la rotación, y con la aparición de información nueva por los bordes de la PhFT que corresponde a las partes que antes no se podían ver. De la misma manera va desapareciendo información por los bordes, correspondiente a lo que deja de ser visible. En las siguientes figuras ilustramos este efecto:

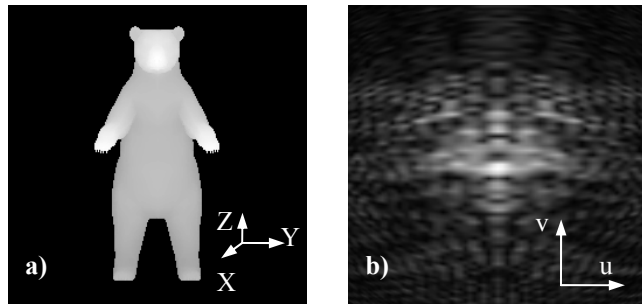


Figura 3.5 Objeto sin rotar en a), y su PhFT en b).

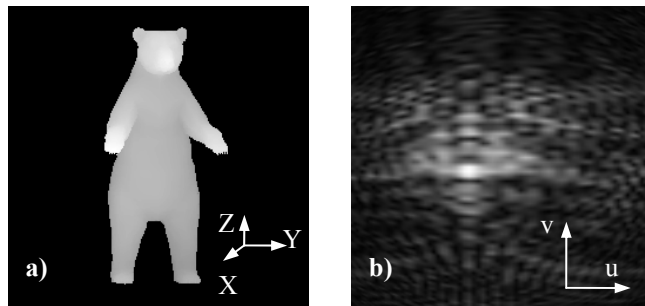


Figura 3.6 Objeto rotado 8 grados en torno a Z en a), y su PhFT en b).

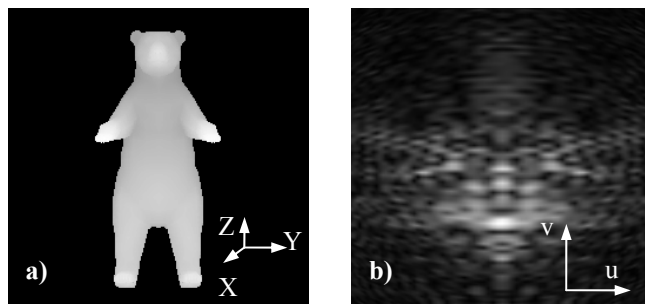


Figura 3.7 Objeto rotado 8 grados en torno a Y en a), y su PhFT en b).

Se comprueba que las figuras anteriores confirman lo anteriormente mencionado: que una rotación del objeto 3D, respecto a cualquier eje coordenado, se traduce, a primer orden, en una traslación de la PhFT asociada.

Veamos qué ocurre ahora si en vez de someter al objeto a una rotación respecto a un eje, lo generalizamos a rotaciones con respecto a todos los ejes.

3.2.3 3DOOM

Tal y como hemos mencionado en la introducción, el método propuesto tiene una fase en el que se codifica la información 3D para una posterior utilización de algoritmos de reconocimiento. Esta codificación se basa en calcular las PhFTs de las distintas imágenes de rango obtenidas a partir de la rotación sucesiva del objeto tridimensional de entrada. Cada una de estas PhFTs se convierte en una ligera traslación de la anterior más una pequeña información adicional perteneciente a las partes de objeto que ahora son visibles y antes no.

Llevando al límite el procedimiento de rotar y calcular PhFTs, nos podemos imaginar que cada vez que evaluamos la PhFT de una vista parcial del objeto rotado, lo único que hacemos es calcular parte de una función recíproca al objeto 3D, denominado 3DOOM, que contiene toda la información geométrica del objeto de partida. Si la rotación se restringiese a un único eje, como el Z, el 3DOOM tendría la forma de una banda cilíndrica y paralela a dicho eje. Si se toma el caso de rotaciones completamente generales el 3DOOM se define en la superficie de una esfera.

Este mapa contiene codificada toda la información geométrica de un objeto y se puede entender como un tapiz hecho a base de parchear las PhFTs de las respectivas vistas parciales al ir rotando el objeto, como se muestra en un caso parcial en [Esteve-Taboada-03].

El procedimiento para construirlo es como sigue: Para un objeto dado se selecciona una esfera de radio arbitrario centrada en el propio objeto. Con la esfera definida, consideramos imágenes de rango cuya orientación viene dada por un muestreo de puntos pertenecientes a la esfera. Una vez obtenidas las distintas imágenes de rango calculamos las PhFTs y con éstas vamos construyendo todas las áreas del 3DOOM completo, que será nuestra codificación del objeto. Una vez introducida la codificación 3DOOM resulta conveniente cambiar nuestra notación. En la ecuación (3.6) hemos notado como recorrido de definición de la PhFT un espacio recíproco (u, v) . Como, en lo sucesivo, la PhFT estará asociada a una superficie esférica, resulta conveniente cambiar a unas coordenadas más apropiadas. De acuerdo con la siguiente figura:

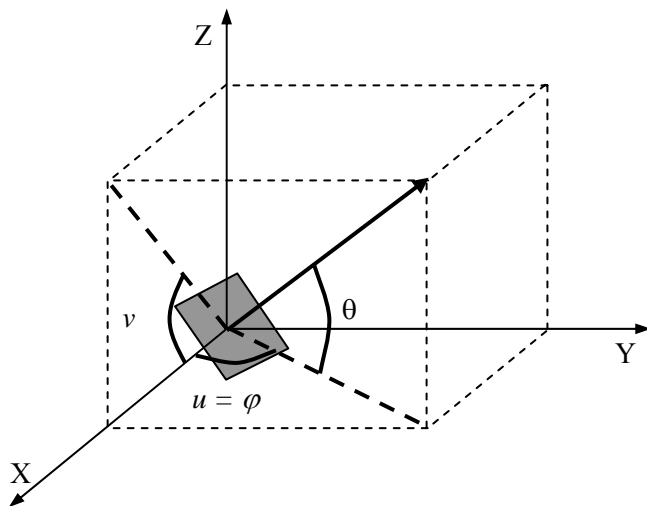


Figura 3.8 Relación entre coordenadas rectangulares y esféricas

Se deduce la siguiente relación las coordenadas (u, v) y (φ, θ) :

$$\begin{aligned} u &= \varphi \\ \operatorname{tg}(v) &= \frac{\operatorname{tg}(\theta)}{\cos(\varphi)} \end{aligned} \quad (3.11)$$

En lo que sigue consideraremos el 3DOOM en las coordenadas (φ, θ) , aun cuando a la hora de representarlo gráficamente lo hagamos indistintamente en una esfera, Fig. 3.9 (b), o en una proyección plana Fig. 3.9 (c), como se muestra en la siguiente figura:

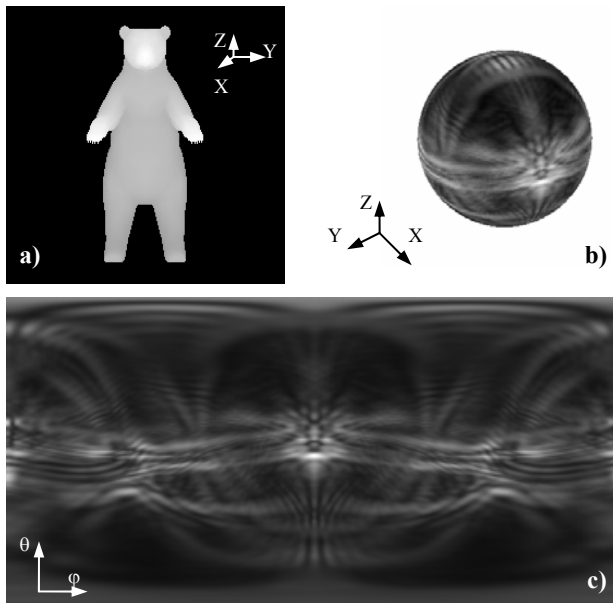


Figura 3.9 Imagen de rango de un objeto 3D en a) y su 3DOOM, representado en forma esférica b), y plana c)

En la Figura c) se observa que la proyección plana del 3DOOM muestra la totalidad de esta función, mientras que la imagen b) de la figura muestra el 3DOOM en la superficie de una esfera que constituye su espacio de definición. Esta representación no puede mostrar la totalidad de la función, al quedar ocultas ciertas zonas.

El convenio usado para representar el 3DOOM como una proyección plana es el siguiente: teniendo en cuenta que $\varphi = [0, 2\pi)$, $\theta = [-\pi, \pi]$, la zona de $\theta = 0$ es la línea horizontal intermedia y, además, los bordes laterales están unidos, ya que la coordenada φ es continua en su intervalo de definición.

Insistimos en que el 3DOOM no es una función plana de \mathbb{R}^2 , aún cuando se pueda representar de forma similar por comodidad, sino una función definida en la superficie de una esfera, en S^2 . Por ejemplo, a la hora de construirlo, mediante la yuxtaposición de las PhFTs de diversas zonas, sólo las correspondientes a la zona central tienen una forma rectangular en la proyección plana, el resto tiene diferente forma, tal y como ilustra la siguiente figura:

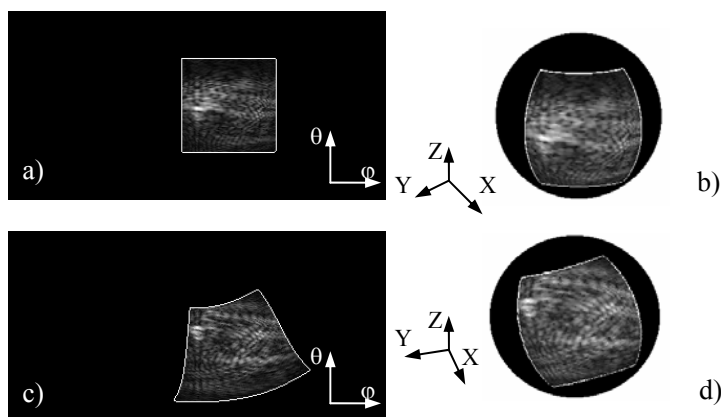


Figura 3.10 Partes de un 3DOOM en una proyección plana en a) y c),
y las mismas partes representadas en forma esférica en b) y d)

Nótese como la forma del parche correspondiente a ambas PhFTs tiene la misma forma en S^2 , ver Fig. 3.10 (b) y (d). El cambio de forma en la representación plana no es más que un defecto de ésta, no una alteración real de los datos.

La Figura 3.10 sirve también para ilustrar el proceso de generación del 3DOOM mediante el parcheo de diferentes PhFTs. Lo que se desea es fijar un sistema de coordenadas al objeto y rotarlo. Tras esto, para cada orientación del objeto 3D, se calcula la PhFT como una función plana definida en (u,v) según la ecuación (3.6). A continuación se debe pasar a las coordenadas esféricas (φ,θ) con la transformación de coordenadas de la ecuación (3.11). Y, por último, se deshace la rotación inicial, mediante la aplicación de las matrices de rotación convencionales y se pega la información de la PhFT en el 3DOOM definido en el espacio S^2 .

En la siguiente figura representamos dos sistemas de coordenadas estáticos con los pares de coordenadas (u,v) y (φ,θ) :

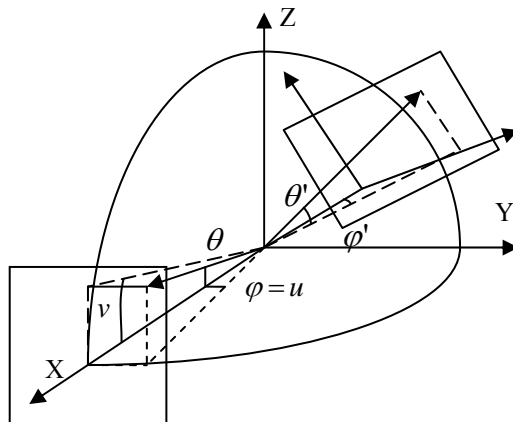


Figura 3.11 Cambio de coordenadas entre sistemas de coordenadas estáticos

Iterando este proceso, generamos el 3DOOM completo, ver Fig. 3.9 b), que contiene toda la información geométrica del objeto 3D. Esta es la codificación que usaremos a lo largo de este capítulo.

3.3 Correlaciones y análisis en SO(3)

Ahora, una vez vista la codificación y siguiendo el esquema que hemos explicado en la introducción, hemos de ver cómo se opera con esta información codificada.

El 3DOOM contiene todas las posibles PhFTs de un objeto correspondientes a todas las posibles rotaciones a la que este se pueda someter. Así, usando esta codificación, y dado un objeto 3D con una rotación determinada el problema de la detección con invariancia a rotaciones se ha reducido a la detección de una PhFT, determinada por el giro del objeto 3D, en el 3DOOM.

3.3.1 Correlación en SO(3)

Dado que queremos encontrar un parche (o una PhFT) en la superficie del 3DOOM, hemos de definir una correlación adaptada a la superficie de la esfera. Esto es análogo al problema básico del reconocimiento de objetos en dos dimensiones. Si en dos dimensiones, para detectar un objeto $f(x, y)$ en una escena $g(x, y)$, hemos de hacer la correlación en \mathbb{R}^2 , en S^2 habremos de definir una operación equivalente a la correlación entre dos funciones $f(\varphi, \theta)$ y $g(\varphi, \theta)$ definidas sobre la superficie de una esfera.

En \mathbb{R}^2 la correlación se define como:

$$\begin{aligned}
 f(x, y) * g(x, y) &= \int_{\mathbb{R}^2} f^*(x' - x, y' - y) \cdot g(x', y') dx' dy' = \\
 \int_{\mathbb{R}^2} T(f(x', y'); x, y)^* \cdot g(x', y') dx' dy' &= \langle T(f; x, y) | g \rangle
 \end{aligned}
 \tag{3.12}$$

donde por $T(f; x, y)$ notamos el operador de traslación que centra la función f en el par de coordenadas (x, y) . Es decir, la correlación convencional es una operación que traslada una función con relación a otra, las multiplica e integra el producto.

La correlación en S^2 será ,por analogía, y de acuerdo con [Kostelec-03], una operación definida entre dos funciones $f(\varphi, \theta)$ y $g(\varphi, \theta)$ definidas a su vez sobre la esfera, tal que la primera función se rote una determinada cantidad con relación a la otra, se multipliquen y se integre el producto:

$$\begin{aligned}
 f(\theta, \varphi) * g(\theta, \varphi) &\equiv \int_{S^2} R(f; \alpha, \beta, \gamma)^* g(\theta, \varphi) \text{sen}(\theta) d\theta d\varphi = \\
 \langle D(f; \alpha, \beta, \gamma) | g \rangle &= (f *_{SO(3)} g)(\alpha, \beta, \gamma)
 \end{aligned}
 \tag{3.13}$$

donde $R(f; \alpha, \beta, \gamma)$ es la notación que empleamos para indicar la rotación de la función $f(\theta, \varphi)$, $D(f; \alpha, \beta, \gamma)$ es el vector resultante de la aplicación del operador rotación sobre $f(\theta, \varphi)$ y el trío (α, β, γ) son los ángulos de Euler, de los que daremos más detalle en el siguiente apartado.

Lo más relevante de la ecuación (3.13) es que la correlación en S^2 es una función cuyas variables son los ángulos de Euler. Esto es, la correlación de

dos funciones en S^2 está definida en el grupo $SO(3)$ de rotaciones en \mathbb{R}^3 [Kostrikin-78]. Además, ésta no es la única definición de correlación en S^2 . Usando un criterio diferente, es posible definir una convolución y a partir de ahí una correlación:

$$f(\theta, \varphi) \otimes g(\theta, \varphi) \equiv \sum_{j,m} Y_{j,m}(\theta, \varphi) CONV_{j,m} \quad (3.14)$$

$$\text{con } CONV_{j,m} \equiv 2\pi \sqrt{\frac{4\pi}{2j+1}} F_{j,m} \cdot G_{j,0}$$

donde los $F_{j,m}$ y $G_{j,0}$ son análogos a la transformada de Fourier en \mathbb{R}^2 y se definen en detalle en (3.20).

Pero la expresión (3.14) no resulta útil para el reconocimiento de objetos ya que no da un criterio de semejanza entre $f(\theta, \varphi)$ y $g(\theta, \varphi)$. A pesar de ello, esta definición tiene su ámbito de uso y posee interés teórico por sí misma [Healy-98].

Así pues, será la ecuación (3.13), junto con la codificación 3DOOM, la que usemos para el reconocimiento de objetos y tiene las siguientes características:

- El 3DOOM tiene toda la información de la superficie externa del objeto. Luego estamos aprovechando la mayor parte de la información disponible.
- Siendo el módulo de la PhFT invariante a traslaciones, también lo es el 3DOOM. Luego nuestro método es invariante a traslaciones de forma automática.

- El problema del reconocimiento 3D invariante a rotaciones tiene seis grados de libertad si se aborda de forma directa. Con nuestro planteamiento, estos grados de libertad se reducen a la mitad. Con ello se reduce la dimensionalidad del problema en gran medida.
- De forma adicional, si la correlación en \mathbb{R}^2 nos da la posición en el plano del objeto que buscamos, la correlación en S^2 nos da la rotación a que ha sido sometido. En sentido estricto el método que proponemos es un sistema lineal frente a rotaciones, más que invariante. Es decir, tendremos que el valor de pico de la salida es constante y su posición varía de forma proporcional a la rotación del objeto a reconocer, con lo que podemos estimar dicha rotación del objeto. Esto puede interpretarse como una ventaja si lo comparamos con un método estrictamente invariante.

Por contra, la correlación en S^2 en la forma (3.13) presenta el mismo problema que el uso de la correlación convencional según la ecuación (3.12), el iterar el proceso rotación-multiplicación-integración es pesado matemáticamente.

3.3.2 Análisis de Fourier en SO(3)

En la práctica, y más en el campo del reconocimiento de objetos, la correlación entre dos funciones no se calcula mediante una iteración de traslaciones, etc., sino que se calcula de forma rápida mediante el uso de transformadas de Fourier como:

$$f(x, y) * g(x, y) = F^{-1} \{ F^* \{ f \} \cdot F \{ g \} \} \quad (3.15)$$

En el capítulo 2 se introdujo una estructura de espacio vectorial sobre el conjunto de funciones en \mathbb{R}^2 . Usando esa estructura, la transformada de Fourier se puede reinterpretar de forma vectorial:

$$F\{f(x, y)\} = \int_{\mathbb{R}^2} e^{-i2\pi(uv+vy)} f(x, y) dx dy = \langle u, v | f \rangle \quad (3.16)$$

en donde $|u, v\rangle = e^{i2\pi(xu+yv)}$.

Nótese que la ecuación (3.16), considerada para todo el espacio frecuencial (u, v) , no es otra cosa que el cambio de coordenadas en la base $\{|u, v\rangle = e^{i2\pi(xu+yv)}, \forall u, v \in \mathbb{R}^2\}$ ¹. Es decir, la transformada de Fourier en \mathbb{R}^2 es un cambio de la base canónica $\{\delta(x-x', y-y'), \forall x', y' \in \mathbb{R}^2\}$ a la $|u, v\rangle$.

Si una vez más usamos la analogía entre \mathbb{R}^2 y S^2 , el equivalente de la expresión (3.15), que permite calcular correlaciones de forma rápida, será también un cambio de coordenadas, pero ahora en el conjunto de las funciones definidas en la superficie de una esfera. A este espacio se le puede dar estructura de espacio vectorial mediante la definición de un producto escalar análogo al caso de \mathbb{R}^2 , tal y como muestra la siguiente expresión:

$$\langle f | g \rangle \equiv \int_{S^2} f^*(\theta, \varphi) g(\theta, \varphi) \cdot \text{sen}(\theta) d\theta \cdot d\varphi \quad (3.17)$$

en donde estamos siguiendo la convención de coordenadas de la Figura 3.8.

¹ Nótese que los $|u, v\rangle$ forman una base ortonormal ya que:

$$\langle u_1, v_1 | u_2, v_2 \rangle = \int_{\mathbb{R}^2} e^{-i2\pi(xu_1+yv_1)} e^{i2\pi(xu_2+yv_2)} dx dy = \delta(u_2 - u_1, v_2 - v_1)$$

Es trivial comprobar que este producto escalar cumple con la axiomática necesaria expresada en las ecuaciones (2.8.1-4), con lo que ahora ya se puede hablar de las funciones en S^2 como un espacio vectorial de forma rigurosa.

De la misma forma que en \mathbb{R}^2 , podemos decir que las funciones $f(\varphi, \theta)$ son vectores $|f\rangle$ expresados en coordenadas de una base canónica formada por deltas de Dirac:

$$|\varphi, \theta\rangle = \delta(\varphi, \theta) \quad (3.18)$$

Ahora el siguiente paso es buscar una base ortonormal de vectores en S^2 , de forma similar a los $e^{i2\pi(xu+yv)}$, para que así puedan generar todo el espacio, sean fáciles de calcular y tengan unas propiedades bien conocidas.

De entre todas las familias de funciones que existen en S^2 , hemos escogido los armónicos esféricos, ya que cumplen con todos los requisitos expuestos anteriormente. Son ortonormales, generan todo el espacio de funciones en S^2 , y son funciones muy bien conocidas y que tienen propiedades matemáticas muy interesantes para trabajar con problemas que involucren rotaciones [Brezis-84]. Además, siguiendo con la analogía entre \mathbb{R}^2 y S^2 , si $e^{i2\pi(xu+yv)}$ es la expresión de las ondas planas, los armónicos esféricos son la solución de la ecuación de ondas en coordenadas esféricas, [Jackson-99]. La definición concreta que vamos a usar es la siguiente:

$$Y_{j,m}(\theta, \varphi) \equiv \sqrt{\frac{2j+1}{4\pi} \frac{(j-m)!}{(j+m)!}} P_j^m(\cos(\theta)) e^{im\varphi} = |j, m\rangle \quad (3.19)$$

con $j=0,1,\dots$ y $m=-j,\dots,0,\dots,+j$, y siendo P_j^m los polinomios asociados de Legendre de orden j y grado m , funciones de propiedades también muy conocidas [Zwillinger-96]. En la siguiente figura mostramos cuál es el aspecto gráfico de algunos de estos armónicos:

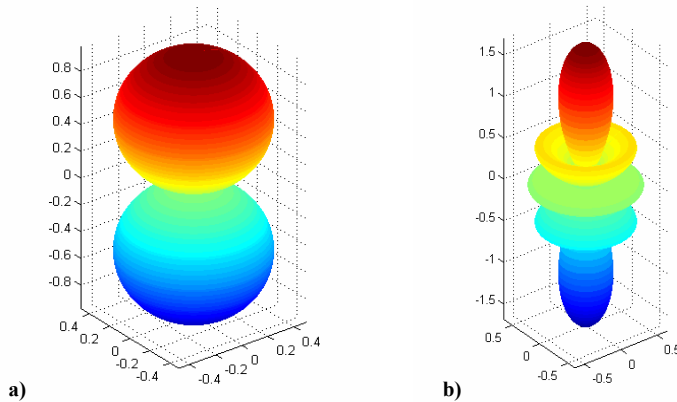


Figura 3.12 Parte real de $|1,0\rangle$ en a) y de $|4,0\rangle$ en b)

Con nuestra nueva base, ya podemos definir una transformada de Fourier en S^2 :

$$F\{g(\theta,\varphi)\} \equiv G_{j,m} \equiv \langle j,m|g\rangle = \int_{S^2} Y_{j,m}^*(\theta,\varphi) g(\theta,\varphi) \cdot \text{sen}(\theta) d\theta \cdot d\varphi \quad (3.20)$$

Y la transformada inversa:

$$F^{-1}\{G_{j,m}\} \equiv \langle \theta,\varphi|f\rangle = \sum_{j,m} \langle \theta,\varphi|j,m\rangle \langle j,m|f\rangle = \sum_{j,m} Y_{j,m}(\theta,\varphi) G_{j,m} \quad (3.21)$$

Nótese que, aunque las funciones en S^2 se definen en términos de las coordenadas esféricas (θ, φ) , la correlación entre dos funciones de este tipo se define en un espacio diferente, el de rotaciones. Ahora falta expresar la correlación en $SO(3)$, dada por la ecuación (3.13) en términos de esta base. Se puede comprobar que toda rotación en torno a cualquier eje se puede descomponer en tres rotaciones en torno a los ejes coordenados [Sakurai-85]. Cada una de estas rotaciones es de una amplitud que viene expresada por los llamados ángulos de Euler. En la siguiente figura mostramos la convención usada en este trabajo:

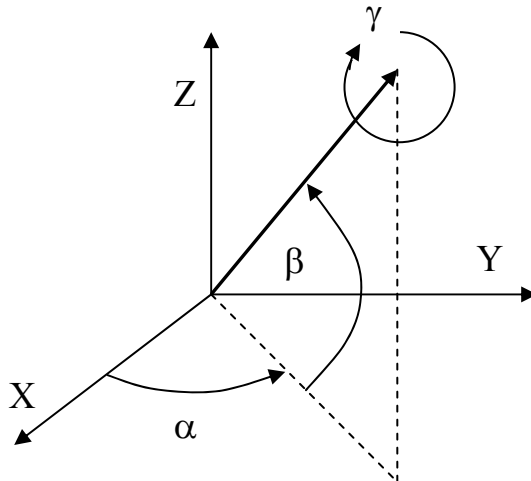


Figura 3.13 Definición de los ángulos de Euler

Es decir una rotación $R(\alpha, \beta, \gamma)$, se puede descomponer en la composición de rotaciones en torno a los ejes coordenados:

$$R(\alpha, \beta, \gamma) \equiv R_z(\gamma)R_y(\beta)R_z(\alpha) \quad (3.22)$$

con $\alpha \in [-\pi, \pi)$, $\beta \in [-\pi, \pi]$ y $\gamma \in [-\pi, \pi)$.

Esta descomposición no es única. En este trabajo seguiremos la convención usual en la mecánica cuántica; en la mecánica clásica la segunda rotación es en torno al eje X' [Goldstein-02, Marion-75]. Ambas convenciones son válidas y usar una u otra es cuestión de preferencia. Además estas rotaciones tienen lugar en torno a los ejes que se mueven solidariamente con el objeto (con prima). Convendría que las rotaciones se efectuasen en torno a los ejes fijos del espacio como [Sakurai-85]:

$$R(\alpha, \beta, \gamma) \equiv R_z(\alpha) R_y(\beta) R_z(\gamma) \quad (3.23)$$

Siendo los intervalos de definición de los ángulos los mismos que en (3.22)

Una vez caracterizadas las rotaciones en forma de los ángulos de Euler, vamos a operar sobre la definición de correlación en S^2 :

$$f(\theta, \varphi) * g(\theta, \varphi) = CORR(\alpha, \beta, \gamma) \equiv \int_{S^2} R(f; \alpha, \beta, \gamma)^* g(\theta, \varphi) \text{sen}(\theta) d\theta d\varphi \quad (3.24)$$

Si introducimos la definición de los ángulos de Euler, queda:

$$\begin{aligned} \int_{S^2} \left((R_z(\alpha) R_y(\beta) R_z(\gamma)); f \right)^* g(\theta, \varphi) \text{sen}(\theta) d\theta d\varphi = \\ \left\langle (D_z(\alpha) D_y(\beta) D_z(\gamma)) f \middle| g \right\rangle = \left\langle D(\alpha, \beta, \gamma) f \middle| g \right\rangle \end{aligned} \quad (3.25)$$

Expresión en la que seguimos la notación más usual en mecánica cuántica, en la que los operadores de rotación actuando sobre un vector se representan por la letra D .

Ahora veremos cómo actúan estos operadores sobre los vectores $|f\rangle$. Por suerte éste es un problema básico de la mecánica cuántica [Dirac-58, Galindo-89, Sakurai-85], por tanto está resuelto y simplemente podemos aprovechar todo el soporte matemático desarrollado para las rotaciones. Además, el haber escogido la base de armónicos esféricos simplificará el cálculo.

La expresión del operador rotación para una rotación en torno a un eje cualquiera expresado por un vector \hat{n} , es la siguiente:

$$D(\phi; \hat{n}) = e^{-i\phi \vec{J} \cdot \hat{n}} \quad (3.26)$$

donde $\vec{J} = J_x \hat{i} + J_y \hat{j} + J_z \hat{k}$ es un operador análogo al momento angular. Con esto queremos expresar que \vec{J} es una operación matemática tal que cumple las mismas propiedades que el momento angular cuántico, pero sin que tenga asociada ninguna magnitud observable. Para nosotros \vec{J} es un operador del que conocemos sus propiedades, y que nos permite realizar los cálculos que siguen.

Aclarado este punto, las propiedades del operador momento angular actuando sobre los armónicos esféricos son conocidas: los armónicos esféricos son los autovalores de J_z , la componente Z del momento angular, y de J^2 , el módulo al cuadrado de éste, así que son una base idónea para trabajar con este tipo de problemas; ésta fue otra de las razones que nos llevaron a escogerla:

$$J_z |j, m\rangle = m |j, m\rangle \quad (3.27.1)$$

$$J^2 |j, m\rangle = (j+1)j |j, m\rangle \quad (3.27.2)$$

Hacemos notar que con lo anterior estamos incurriendo en un pequeño abuso del lenguaje. Hablamos de los armónicos esféricos como si fueran todos los $|j, m\rangle$ del formalismo cuántico, cuando en verdad no es exactamente así. Los $|j, m\rangle$ en cuántica pueden tener j tanto entero como semientero, y los armónicos esféricos sólo cubren el caso entero. Esta imprecisión en el lenguaje no es relevante, ya que los armónicos esféricos forman base de las funciones de S^2 . Además los armónicos esféricos son las componentes de los vectores $|j, m\rangle$ en la base usual del espacio:

$$Y_{j,m}(\theta, \varphi) = \langle \hat{n}(\theta, \varphi) | j, m \rangle \quad (3.28)$$

Como es más práctico decir “armónico esférico” que “coordenadas de los vectores $|j, m\rangle$ en la base canónica”, hemos empleado esa notación poco precisa.

Consideraremos de nuevo la expresión de la correlación entre dos funciones de S^2 . Dada la ecuación (3.25) e introduciendo una identidad de Parseval $\sum_{j,m} |j, m\rangle \langle j, m|$, para poder pasar a la base de los vectores $|j, m\rangle$:

$$\begin{aligned} f(\theta, \varphi) * g(\theta, \varphi) &= \dots = \langle D(\alpha, \beta, \gamma) f | g \rangle = \dots \\ &= \langle D(\alpha, \beta, \gamma) f | g \rangle = \sum_{j,m} \langle D(\alpha, \beta, \gamma) f | j, m \rangle \langle j, m | g \rangle = \dots \end{aligned} \quad (3.29)$$

Recordando que $\langle j, m | g \rangle \equiv G_{j,m}$ es la transformada de Fourier, en S^2 de $g(\varphi, \theta)$, cambiando de orden el primer término y usando los ángulos de Euler:

$$\dots = \sum_{j,m} \langle j,m | D(\alpha, \beta, \gamma) | f \rangle^* G_{j,m} = \sum_{j,m} \langle j,m | D_z(\alpha) D_z(\alpha) D_z(\alpha) | f \rangle^* G_{j,m} = \dots \quad (3.30)$$

Introducimos otra identidad de Parseval:

$$\begin{aligned} \dots &= \sum_{j,j',m,m'} \langle j,m | D_z(\alpha) D_z(\alpha) D_z(\alpha) | j',m' \rangle^* \langle j',m' | f \rangle G_{j,m} = \\ &\sum_{j,j',m,m'} \langle j,m | D_z(\alpha) D_z(\alpha) D_z(\alpha) | j',m' \rangle^* F_{j',m'} G_{j,m} = \dots \end{aligned} \quad (3.31)$$

Ya sabemos que $D_i(\phi) = e^{-iJ_i\phi}$ y en el caso de las rotaciones en torno al eje Z el resultado es inmediato aplicarlas a un $|j,m\rangle$ según la Ec. (3.27.1), $D_z(\phi) |j,m\rangle = e^{-iJ_z\phi} |j,m\rangle = e^{-im\phi} |j,m\rangle$. Sustituyendo en lo anterior:

$$\dots = \sum_{j,j',m,m'} \langle j,m | e^{-iJ_y\beta} | j',m' \rangle^* F_{j',m'} G_{j,m} e^{im\gamma} e^{im\alpha} \quad (3.32)$$

Donde sólo nos queda por conocer el valor de los $\langle j,m | e^{-iJ_y\beta} | j',m' \rangle$, que también están estudiados. Se los conoce como los elementos de matriz reducida, matrices de Wigner o símbolos d . Las propiedades más importantes son las siguientes.

Se representan por $d_{m,m'}^j(\beta)$, son reales y sólo tienen los índices, m , m' y j . En nuestro caso desaparecerá un sumatorio en j o j' . Es posible hallar una expresión completa para ellos, por ejemplo en [Sakurai-85] se da esta:

$$d_{m',m}^j(\beta) = \sum_k (-1)^{k-m+m'} \frac{\sqrt{(j+m)!(j-m)!(j+m')!(j-m')!}}{(j+m-k)!k!(j-k-m')!(k-m+m')!} \left(\cos\left(\frac{\beta}{2}\right)\right)^{2j-2k+m-m'} \left(\operatorname{sen}\left(\frac{\beta}{2}\right)\right)^{2k-m+m'} \quad (3.33)$$

Con k cualquier entero para el que estén definidos los factoriales. La dificultad para calcular esta expresión en la práctica es que es imposible obtener los valores de factoriales demasiado altos, ya que a partir de un valor relativamente pequeño el valor de los factoriales es muy elevado para calcularlos numéricamente. Aparte de no ser posible calcularlos a partir de un cierto orden, la expresión (3.33) implica el calcular cocientes entre ellos, con lo que la precisión de los que pudiésemos obtener en los términos calculables sería baja.

Una expresión alternativa y más práctica es la aportada por Kostelec y Rockmore, [Kostelec-03] que es la siguiente relación de recurrencia y será la que usaremos en este trabajo:

$$0 = \frac{\sqrt{[(j+1)^2 - m^2]} \cdot \sqrt{[(j+1)^2 - m'^2]}}{(j+1)(2j+1)} d_{m,m'}^{j+1}(\beta) + \left(\frac{mm'}{j(j+1)} - \cos(\beta)\right) d_{m,m'}^j(\beta) + \frac{\sqrt{[j^2 - m^2]} \cdot \sqrt{[j^2 - m'^2]}}{j(2j+1)} d_{m,m'}^{j-1}(\beta) \quad (3.34)$$

Como condición de inicio de la relación de recurrencia se usa la ecuación (3.33) para los elementos de la forma $d_{j,m}^j(\beta)$, $d_{-j,m}^j(\beta)$, etc.

Con esto, retomando (3.32) la expresión de la correlación en SO(3) queda de la siguiente forma:

$$f(\theta, \varphi) * g(\theta, \varphi) = \dots = \sum_{j,m,m'} d_{m,m'}^j(\beta) \cdot F_{j,m'} G_{j,m} e^{im'y} e^{im'x} \quad (3.35)$$

Esta es la expresión de la correlación en S^2 en el espacio de Fourier y será la que usemos para implementar nuestro método, ya que es más rápida que la implementación directa de la ecuación (3.13), al igual que sucede en el caso de \mathbb{R}^2 .

3.4 Implementación y resultados

En este apartado analizaremos cómo implementar el método descrito en los apartados anteriores, además de los resultados obtenidos con él. Antes de comenzar indicaremos que el método ha sido implementado de forma digital, usando la versión de Matlab 6.5 r13. Para facilitar la comprensión del algoritmo de cálculo mostramos el siguiente diagrama de flujo de la figura:

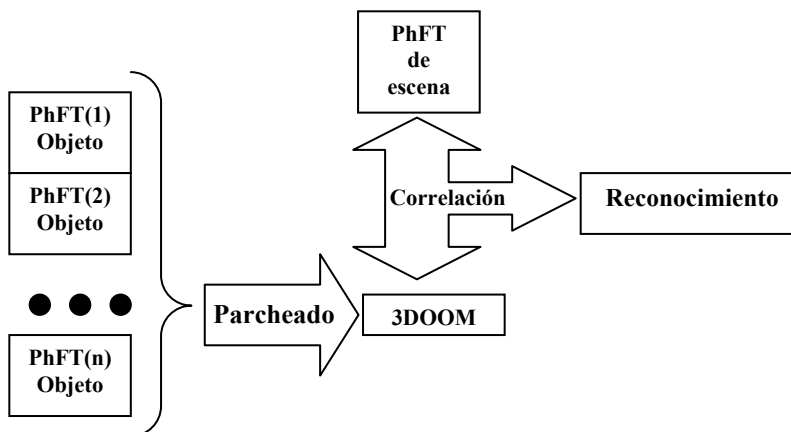


Figura 3.14 Diagrama de flujo del método

A lo largo de este capítulo seguiremos ese mismo orden y empezaremos explicando como hallar la PhFTs de un objeto 3D.

3.4.1 Cálculo de PhFTs y composición del 3DOOM

Calcular la PhFT de forma digital es inmediato si se tiene la imagen de rango correspondiente. Así, en un primer paso hemos de obtener las imágenes de rango de objetos tridimensionales para posteriormente calcular las PhFTs.

Como único dato de partida contamos con los objetos 3D codificados en el formato de facetas y vértices. En la introducción hemos indicado que con este método de representar objetos tridimensionales podemos hacer todas las operaciones matemáticas que necesitamos. Así, una rotación del objeto tridimensional se reduce a la aplicación de una matriz de rotación convencional sobre todos los elementos del array de vértices; una traslación es una suma de un vector constante a todos vértices, etc. La imagen de rango será calculada utilizando la codificación facetas-vértices asignando como paleta para el tono de gris la coordenada x de los ejes cartesianos. Con eso ya tenemos una imagen para la cual cada tono indicará la distancia del objeto 3D a un plano de referencia dado, es decir, una imagen de rango como las que mostramos en la siguiente figura:

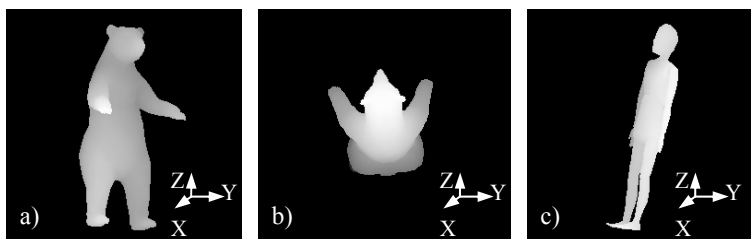


Figura 3.15 Diferentes imágenes de rango, a) y b) representan el objeto “oso” rotado (20,12,10) y (5,80,180) y c) el objeto “test 1” rotado (10,85,34)

Una vez generadas las imágenes de rango en forma digital, la obtención de las PhFTs es trivial y sólo implica el codificar en una fase la imagen y calcular su transformada de Fourier mediante una FFT. La figura siguiente muestra las PhFTs correspondientes a las imágenes de rango de la Figura 3.15:

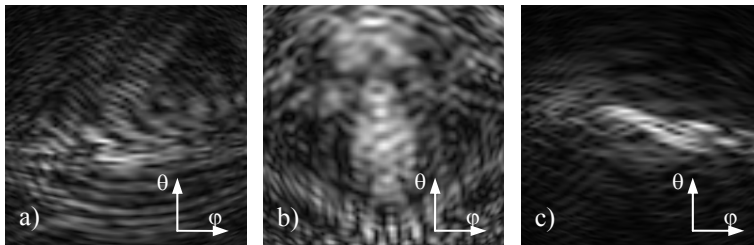


Figura 3.16 PhFTs correspondientes a las imágenes de rango de la Figura 3.15

Siguiendo el esquema de la Figura 3.14, hemos de construir el 3DOOM del objeto que queremos reconocer, es decir, del objeto que tomemos como referencia, por lo que mostraremos a continuación el proceso de generación del 3DOOM.

Como ya explicamos anteriormente, el 3DOOM es un parcheado de PhFTs de un objeto desde diferentes vistas. Así que para calcularlo seguiremos el proceso de rotar el objeto, calcular PhFT y pegar la PhFT en la zona del futuro 3DOOM que le corresponda. Después, haremos una nueva rotación y repetiremos hasta tener un muestreo apropiado de todo S^2 . En las siguientes figuras ilustramos esta operación:

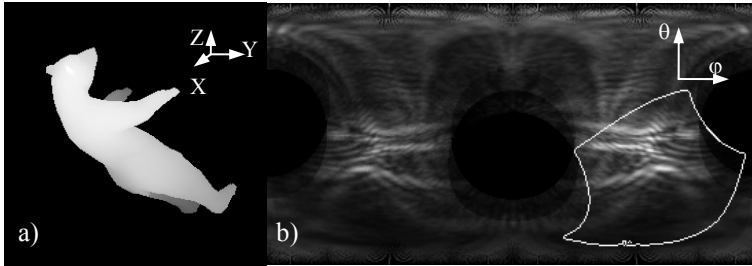


Figura 3.17 El proceso de construcción del 3DOOM en un estado intermedio, la figura a) muestra una imagen de rango y b) 3DOOM que ya ha sido completado hasta llegar ahí.

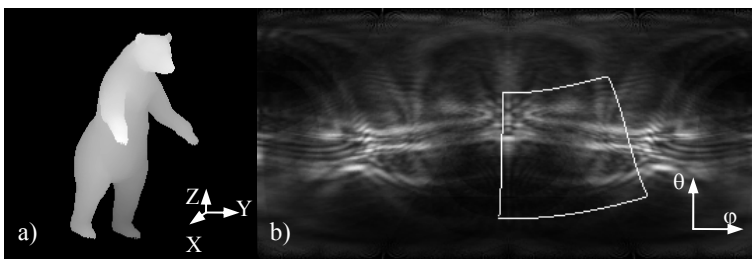


Figura 3.18 Ídem de 3.17, pero en un estado mucho más avanzado.

El proceso iterativo de rotaciones, cálculos de PhFTs y pegados continúa hasta completar el 3DOOM, la figura siguiente muestra el resultado para dos objetos diferentes:

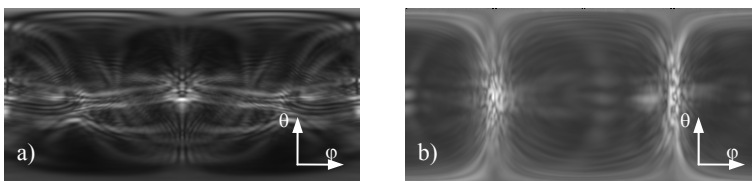


Figura 3.19 Dos 3DOOMs completos, a) se corresponde al objeto “oso”, y b) al objeto “test 1” mostrado en la Figura 3.15 c)

Una vez que ya hemos visto como construir PhFTs y 3DOOMs el siguiente paso a resolver es el cálculo de la correlación en S^2 . Para ello usaremos la ecuación (3.35), lo que implica el cálculo de las transformadas de Fourier en

S^2 , esto es, desarrollar las funciones, $f(\varphi, \theta)$ y $g(\varphi, \theta)$, en términos de los armónicos esféricos.

3.4.2 Desarrollo de funciones en S^2

Normalmente no es factible trabajar con los infinitos términos de la ecuación (3.35), sino que se ha de hacer un desarrollo finito. Por tanto, tenemos que averiguar hasta que valor de j se ha de hacer el desarrollo de las funciones para obtener una reconstrucción apropiada. Para estimar ese valor es necesario estudiar el comportamiento de los armónicos esféricos definidos en la Ec. (3.19). El análisis de esa ecuación nos lleva a analizar el comportamiento de los polinomios asociados de Legendre $P_j^m(\cos(\theta))$ [Mathworld-99b]. Dichos polinomios presentan un serie de ceros, cuya cantidad es igual al orden j . Esto implica que existen un número de lóbulos paralelos al ecuador igual a $j+1$, como muestra la Figura 3.12 [Vallés-04].

Como un lóbulo no es más que un máximo rodeado de una zona de mínimo, si usamos estos polinomios para hacer un desarrollo podemos estimar la resolución angular del desarrollo como $2\pi/j$, ya que cuantos más lóbulos, más detalles finos podremos reproducir. Teniendo en cuenta que, en este caso particular, el 3DOOM se representa por una imagen (φ, θ) de 359x180 píxeles, se llega la conclusión de que un desarrollo hasta $j_{\max} \approx 100$ será difícilmente distinguible de la función original. La Figura 3.20 muestra diferentes reconstrucciones del 3DOOM del objeto, tomando para cada una distintos valores de j_{\max} , y la Figura 3.21 las de una PhFT:

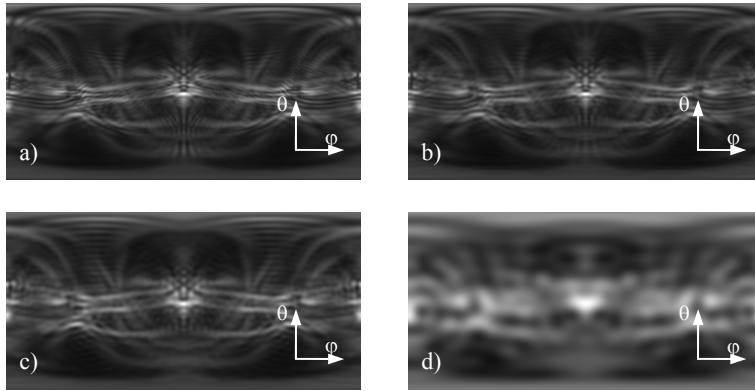


Figura 3.20 Reconstrucción del 3DOOM del objeto “oso”, para $j_{\max}=120, 80, 60$ y 20 , en a), b), c) y d) respectivamente.

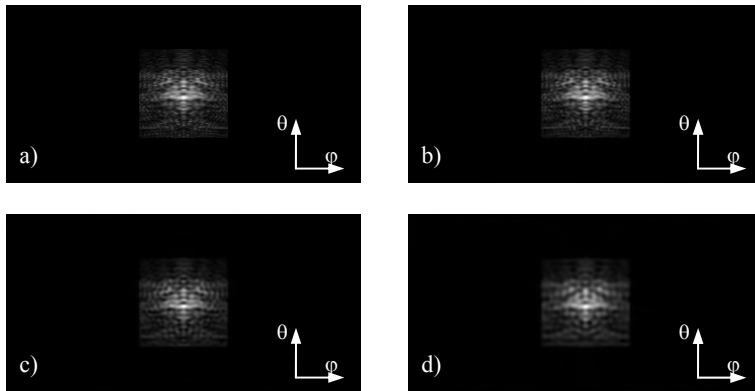


Figura 3.21 En a) PhFT del objeto “oso” sin rotar, y en b), c) y d) reconstrucciones a $j_{\max}=120, 80$ y 60 , respectivamente.

Como se puede ver en las figuras anteriores, las versiones reconstruidas hasta $j_{\max} = 120$, son prácticamente indistinguibles del mapa, mostrado en la Figura 3.19 a), y la PhFT real. Asimismo es también muy difícil distinguir a simple vista la reconstrucción con $j_{\max} = 80$, por lo que ese será el valor con el que trabajemos, buscando un equilibrio entre tiempo y precisión de cálculo.

3.4.3 Cálculo de correlaciones en SO(3)

El último paso de la implementación del método es la propia correlación, donde la mayor dificultad se encuentra en las matrices de Wigner, ya que el resto de la ecuación (3.35) es un sumatorio.

En el apartado previo, vimos una expresión directa para calcular las matrices de Wigner, pero ya dijimos que es imposible de usar para valores de j altos. También existe una relación de recurrencia para el mismo fin, la ecuación (3.34). Con esa relación de recurrencia podemos calcular todos los elementos de la forma $d_{m,m'}^j$ para todos los valores de j ; por ejemplo calcularíamos $d_{0,0}^0, d_{0,0}^1, \dots$ y luego $d_{1,0}^1, d_{1,0}^2, \dots$, con lo que, a nivel de implementación, el sumatorio (3.35) conviene que lo ordenemos de tal forma que simplifiquemos la manera de calcular los $d_{m,m'}^j$.

El otro factor de dificultad es el orden j hasta el cual desarrollaremos nuestras funciones y cómo influye esto en la correlación obtenida. En la siguiente figura mostramos la autocorrelación del 3DOOM del objeto “oso”, eso deberá darnos un pico en $(\alpha = 0, \beta = 0, \gamma = 0)$ cada vez más nítido:

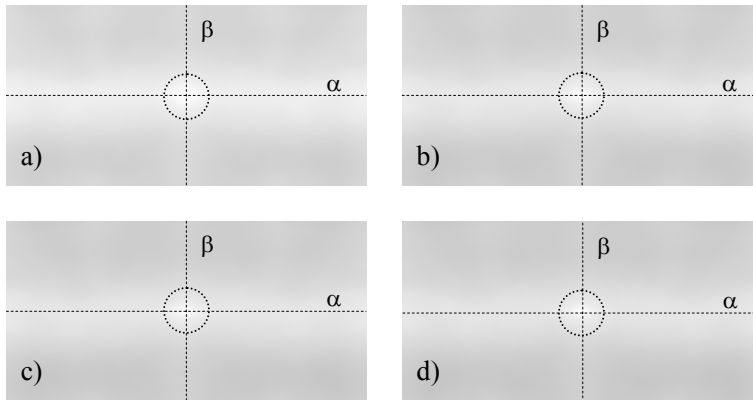


Figura 3.22 Autocorrelación del 3DOOM del “oso”, plano ($\gamma = 0$),

calculamos sólo hasta $j_{\max} = 20, 40, 60$ y 80 , para a), b), c) y d).

Para facilitar la observación, se ha añadido un círculo en torno al pico

La representación de los resultados de la correlación en función de los ángulos de Euler es tridimensional. Con el fin de facilitar la interpretación de los resultados, mostraremos únicamente los resultados en forma simplificada. En la Figura 3.22 la representación es de forma bidimensional, es decir en función de (α, β) así consideraremos que el plano $\gamma = cte$, es cero en este caso particular. Además, en el desarrollo teórico hemos dicho que $\alpha \in [-\pi, \pi)$ y $\beta \in [-\pi, \pi]$, así que en las figuras de este capítulo elegiremos el origen de α en la vertical intermedia de la correlación, y el origen β en la línea horizontal intermedia de la correlación para mejorar así la visibilidad de las figuras. Siguiendo este criterio en la Figura 3.22 observamos un pico de correlación en el centro, tal y como era de esperar, aunque la respuesta es un tanto pobre ya que si bien el valor máximo es alto, el valor promedio de la respuesta es cercano a él, dificultando el reconocimiento y también la interpretación de los resultados. Por ello, en la figura siguiente se muestra un perfil, con respecto al eje $\beta = 0$, de la autocorrelación para el caso con $j_{\max} = 80$:

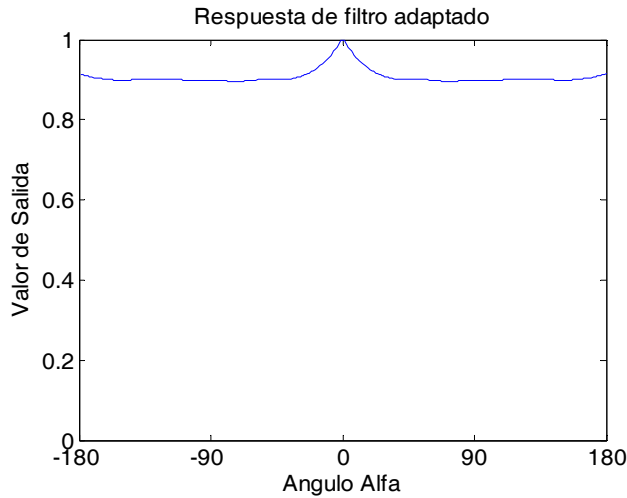


Figura 3.23 Perfil de la autocorrelación del 3DOOM para el caso $j_{\max}=80$.

Para mejorar este resultado podemos aplicar técnicas similares a las del reconocimiento en \mathbb{R}^2 , concretamente se ha usado el filtrado sólo de fase, que se describe a continuación.

Si existe un filtro de fase en \mathbb{R}^2 , podemos diseñar uno similar en S^2 , el cual esperamos que mejore la respuesta de la correlación. En el caso de un filtro de fase, POF¹, en \mathbb{R}^2 lo que hacemos es sustituir en el filtro adaptado $F^*\{f\}$ por su fase, [Horner-84] aumentando así la discriminación y mejorando la forma del pico de correlación. En el caso de S^2 vamos a hacer lo mismo, sustituir en la expresión (3.35) $F_{j,m'}$ por su fase. En las siguientes figuras mostramos los resultados de un POF variando el j máximo para el que desarrollamos y un perfil para la autocorrelación con $j_{\max}=80$, respectivamente:

¹ Phase only Filter, en inglés.

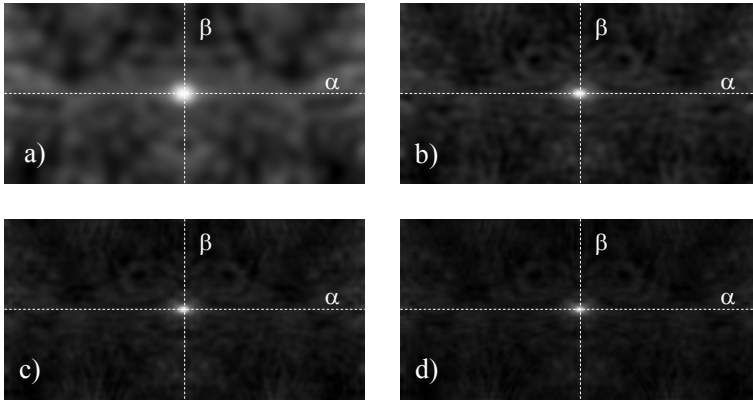


Figura 3.24 La misma figura que en 3.22, pero ahora con un filtro de fase, a), b), c) y d) son las correlaciones hasta $j_{\max}=20, 40, 60$ y 80 respectivamente

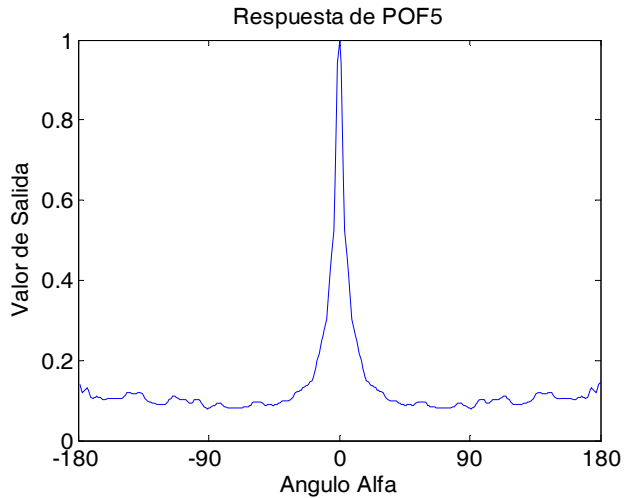


Figura 3.25 Perfil de la autocorrelación del 3DOOM, con un filtro de fase y para el caso $j_{\max}=80$.

Como observamos en ambas figuras, la mejora del pico de correlación es notable. Con respecto al orden de desarrollo, de la Figura 3.24 llegamos a la misma conclusión obtenida en el desarrollo de funciones. Usar $j_{\max}=80$ ofrece un buen equilibrio entre tiempo de cálculo y precisión.

3.4.4 Resultados de la correlación en S^2

Para empezar, veremos como se puede usar la correlación en S^2 en la implementación de filtrajes POF aplicada al reconocimiento de objetos 3D rotados. Las Figuras 3.26 a la 3.28 proporcionan un grupo de resultados:

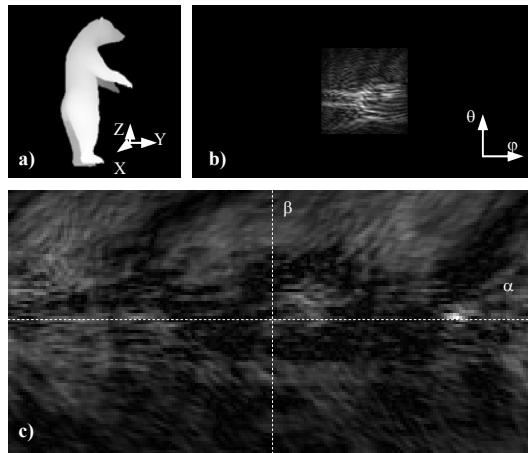


Figura 3.26 Objeto rotado (78,5,24) en a), su PhFT en b)
y el plano ($\gamma=24$) de la correlación ($j_{\max}=80$) en c)

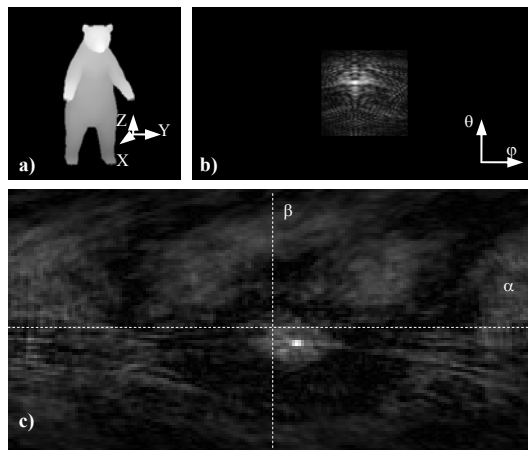
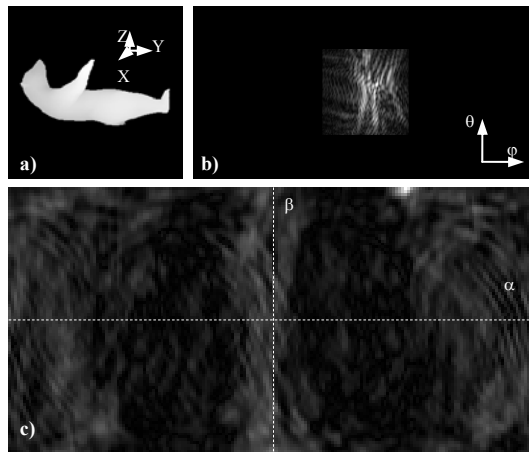


Figura 3.27 Objeto rotado (3,-15,7) en a), su PhFT en b)
y el plano ($\gamma=7$) de la correlación ($j_{\max}=80$) en c)



**Figura 3.28 Objeto rotado (0,90,90) en a), su PhFT en b)
y el plano ($\gamma=90$) de la correlación ($j_{\max}=80$) en c)**

Recordemos, como hemos explicado antes, que sólo representamos planos definidos por un valor de $\gamma=\text{cte}$ correspondientes al volumen 3D total que es la correlación. De las figuras anteriores se observa que el método es capaz de reconocer el objeto y estimar su orientación de forma correcta.

A continuación vamos a estudiar la capacidad de discriminación del método cuando se introduce un objeto 3D falso y se realiza la correlación utilizando el 3DOOM anterior. Las Figuras 3.29 a 3.31 muestran los resultados de discriminación entre un objeto y otros tres diferentes:

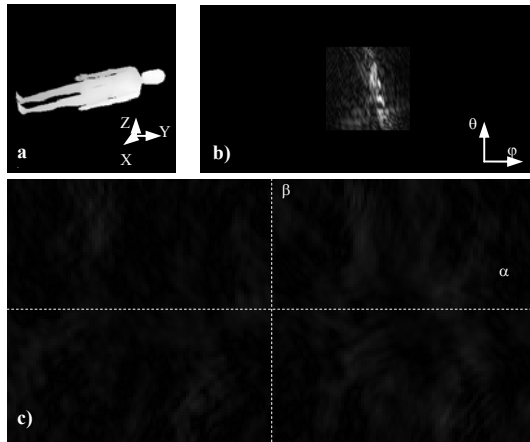


Figura 3.29 Objeto “test 1” en a) , girado (67,34,12),

la PhFT correspondiente a la imagen de rango a),

y la correlación ($j_{\max}=80$) con el 3DOOM del “oso” para el plano ($\alpha,\beta,\gamma=12$) en c).

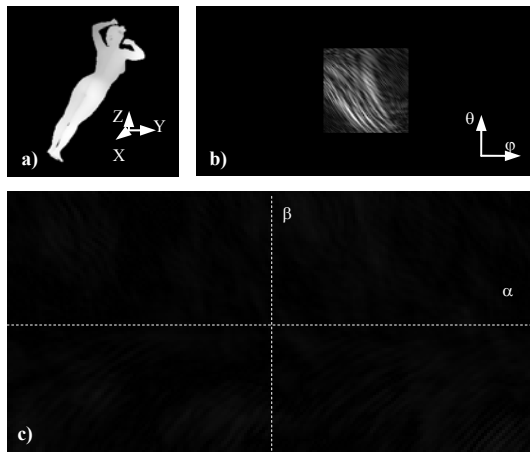


Figura 3.30 Objeto “test 2” en a) , girado (29,85,24),

la PhFT correspondiente a la imagen de rango a),

y la correlación ($j_{\max}=80$) con el 3DOOM del “oso” para el plano ($\alpha,\beta,\gamma=24$) en c)

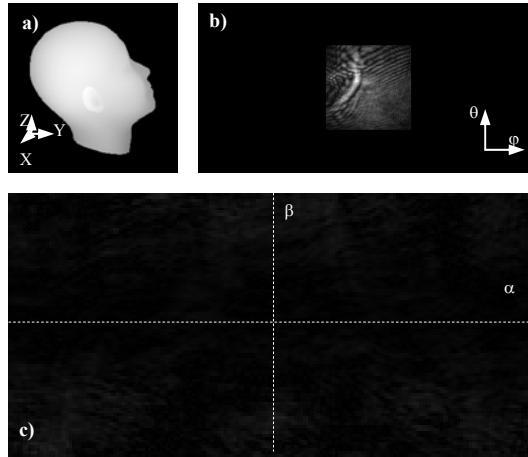


Figura 3.31 Objeto “cara 2” en a) , girado $(0,0,90)$,

la PhFT correspondiente a la imagen de rango a),

y la correlación ($j_{\max}=80$) con el 3DOOM del “oso” para el plano $(\alpha,\beta,\gamma=90)$ en c)

Para representar correctamente estos resultados de discriminación, el valor de la correlación ha sido normalizado con la obtenida para el objeto correcto. Se observa que la respuesta de todo el plano de correlación es prácticamente nula y, por tanto, el método propuesto es capaz de discriminar de forma satisfactoria entre objetos 3D diferentes que se encuentran en diferente orientación.

Para concluir con este apartado de resultados mencionaremos alguna dificultad del método que proponemos debida a la caracterización de las rotaciones utilizando los ángulos de Euler. Hay que tener en cuenta que éstos no son unívocos, y para el caso de que la segunda rotación sea nula $(\alpha,\beta=0,\gamma)$ hay infinitos tríos de Euler que dan lugar a esa misma rotación, $(\alpha + \lambda,\beta=0,\gamma - \lambda)$ para cualquier λ . Por tanto para cada plano de $\gamma = cte$ tenemos un pico en una posición diferente, de hecho tal pico no existe, sino que es un cilindro en el espacio (α,β,γ) , por eso al cortarlo en diferentes

plano de $\gamma = cte$ obtenemos varios picos. En la Figura 3.32 mostramos un objeto sometido a una de las rotaciones problemáticas, y en la 3.33 mostramos los resultados de la correlación para diferentes planos de $\gamma = cte$.

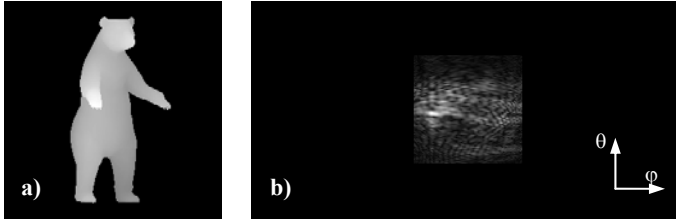


Figura 3.32 Objeto rotado (20,0,10) en a), y su PhFT en b).

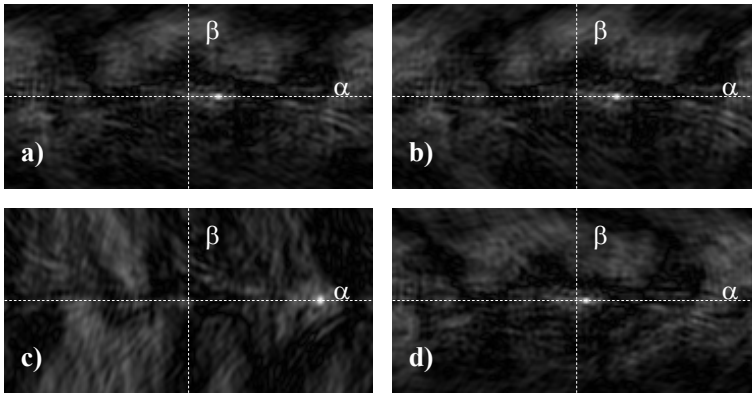


Figura 3.33 Correlación ($j_{max}=60$) entre la PhFT mostrada en la Figura 3.30 y el 3DOOM, para 4 planos de de $\gamma=0,10,100,240$ respectivamente

En la Figura 3.33 se puede observar claramente el desplazamiento del pico en función del plano que consideremos. Este es un problema de base del método que no tiene solución sencilla. Pero no es demasiado grave ya que el objeto se sigue detectando y aún cuando se detecta para diferentes conjuntos de ángulos de Euler, todos ellos son correctos. El método no obtiene falsos positivos, solamente hay una indeterminación a la hora de caracterizar la salida, lo cual puede afectar a la estimación del ángulo al que ha sido rotado.

3.5 Conclusiones

En este capítulo hemos definido un método de reconocimiento de objetos 3D invariante a rotaciones en todo el espacio tridimensional según dos etapas: una codificación de la información tridimensional en una distribución al que se ha denominado 3DOOM y la definición de correlaciones en S^2 para la detección. Hemos conseguido reducir la dimensionalidad del problema de seis, tres espaciales más tres ángulos de un planteamiento directo, a tres. La codificación considerada utiliza la mayor parte de la información geométrica del objeto, dejando sin usar sólo la configuración interna, una limitación no demasiado grave en la mayor parte de las aplicaciones. Además hemos conservado la invariancia a traslaciones y el método no sólo reconoce objetos con invariancia a rotaciones y discrimina entre diferentes objetos, sino que nos permite estimar la rotación que ha sufrido éste.

Los resultados obtenidos con este método aparecen en [García-03].

Capítulo 4

Reconocimiento de objetos tridimensionales con invariancia a cambios de escala

4.1 Introducción

En este apartado veremos cómo una de las propiedades de la PhFT permitirá reducir el problema del reconocimiento de objetos 3D con invariancia a escala a un proceso de reconocimiento de objetos 2D con invariancia a cambios de intensidad. Estos cambios de iluminación o intensidad fueron tratados en el capítulo 2 del trabajo y consistían en averiguar la pertenencia o no de un determinado vector a un subespacio dado, utilizando la aplicación de un filtro al que denominamos LACIF. Aplicaremos todas estas ideas al caso de objetos 3D que sufren distintos cambios de escala.

Este capítulo presenta ciertas diferencias frente al resto de los tratados hasta ahora. En el caso de rotaciones (capítulo 3), el total de la energía asociada al objeto que queremos reconocer permanece constante. También es así, en primera aproximación, para el caso de invariancia a iluminaciones (capítulo 2) siempre que no cambiemos la intensidad de la fuente luminosa. Por contra, un cambio de escala siempre variará la integral de la intensidad de la imagen del objeto. En el caso bidimensional, la energía varía con el cuadrado del factor de escala, dificultando el reconocimiento general mediante un filtro adaptado convencional.

Existe una amplia bibliografía que aborda estos problemas en el caso de objetos 2D. Ferreira *et al.*, [Ferreira-89], propusieron el uso de correladores

anamórficos múltiples adaptados a un objeto a diferentes escalas. Mediante el uso de partes móviles en el sistema óptico, se puede obtener una respuesta invariante a los cambios de escala.

Frente a este ejemplo, se han propuesto otros métodos que no implican la existencia de partes móviles en el montaje. Así, otro camino seguido es el de usar transformaciones matemáticas sobre la imagen del objeto o usar desarrollos en series de funciones para adaptar filtros a parte de la información y obtener así invariancia a escala. Estos métodos son análogos a la utilización de la descomposición en armónicos circulares para detección de objetos bidimensionales con invariancias a rotaciones.

Uno de los primeros trabajos sobre este tema fue desarrollado por Casasent y Psaltis, [Casasent-76b], que aplicaron la transformada de Mellin [Bracewell-63] al caso de invariancia a escala. Este método se basa en una transformación previa desde las coordenadas originales (x, y) a otras del tipo $(\ln r, \theta)$, donde r y θ son las coordenadas polares. Tras esto se aplica la transformada de Mellin, consiguiendo el objetivo de reconocer de forma invariante a escala pero a costa de perder la invariancia a traslaciones.

Para solventar esta última dificultad, Mendlovic *et al.*, [Mendlovic-88], propusieron usar un desarrollo en armónicos radiales de Mellin, que son funciones de la forma $\left\{ r^{i2\pi M-1} \right\}_{M \in \mathbb{Z}}$. Con este desarrollo, adaptando un filtro convencional a un término del desarrollo en este conjunto de funciones, se puede obtener invariancia a escala, sin perder la invariancia a traslaciones.

De forma similar al caso de los armónicos circulares, en el uso de las funciones de Mellin es crítica la definición del origen de coordenadas que se use en el desarrollo, por ello existen en ambos casos métodos similares para

la obtención de dicho centro, como el propuesto por Moya *et al.* en [Moya-93], basado en un algoritmo iterativo, donde se maximiza el valor de pico de energía obtenida en el desarrollo.

Otra dificultad asociada al uso de armónicos radiales de Mellin es su difícil generación. Aunque es posible generarlos de forma óptica, [Konforti-90], se suelen construir mediante holografía por ordenador como en [Lohmann-67]. Por ello, Tajahuerce *et al.*, [Tajahuerce-94], propusieron un filtro real basado en los armónicos radiales de Mellin, con lo que las dificultades de generar estos se reduce y se obtiene reconocimiento invariante a cambios de escala y traslaciones.

Este método guarda similitudes con el propuesto por Szoplik y Arsenault [Szoplik-85a, Szoplik-85b], que obtuvieron invariancia a escala, si bien sólo parcial, usando un desarrollo en armónicos circulares que sirve como entrada a un correlador anamórfico, un correlador óptico en el que el aumento en los ejes es diferente.

Los métodos que usan armónicos radiales de Mellin suelen adolecer de variaciones de energía en los objetos con diferentes escalas, circunstancia parcialmente solventada por Rosen *et al.*, [Rosen-89], modificando la definición de dichos armónicos de forma oportuna, pero perdiendo la ortogonalidad del conjunto en el proceso.

También en dos dimensiones, García *et al.*, [García-92], propusieron un filtro sólo de fase que se obtiene mediante una transformación de coordenadas elípticas. Con ello se aprovecha de forma más útil la información del espectro de Fourier, obteniendo invariancia a traslaciones y a un rango limitado de variaciones de escala.

Por su parte Cojoc *et al.*, [Cojoc-97], propusieron un método basado en la obtención de un filtro sólo de fase del objeto a reconocer para una escala dada. Partiendo de éste, se define un origen de ángulos y luego se calcula la integral de la energía de dicho filtro entre el origen y un ángulo dado. Es decir, se calcula la distribución angular acumulada de energía que le corresponde al filtro inicial y a partir de esta distribución, se puede obtener otro filtro que sí ofrece invariancia a cambios de escala y traslaciones.

El uso de descriptores propuesto por Sheng *et al.*, [Sheng-86a, Sheng-86b], constituye otra alternativa y consiste en asociar al objeto a reconocer un conjunto de funciones características, llamadas descriptores. Estos se definen de tal manera que su variación frente a una alteración del objeto sea bien conocida. Concretamente, Sheng propuso unos descriptores de la forma

$$M_{s,l} = \int_0^{\infty} \int_{-\pi}^{\pi} r^{s-1} g(r, \theta) \exp(-il\theta) dr d\theta, \text{ donde } g(r, \theta) \text{ es el objeto a reconocer}$$

escrito en coordenadas polares. Si el objeto cambia de escala, los $M_{s,l}$ normalizados no cambian, obteniéndose la invariancia.

Hasta ahora, todas las alternativas expuestas tiene en común el estar enfocadas al caso bidimensional. Un método que sí está adaptado al caso de objetos tridimensionales, fue el planteado por Esteve-Taboada *et al.*, [Esteve-Taboada-00], en la que la invariancia a escala se consigue usando un montaje óptico adaptado a una determinada escala, con luz policromática. Usando el hecho de que a diferentes longitudes de onda, se tienen diferentes aumentos del sistema, se consigue invariancia a escala mediante el multiplexado en longitud de onda.

Li y Rosen, [Li-03], propusieron un método basado en aplicar el desarrollo de armónicos radiales de Mellin a un conjunto de imágenes diferentes de un

mismo objeto 3D, con lo que obtuvieron un reconocimiento invariante a escalas de un objeto tridimensional.

Otra posibilidad es la propuesta por Paquet *et al.* en [Paquet-95], donde se toma como información del objeto PhFTs de imágenes de rango del mismo. Mediante el uso de redes neuronales se obtiene invariancia a cambios de escala, rotaciones y desplazamientos.

El uso de redes neuronales para la obtención de invariancia a escala en el caso 3D ha sido ampliamente usado por diversos autores. Por ejemplo, Dotsenko, [Dotsenko-88], propone la extracción de características de objetos y la posterior aplicación de redes neuronales a fin de obtener reconocimiento invariante a cambios de escala, traslaciones y rotaciones.

Similar es la propuesta de Perantonis *et al.*, [Perantonis-92], cuyas principales diferencias están en el tipo de red neuronal usada y la entrada con que se alimenta ésta, clasificadores en lugar de características. Con ello, se consigue reconocimiento 2D invariante a escala, rotación y traslaciones.

Por último, un subcampo de interés práctico en el reconocimiento usando redes neuronales se encuentra en los OCR¹, como por ejemplo en [Paulpandian-93], donde se propone un método capaz de reconocer escritura a mano, incluyendo cambios de escala en este problema.

Frente a todo lo anterior, en este capítulo proponemos un método para detectar objetos tridimensionales con invariancia a escala codificando la imagen de rango según la PhFT definida en el capítulo 3. Estudiaremos con detalle las propiedades de la PhFT cuando se somete la imagen de rango a un

¹ Optical Character Recognition, reconocimiento óptico de caracteres.

cambio de escala y veremos como el problema de la detección tridimensional pasará a ser un problema de detección bidimensional.

4.2 Cambios de escala y PhFT

Comenzaremos analizando cómo afecta un cambio de escala a la PhFT de una imagen de rango. Del capítulo 3 se recupera el hecho de que la superficie externa de un objeto 3D se puede entender como compuesta, a nivel infinitesimal, por un conjunto de facetas planas, cada una de ellas con su orientación y forma. En ese mismo capítulo se analizó cómo dada una faceta de la forma $f(y, z) = (Ay + Bz + C) \cdot w(y, z)$, donde, recordemos, $Ay + Bz + C$ es la expresión de un plano que define la orientación de la faceta y $w(y, z)$ es un factor de forma que representa una determinada región de interés en el entramado de la distribución 3D, la PhFT de la distribución $f(y, z)$ era igual a:

$$PhFT(u, v) = e^{ic} \delta \left(u - \frac{a}{2\pi}, v - \frac{b}{2\pi} \right) \otimes W(u, v) \quad (4.1)$$

Ahora consideremos un cambio de escala, con factor de escala k , en el objeto $f(y, z)$. Esto afectará a cada una de las facetas de la forma:

$$f'(y, z) = (Ay + Bz + C) \cdot w(y/k, z/k) \quad (4.2)$$

Hacemos notar que la orientación de la faceta, dada por el término $Ay + Bz + C$, no varía pero sí que lo hará el factor de forma $w(y/k, z/k)$. Además nótese que un cambio de escala de factor k , por ejemplo para

$k = 0.8$, indica que se tiene una versión del objeto cuyo tamaño es el 80% del original.

Si ahora calculamos la PhFT de la faceta escalada obtenemos:

$$\begin{aligned} PhFT_{f'}(u,v) &= e^{ic} \delta\left(u - \frac{A}{2\pi}, v - \frac{B}{2\pi}\right) \otimes FT\{w(y/k, z/k)\} = \\ &= k^2 e^{ic} \delta\left(u - \frac{A}{2\pi}, v - \frac{B}{2\pi}\right) \otimes W(ku, kv) \end{aligned} \quad (4.3)$$

Nótese que hemos indicado con un subíndice que se trata de la PhFT del objeto escalado. El pico asociado a la faceta escalada está centrado en el mismo punto que el de la versión sin escalar, pero convolucionado con un factor de forma escalado. En la siguiente figura mostramos la evolución del valor de pico de la PhFT para una faceta cuadrada en la que vamos variando el factor de escala:

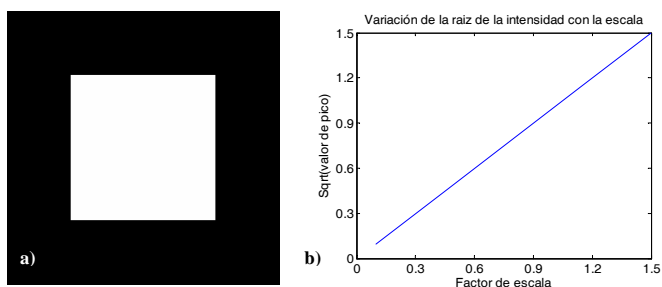


Figura 4.1 En a) mostramos el objeto del que calculamos su PhFT, en b) mostramos la evolución de la raíz cuadrada del pico normalizado en función de la escala

Se observa que la relación encontrada en la ecuación (4.3) se cumple de forma exacta, la relación del pico de PhFT de una faceta y la escala es cuadrática de forma exacta.

Consideremos ahora una imagen de rango sencilla, una pirámide compuesta por cuatro facetas. En la siguiente figura mostramos cuál es el efecto del cambio de escala sobre la PhFT de dicho objeto a diferentes escalas:

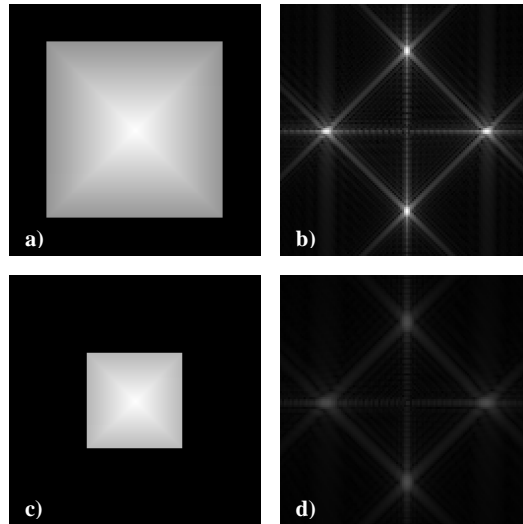


Figura 4.2 En a) mostramos una faceta a su escala original y en b) su PhFT. En c) y d) mostramos el objeto cambiado de escala y su PhFT respectivamente.

Se comprueba que estas cuatro facetas del objeto se corresponden con cuatro picos en el plano de Fourier. Si la curvatura de la superficie del objeto fuese continua (un objeto suave) la PhFT sería continua en lugar de estar centrada en picos discretos.

Ahora centrémonos en el efecto del cambio de escala en la PhFT del objeto; de la Figura 4.2 se observa que la forma general de ésta es la misma, salvo un factor multiplicativo en la intensidad. Es decir, la PhFT del objeto escalado se puede aproximar de la forma siguiente:

$$PhFT_f(u, v) \cong k^2 e^{ic} \delta\left(u - \frac{A}{2\pi}, v - \frac{B}{2\pi}\right) \otimes W(u, v) = k^2 PhFT_f(u, v) \quad (4.4)$$

Es decir, tenemos una relación cuadrática entre escala y amplitud de la salida, pero sólo en primera aproximación. De la ecuación (4.4) podemos concluir que un cambio de escala en la PhFT implica un cambio de intensidad según el cuadrado del factor de escala. De esta forma, podemos aplicar aquí los algoritmos de detección relacionados con la invariancia a iluminaciones o a intensidad que explicábamos en el capítulo 2.

Sin embargo, esta relación en primera aproximación entre las PhFT de las imágenes de rango originales y las escaladas no es tan directa cuando el objeto 3D posee un entramado complejo y con facetas que suelen tener un tamaño bastante reducido. Así, dado que la implementación que haremos será de tipo digital, no será posible trabajar con facetas de tamaño arbitrariamente pequeño. De hecho, al hacer la PhFT de un objeto complejo a bajas escalas, perderemos parte de la información con menor intensidad. Esto es similar a un filtraje de Fourier pasa-baja. En la siguiente figura mostramos el efecto descrito anteriormente para una imagen de rango más compleja:

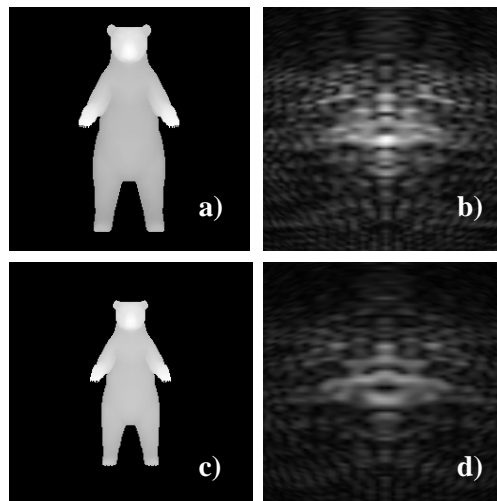


Figura 4.3 Imagen de rango de un objeto en a) y su PhFT en b) mostramos una versión escalada de a) en c) y su PhFT en d)

Se aprecia cómo la forma general de ambas PhFTs es similar, salvo un por un factor multiplicativo constante y por la pérdida de parte de la información, tal y como habíamos argumentado. Con esto, dentro de un cierto margen de aproximación, es posible reducir el problema del reconocimiento invariante a escalas al de reconocimiento con invariancia a intensidad.

4.3 Métodos vectoriales para el reconocimiento invariante a escala

4.3.1 LACIF aplicado a invariancia a escala

Dado que el problema de la invariancia a escala ha quedado reducido a un problema de invariancia a intensidad, vamos a aplicar el filtraje LACIF, [Lefebvre-02], [Lefebvre-03], [Roy-04] y [Arsenault-05], que utilizamos en el capítulo 2. El estudio realizado en ese capítulo describía la aplicación del filtraje al caso de imágenes que sufrían alteraciones en la intensidad o iluminación. Nótese que ahora, las imágenes de referencia y de entrada que utilizaremos en los filtros vienen dadas no por imágenes convencional sino por las PhFT de imágenes de rango, ya que son éstas las que están afectadas de los cambios de iluminación.

En nuestro caso la PhFT del objeto original se puede escribir como:

$$PhFT_{x'}(u,v) = \alpha PhFT_x(u,v) + \beta v_0(u,v) \quad (4.5)$$

donde $v_0(u,v)$ es el soporte de la PhFT del objeto que queremos reconocer, dado por $PhFT_x(u,v)$, y α y β son dos constantes arbitrarias. Con esto definimos $PhFT_x^o(u,v)$ como:

$$PhFT_x^o(u, v) = PhFT_x(u, v) - \mu_f v_0(u, v) \quad (4.6)$$

Nótese que $PhFT_x^o(u, v)$ es la versión de media cero de la PhFT original. Con esto el conjunto base formado por los vectores o imágenes $\{PhFT_x^o(u, v), v_0(u, v)\}$ se puede ortonormalizar para aplicar el filtraje LACIF, teniendo como base ortonormal:

$$\{\phi_0(u, v) = v_0(u, v) / \|v_0(u, v)\|; \phi_1(u, v) = PhFT_x^o(u, v) / \|PhFT_x^o(u, v)\|\} \quad (4.7)$$

donde la norma de los vectores está definida como en el capítulo 2, siendo:

$$\|PhFT_x^o(u, v)\| = \sqrt{(PhFT_x^o * PhFT_x^o)(0, 0)} \quad \text{y} \quad \|v_0(u, v)\| = \sqrt{(v_0 * v_0)(0, 0)} .$$

Con esta nueva base $\{\phi_0(u, v), \phi_1(u, v)\}$ ya podemos aplicar el filtraje LACIF a nuestro caso, quedando:

$$LACIF_{PhFT_s, \phi}(u, v) = \frac{(PhFT_s * \phi_0)^2(u, v)}{\sqrt{N} (PhFT_s^2 * \phi_1)(u, v) - (PhFT_s * \phi_1)^2} \quad (4.8)$$

siendo $PhFT_s(u, v)$ la PhFT de la imagen de rango de la escena de entrada y N el número de píxeles del soporte $v_0(u, v)$.

Si ahora aplicamos el LACIF a una escena dada y ésta es de la forma:

$$PhFT_x(u, v) = \alpha' \cdot \phi_0(u, v) + \beta' \cdot \phi_1(u, v) \quad (4.9)$$

donde α' y β' son dos constantes arbitrarias. La respuesta de la ecuación (4.8) cuando utilizamos como escena de entrada la ecuación (4.9), será igual a la unidad cuando el objeto es el verdadero y menor en cualquier otro caso. Nótese que la conexión existente entre el factor de escala de la imagen de rango, k , según la ecuación (4.4), y el factor de intensidad es $\alpha = k^2$. Esto no plantea ninguna dificultad, ya que el LACIF ha sido diseñado para operar con cualesquiera valores arbitrarios de α y β como se muestra en [Lefebvre-02], [Lefebvre-03] y [Arsenault-05], no siendo necesaria ninguna restricción al respecto.

En cuanto al término β , éste sería el que multiplicase a $\phi_1(u, v)$. Este vector es la versión normalizada del soporte de la PhFT. El objeto que queremos reconocer puede tener facetas en cualquier orientación, pero no todas éstas son observables desde el punto de observación. Ello determina que la PhFT no existe en un dominio de definición infinito, sino en un cuadrado en el espacio (u, v) , limitado por las orientaciones máximas, correspondientes a facetas normales al plano de observación, que podamos percibir.

Para calcular la ecuación (4.8) elegiremos una base $\{\phi_0(u, v), \phi_1(u, v)\}$. Como en la introducción hemos indicado que $PhFT_f(u, v) \cong k^2 PhFT_f(u, v)$ tomaremos como elemento $\phi_1(u, v)$ el resultado de normalizar la PhFT del objeto que queremos reconocer en una escala que tomaremos como referencia. Para el elemento de la base $\phi_0(u, v)$ hemos considerado un cuadrado homogéneo dentro del soporte de definición.

Vamos a realizar una serie de experimentos para verificar el buen funcionamiento del filtraje LACIF aplicado sobre distribuciones de PhFTs

cuando el objeto 3D se somete a cambios de escala. Para ello variaremos la escala del objeto 3D “oso” entre un factor 0.6 y 1.2, y tomaremos como base la PhFT del objeto a escala 0.9 y un soporte cuadrado. La elección del factor de escala 0.9 es debido a que ese factor es el intermedio dentro del rango de variación de escala considerado. Los resultados normalizados al valor máximo se muestran en la siguiente figura:

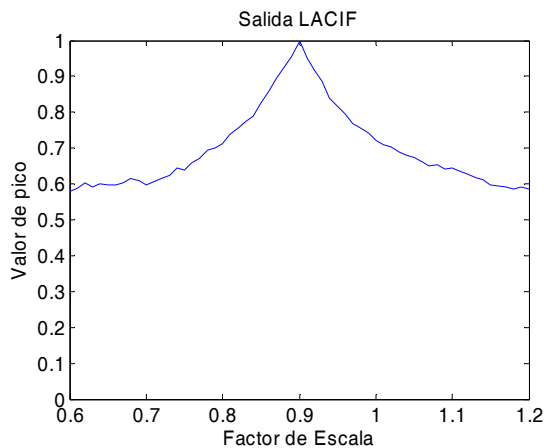


Figura 4.4 Resultado del LACIF aplicado al objeto “oso”, tomando como base la PhFT del objeto a escala 0.9

Se observa de la figura que los resultados no son especialmente buenos, ya que si bien todos los resultados quedan por encima del valor 0.6, el LACIF sólo da un valor igual a la unidad en el caso de tener el objeto con una escala exactamente igual a la de referencia. En el resto de casos la respuesta difiere de forma notable del valor máximo.

La explicación para este resultado es sencilla: hemos dicho que por efecto del muestreo, que nos impide representar facetas de tamaño muy pequeño, y de la aproximación hecha en la ecuación (4.4), sólo tenemos de forma aproximada una relación cuadrática entre escala y la intensidad de la PhFT. Dicho de forma vectorial: en el rango de escalas en que vamos a trabajar, la

base $\{\phi_0(u,v), \phi_1(u,v)\}$, con $\phi_1(u,v)$ una PhFT de referencia y $\phi_0(u,v)$ el soporte entero de la PhFT no es completa. Para mejorar estos resultados hemos de encontrar una base que represente mejor el conjunto de PhFT correspondientes a diferentes escalas. Esa base la encontraremos utilizando la transformada de Karhunen-Loeve que ya definimos y utilizamos en el capítulo 2.

4.3.2 Transformación de Karhunen-Loeve aplicada a la invariancia de escala

La KLT es una transformación de una base determinada en sus componentes principales, [Karhunen-47] y [Loeve-78]. Con el fin de obtener una base representativa de todas las posibles escalas del objeto, la KLT se aplicará sobre el conjunto escalado de PhFTs.

Tomaremos como conjunto de muestra todas las PhFTs del objeto “oso” entre las escalas 0.6 y 1.2. Tras aplicar la KLT se obtienen como resultado los siguientes autovalores:

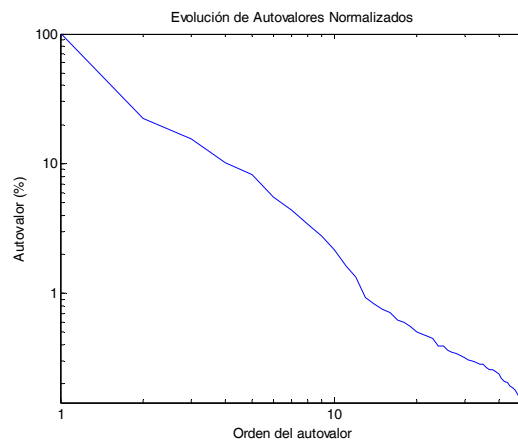


Figura 4.5 Primeros 50 autovalores asociados al objeto “oso”, rango de escalas [0.6-1.2]

Una simple comparación visual entre esta figura y la Figura 2.16 ilustra las diferencias entre ambos problemas. Mientras que en el caso de las variaciones de iluminación de objetos 3D (capítulo 2), el conjunto presentaba 4 valores que generaban casi todo el espacio, equivalentes al soporte y las 3 componentes del vector normal a la superficie, en el caso de los cambios de escala, tenemos que los 15 primeros autovectores tienen un valor apreciable y por tanto nada desestimable. Así, el hecho de perder detalles a bajas escalas y la no exactitud de la ecuación (4.4) hacen que la dimensionalidad exacta del problema sea muy alta.

De esta forma, para poder reconstruir el espacio de forma aceptable con la base generada por la KLT tendríamos que tomar del orden de 15 vectores. Esta sería una posible solución, pero en este trabajo se ha descartado por poco práctica, ya que el tiempo de cálculo sería muy elevado.

En lugar del uso del KLT proponemos una solución alternativa para la elección de la base y la mejora de resultados. Dado que el cambiar de escala el objeto, provoca que ciertos detalles finos se pierdan, sugerimos aprovechar este hecho en beneficio nuestro. Por un lado, en lugar de tomar como elemento de la base vectorial, una única PhFT del objeto a una escala dada, sugerimos tomar un promedio de las PhFTs implicadas como referencia. Con esto reduciremos el efecto de baja resolución que tiene el cambio de escala.

Por otra parte, el elemento $\phi_0(u,v)$ está relacionado con el soporte de definición de la PhFT. Si, como hemos dicho, la PhFT sufre un proceso similar a un filtraje pasa-baja o de baja resolución al cambiar de escala, tendremos que no todas las orientaciones (u,v) son igualmente importantes, habrá algunas que no aparecerán en todas las escalas. Así, usando $\phi_0(u,v)$

como una máscara binaria que no sea cuadrada es decir, una función de una forma determinada que multiplica la entrada y cuyos valores son 1 o 0, podremos filtrar parte de la información menos relevante. En concreto la máscara se obtendrá aplicando una función umbral a la PhFT de referencia.

4.4 Implementación y resultados

La implementación del método que proponemos es digital y se ha llevado a cabo usando Matlab 7.1 r13. El esquema de trabajo se representa por el siguiente diagrama:

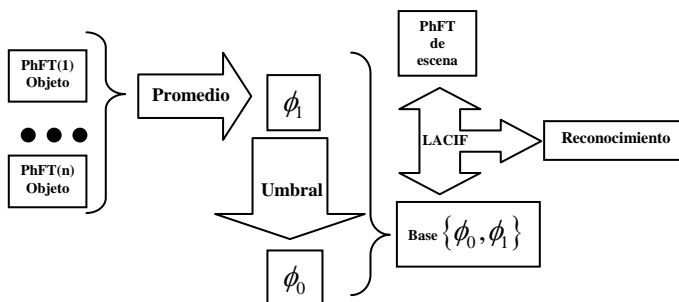


Figura 4.6 Diagrama de flujo del método

Dentro de este esquema mostraremos cuales son los pasos del método y los resultados obtenidos para detectar el objeto “oso” y frente a objetos falsos, en un rango de escala [0.6-1.2], en la Figura 4.7 se tienen los objetos que se usarán en este proceso:

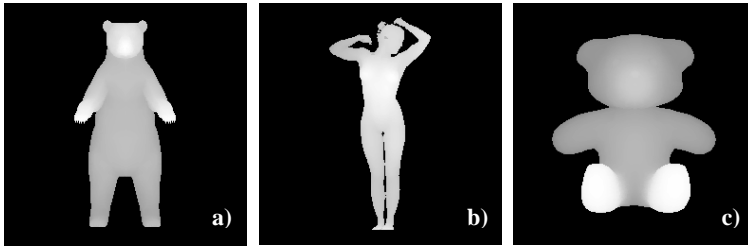


Figura 4.7 Objetos usados en la implementación digital, en a) mostramos el objeto “oso”, que será el objeto a detectar, en b) objeto “test 2” y en c) objeto “test 3”

Para generar el elemento de la base $\phi_1(u,v)$ hemos tomado el promedio de las PhFTs del objeto correcto, la máscara $\phi_0(u,v)$ se calcula tomando como criterio una umbralización de $\phi_1(u,v)$. La justificación para ello se basa en dos efectos ya explicados. El primero es que a diferentes escalas, se tienen PhFTs ligeramente diferentes, por ello $\phi_1(u,v)$ posee la contribución promedio de las PhFT a diferentes escalas. El otro efecto es que a escalas menores que la unidad, la PhFT se comporta como si hubiese pasado por un filtrado pasa-baja debido a la pérdida de resolución. Por ello tomaremos como elemento ventana $\phi_0(u,v)$, el resultado de una umbralización. De esta forma se consigue tener en cuenta, aproximadamente, la información común a las PhFTs de todo el rango de escalas. En la siguiente figura mostramos la base vectorial $\{\phi_0(u,v), \phi_1(u,v)\}$, correspondiente al objeto “oso”, que se genera usando este método:

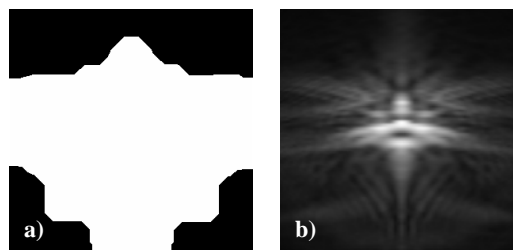
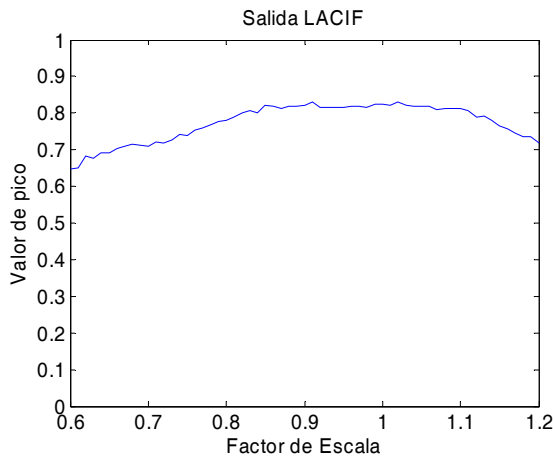


Figura 4.8 Base vectorial que usaremos en esta batería de pruebas en a) mostramos $\phi_0(u,v)$ y en b) $\phi_1(u,v)$

Nótese cómo el contenido en información de los bordes de la PhFT, correspondientes a orientaciones casi normales al plano de referencia, es de menor importancia. Algo previsible, ya que el área visible de una faceta es su proyección y esta tiende a cero para ángulos normales, por ello esta parte de la información es filtrada por la máscara ϕ_0 .

Una vez definida la base $\{\phi_0, \phi_1\}$ se muestran los resultados obtenidos con este método. El objeto 3D a detectar sigue siendo el objeto “oso”, la Figura 4.9 muestra la respuesta para este objeto:



**Figura 4.9 Detección con invariancia a escala del objeto “oso”,
rango de escalas [0.6 – 1.2]**

Nótese como el método mejora dando un aspecto más uniforme a la detección del objeto si lo comparamos con la Figura 4.4 en la que sólo una PhFT se utilizaba como elemento base. El valor promedio de detección para todas las escalas es cercano a 0.75 de pico de salida.

Ahora probaremos la capacidad de discriminación del método, tratando de detectar el objeto “oso” en escenas donde está un objeto falso. La Figura 4.10 proporciona los resultados para el objeto “test 2”, mientras que la 4.11 lo hace para el objeto “test 3”:

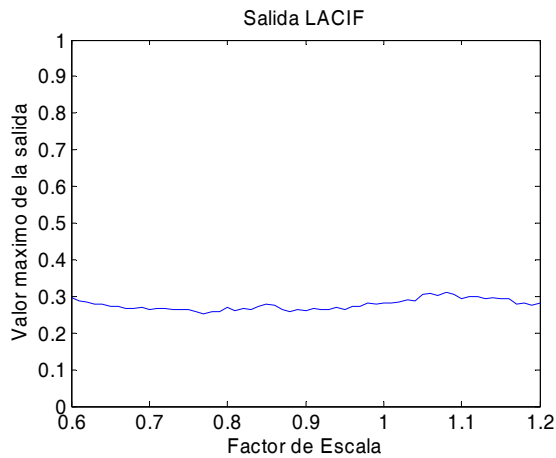


Figura 4.10 Salida del LACIF utilizando el objeto “test 2” como entrada

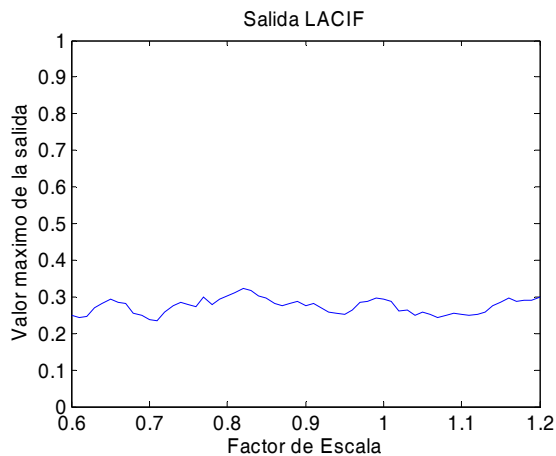


Figura 4.11 Salida del LACIF utilizando el objeto “test 3” como entrada

En ambos casos el método es capaz de discriminar entre objetos distintos a la referencia. La salida típica para un objeto incorrecto está en torno a 0.3 que es más de dos veces inferior a la obtenida con el objeto correcto.

Recordemos que en el capítulo 2, definimos una operación de reconocimiento similar al LACIF, pero que permitía una mayor capacidad de detección cuando el objeto 3D era sometido a cambios de iluminación. Este método, al que denominamos LADC, estaba basado en el cálculo del coseno del ángulo formado por el espacio vectorial generado por la base elegida y un vector que representaba la escena de entrada. Podría pensarse que, al igual que entonces, su utilización en el problema que nos ocupa podría proporcionarnos mejores resultados. Veamos, pues, su aplicación al caso de invariancia a escala. El resultado con el objeto “oso” se muestra en la siguiente figura:

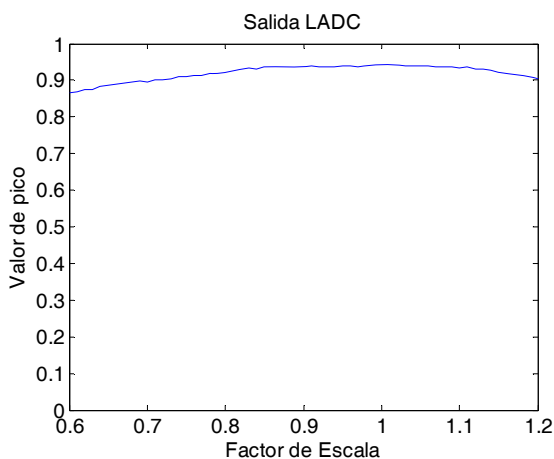


Figura 4.12 Resultado del LADC para el objeto “oso”

Como se puede observar en la figura, los resultados utilizando el LADC son mejores que los obtenidos para el LACIF. De hecho el valor promedio del pico de salida ronda el 0.90. Por el contrario, a la hora de discriminar entre el objeto a detectar y otro distinto, los resultados no son tan halagüeños. Como se pone de manifiesto en la Figura 4.13, donde puede observarse la salida utilizando el filtro LADC para el objeto “test 3”. El valor del pico de correlación no varía apreciablemente en el rango de escalas utilizado, pero es muy alto y muy próximo al valor del pico del objeto de referencia. Por ello la

capacidad de discriminación es muy baja, y, en consecuencia, se seguirá utilizando el LACIF en lo que resta de capítulo.

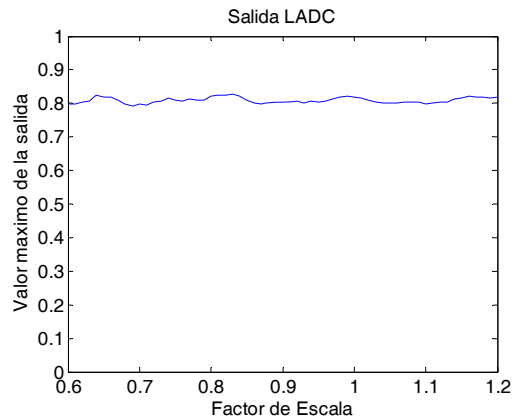


Figura 4.13 Salida del LADC utilizando el objeto “test 3” como entrada

Hasta ahora sólo estamos indicando el valor del pico de correlación. En la siguiente figura mostramos el plano de correlación completo, para observar la apariencia del mismo en cuanto a forma del pico, aparición de lóbulos laterales, etc. A su vez mostramos el plano de correlación cuando se utiliza como entrada una imagen distinta a la de referencia.

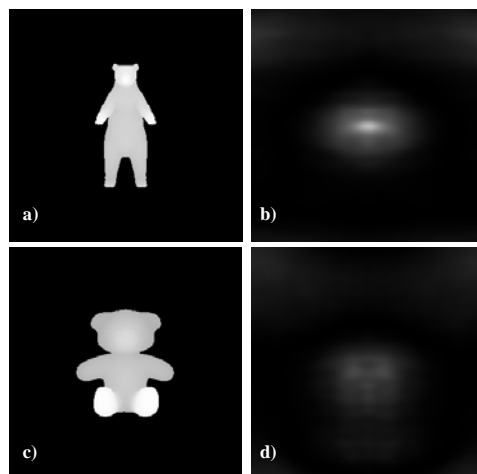


Figura 4.14 En a) mostramos el objeto correcto y en b) el resultado del LACIF en tal caso, en c) tenemos un objeto falso y la salida del LACIF en d), todos los objetos están a escala 0.75

Nótese que el pico de correlación que aparece en la Figura b) está bastante centrado y la forma es suave alrededor del valor máximo. Por otra parte el plano de correlación de la imagen a rechazar muestra un pico con una estructura más dispersa y con una intensidad muy baja.

Otra clase de objetos que hemos utilizado para comprobar el comportamiento del filtro LACIF aplicado a cambios de escala en objetos 3D es el reconocimiento de caras. Este es un tema que está recibiendo una gran atención en los últimos años debido a su gran interés práctico. En el caso 2D existen diferentes propuestas de reconocimiento de caras invariante a cambios de iluminación, como por ejemplo la de Turk *et al.*, [Turk-91a, Turk-91b], [Pentland-94], basados en aplicar una transformación de Karhunen-Loeve al espacio de posibles variaciones de la cara a reconocer.

También hay propuestas de reconocimiento de caras en tres dimensiones como muestra el artículo *review* [Kittler-05], así como diferentes métodos capaces de reconocer caras con invariancia a diferentes alteraciones como cambios de iluminación [Li-04] y [Li-07], escala [Hotta-98] y rotaciones [Chai-07].

Por ello, en las siguientes figuras mostramos otra tanda de resultados, con unos objetos diferentes, en este caso dos caras:

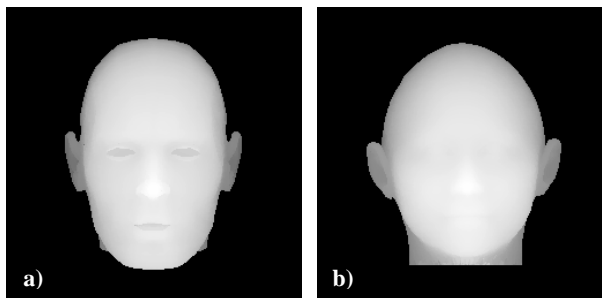


Figura 4.15 En a) mostramos el objeto a reconocer, en b) un objeto falso de prueba

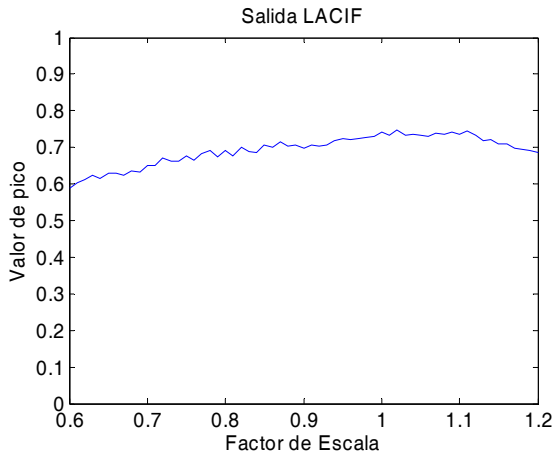


Figura 4.16 Resultados de la detección de la cara de la Figura 4.15 a)

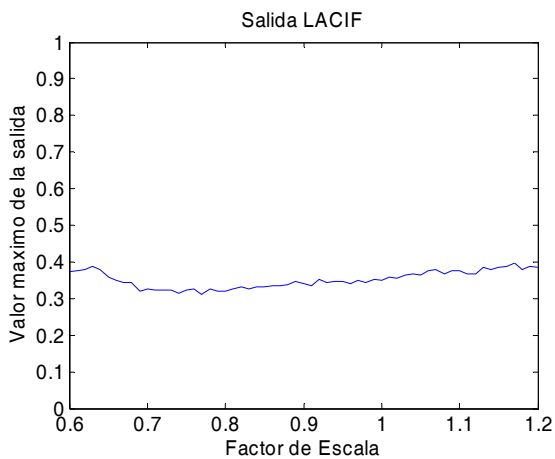


Figura 4.17 Resultados de la correlación LACIF entre las dos caras de la Figura 4.15

Nuevamente se comprueba que el método es capaz de reconocer al objeto correcto con una respuesta media en torno a 0.7, y de discriminar entre objetos, aun cuando los resultados son ligeramente inferiores a los obtenidos con la imagen 3D “oso”.

En estos resultados se ha usado como $\phi_1(u, v)$ un promedio entre las diversas PhFTs, con esto hemos conseguido unos resultados que, aunque aceptables, no son estrictamente invariantes al cambio de escala, ya que existen pequeñas variaciones del valor del pico de unas escala a otras. Estos resultados se pueden mejorar si en lugar de tomar $\phi_1(u, v)$ como el promedio de PhFTs, se puede generar una referencia por medio de un proceso iterativo. Este proceso, llamado ecualizado, empieza tomando el inverso de los valores del pico de correlación mostrados en la Figura 4.9, como pesos para generar el $\phi_1(u, v)$ como un promedio ponderado. Usando este nuevo $\phi_1(u, v)$ se repite el proceso obteniéndose unos nuevos valores de pico, que se invierten tomándose como nuevos picos y se cierra el ciclo de realimentación, obteniendo una salida invariante a escalas al cabo de pocas iteraciones:

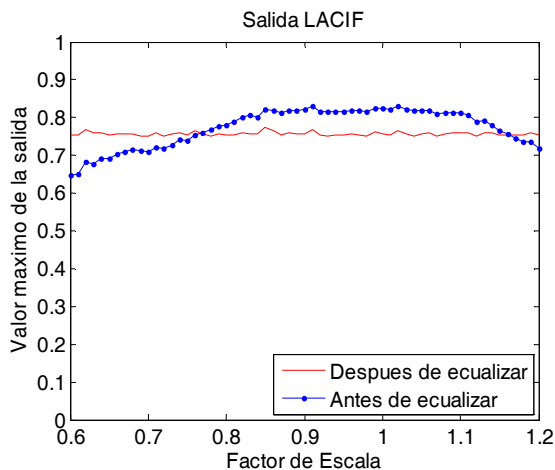
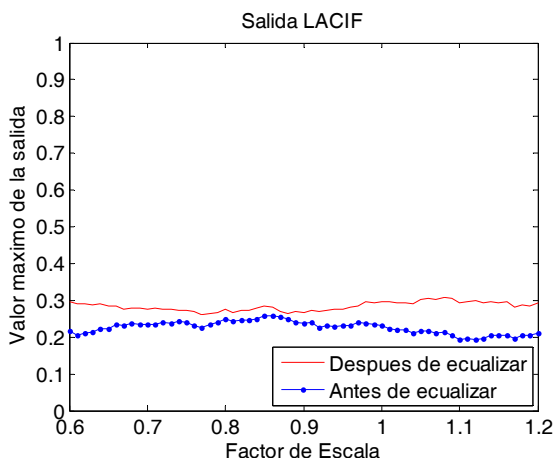


Figura 4.18 Resultados de la detección para el objeto “oso”, generando $\phi_1(u, v)$ mediante un proceso iterativo

Nótese cómo ahora el valor es prácticamente constante e igual a 0.75 para el rango de escalas considerado.

Por el contrario, la aplicación de esta técnica empeora ligeramente la discriminación del método, como se ve en la siguiente figura:



**Figura 4.19 Resultados de la detección para el objeto “test 2”
generando $\phi_1(u, v)$ mediante un proceso iterativo**

Tras el uso de este proceso iterativo el valor promedio para un objeto falso aumenta ligeramente.

4.5 Conclusiones

En este capítulo hemos visto cómo las propiedades de la PhFT permiten reducir, de forma aproximada, el problema del reconocimiento invariante a escalas tridimensional al de reconocimiento bidimensional invariante a intensidad.

Mediante el uso de técnicas vectoriales y de promediado, hemos visto cómo el filtraje LACIF, que introdujimos en el capítulo 2, puede ser aplicado con éxito a este problema, consiguiendo así un método de reconocimiento

invariante a escala. El método está basado en dos etapas, primero la codificación, que retiene la linealidad frente a traslaciones, y segundo el filtraje, basado en correlaciones y por tanto implementable por medio de FFTs en forma digital u ópticamente utilizando correladores.

De forma adicional, este método puede ser ajustado para ofrecer una respuesta invariante en el sentido estricto de la palabra, mediante un método iterativo de selección de la base del espacio vectorial.

En su mayor parte los resultados obtenidos con este método aparecen en [Vallés-07].

Capítulo 5

Reconocimiento de objetos tridimensionales con invariancia a cambios de escala y rotaciones

5.1 Introducción

En este capítulo veremos cómo es posible extender el método aplicado en el capítulo anterior de reconocimiento 3D con cambios de escala a la codificación 3DOOM, para obtener invariancia a escala y rotaciones de objetos 3D de forma simultánea.

Existen varios caminos para abordar el problema general de reconocimiento con invariancia a diversas alteraciones, principalmente en el caso bidimensional. Así existen un conjunto de técnicas basadas en el uso de redes neuronales, como por ejemplo [Perantonis-92].

Un enfoque diferente es que plantean Torres-Mendez *et al.*, [Torres-Mendez-00], proponiendo un método basado en un preprocesado en dos etapas para la extracción de características, con una fase similar a calcular el momento de inercia normalizado de un sólido con lo que consiguen invariancia a traslaciones, rotaciones y escala, más una extracción de características basadas en el trabajo de Watanabe, [Watanabe-85]. Una vez extraídas las características, el reconocimiento se lleva a cabo con un sistema entrenado por un algoritmo holográfico por vecindad, [Sutherland-92], en lugar de usar una red neuronal. La diferencia entre una red neuronal holográfica y otra convencional se encuentra en las neuronas, los elementos básicos de la red. En una red normal se consigue que ésta en su conjunto ofrezca una salida

determinada, para un estímulo de entrada, a partir de un método de entrenamiento. En cambio, en una red holográfica, cada una de sus neuronas es funcionalmente equivalente a una red completa.

Con un enfoque óptico existen también numerosos métodos que presentan múltiples invariancias. Uno de los primeros métodos fue el propuesto por Casasent y Psaltis, [Casasent-76a] y [Casasent-76b], del que ya se ha hablado en el apartado 4.1, con el que consiguieron invariancia a escala y rotación al aplicar una transformada de Mellin, pero perdiendo la invariancia a traslaciones.

Mendlovic *et. al.*, [Mendlovic-95], desarrollaron métodos invariantes a las tres principales alteraciones geométricas, escala, rotación y traslación para el caso 2D. Este método está basado en las ideas descritas en capítulos anteriores y consiste en una descomposición de la señal de entrada en varias etapas, un primer desarrollo en términos de los armónicos radiales de Mellin, $\{r^{i2\pi N-1}\}$, seguido por la selección de un término del desarrollo y un último desarrollo de éste en términos del conjunto $\{|x|^{i2\pi N-1/2}\}$.

Otro método que también es capaz de un reconocimiento invariante a estas tres alteraciones es el propuesto por Fang y Hausler, [Fang-90], que está basado en usar transformaciones matemáticas similares a la transformada Fourier, pero con un kernel¹ integral diferente. Estas transformaciones se diseñan de forma específica para la invariancia deseada, pero tiene el

¹ Se llama así al término $K(\alpha, t)$ en una transformada integral del tipo $g(\alpha) = \int_a^b f(t)K(\alpha, t)dt$, de la que la transformada de Fourier es un caso particular, véase [Mathworld-02b] para más detalle.

inconveniente, ya indicado en el capítulo anterior, de no ser fáciles de computar digitalmente, ya que no se pueden expresar en términos de la FFT.

Otra línea de gran interés se basa en el uso de descriptores. Por ejemplo Duvernoy, [Duvernoy-84], propuso un método de reconocimiento con invariancia a escala, traslación y rotación basado en el uso de descriptores de Fourier.

Así mismo, Mersereau y Morris, [Mersereau-86], propusieron un método basado en el desarrollo en armónicos circulares, lo cual proporcionaba invariancia a rotaciones y traslación usando luz blanca, de modo similar a [Esteve-Taboada-00], ya mencionado en capítulo 4.

Comparado con los métodos existentes, en este capítulo proponemos un esquema en dos etapas, codificación y reconocimiento, en la que la codificación usa la mayor parte de la información del objeto y no se limita a extraer una serie de descriptores o características.

La codificación que nosotros usamos, basada en el 3DOOM, tiene como propiedades el ser invariante a traslaciones y el incluir toda la información sobre las rotaciones de un objeto 3D cualquiera, ver capítulo 3. Como el 3DOOM está definido en términos de PhFTs, estas convertirán los cambios de escala a los que podemos someter al objeto en cambios de contraste, con lo que podemos aplicar un LACIF modificado. Esta es la novedad de este capítulo. Como principal diferencia con los métodos previos, esta codificación está definida en S^2 , la superficie de una esfera, y por tanto las correlaciones necesarias para definir el LACIF estarán definidas en S^2 . Además el pico de correlación del LACIF nos dará una estimación de la

rotación del objeto. En definitiva proponemos un método de reconocimiento invariante a traslación, escala y rotación.

5.2 LACIF y 3DOOM

La primera etapa del método que proponemos está basada en la propiedad de la PhFT de convertir un cambio de escala en una variación de contraste. Esta característica fue explicada con detalle en el apartado 4.2 y es ampliable al 3DOOM ya que esta codificación está formada a partir de PhFTs. Es decir, en el 3DOOM también se cumple que un cambio de escala en el objeto original se transforma en una variación de intensidad, como ilustra la siguiente figura:

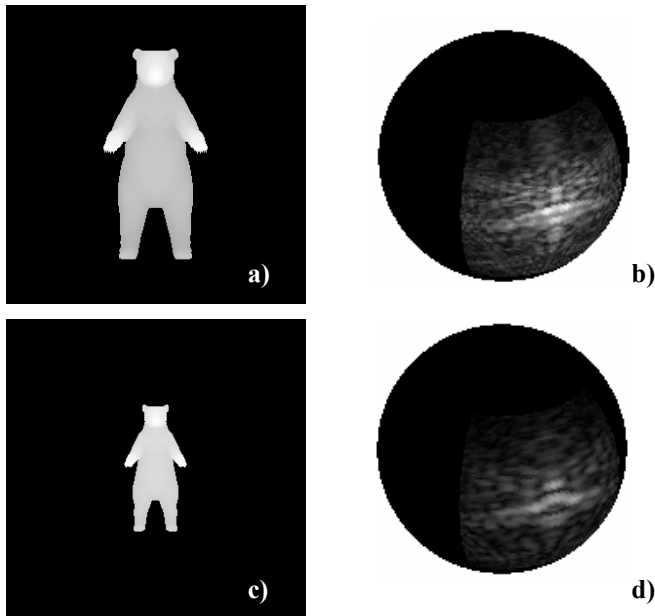


Figura 5.1 En a) y c) mostramos el mismo objeto con escala 1.0 y 0.7, respectivamente, en b) y d) mostramos la PhFT, en S^2 , correspondiente a cada escala

Se observa cómo, de igual manera que en el capítulo anterior, el 3DOOM del objeto escalado es similar al del objeto sin escalar, pero multiplicado por una constante. Asimismo se observa cómo la versión correspondiente al objeto con un factor de escala menor presenta el efecto de filtraje pasa-baja, debido al submuestreo y la no exactitud de la Ec. (4.4).

Con esto es posible plantear el uso del LACIF de forma análoga al caso de invariancia a escala, ya que para el 3DOOM $g(\theta, \varphi)$ de un objeto dado, en una escala concreta, es de la forma:

$$g(\theta, \varphi) = af(\theta, \varphi) + bv_0(\theta, \varphi) \quad (5.1)$$

donde a y b son constantes arbitrarias, $f(\theta, \varphi)$ es la PhFT obtenida desde un punto de vista concreto y $v_0(\theta, \varphi)$ es una función binaria, definida en la superficie de la esfera tal que, es igual a la unidad en el soporte de $f(\theta, \varphi)$ y cero en el resto de S^2 . Recordemos que la PhFT de una imagen de rango no está definida para todo el espacio (θ, φ) , ya que la inclinación máxima de las facetas con respecto al plano de referencia, determina unas coordenadas máximas, obtenidas por medio de la ec.(3.10), que se corresponden con la máscara $v_0(\theta, \varphi)$.

Ortonormalizando el conjunto $\{v_0(\theta, \varphi), f(\theta, \varphi)\}$, podemos generalizar a S^2 el LACIF antes definido en \mathbb{R}^2 , determinando a la salida el resultado siguiente:

$$LACIF_{gf}(\alpha, \beta, \gamma) = \frac{\left[g *_{SO(3)} \hat{f}_o \right]^2(\alpha, \beta, \gamma)}{\sqrt{N} \left[g *_{SO(3)} \hat{v}_0 \right](\alpha, \beta, \gamma) - \left[g *_{SO(3)} \hat{v}_0 \right]^2(\alpha, \beta, \gamma)} \quad (5.2)$$

en donde $\{\hat{v}_0(\theta, \varphi), \hat{f}_o(\theta, \varphi)\}$ es la base que resulta de aplicar el método de ortonormalización de Gramm-Schmidt a $\{v_0(\theta, \varphi), f(\theta, \varphi)\}$ y $*_{SO(3)}$ es la correlación en S^2 definida en la ec. (3.13).

Nótese que N es el número de píxeles de la función soporte $v_0(\theta, \varphi)$ y puede entenderse como el área, en S^2 , de esta función y calcularse en términos de la siguiente integral:

$$N = Area(v_0(\theta, \varphi)) = \int_{S^2} v_0(\theta, \varphi) d\theta d\varphi \quad (5.3)$$

Usando el hecho de que $v_0(\theta, \varphi)$ es binaria y la definición (3.17) del capítulo 3 para el producto escalar en S^2 :

$$N = \int_{S^2} v_0^2(\theta, \varphi) d\theta d\varphi = \langle v_0 | v_0 \rangle = \|v_0\|^2 \quad (5.4)$$

Es decir, el área subtendida por la función en S^2 viene dado por el módulo del vector $w(\theta, \varphi)$. Aplicando (5.4) a la ecuación (5.2) queda la expresión definitiva de la salida que proporciona el LACIF que usaremos en este capítulo:

$$LACIF_{gf}(\alpha, \beta, \gamma) = \frac{\left[g *_{SO(3)} \hat{f}_o \right]^2(\alpha, \beta, \gamma)}{\|v_0\| \left[\left[g *_{SO(3)} \hat{v}_0 \right](\alpha, \beta, \gamma) - \left[g *_{SO(3)} \hat{v}_0 \right]^2(\alpha, \beta, \gamma) \right]} \quad (5.5)$$

5.3 Implementación y resultados

Una vez vistas las propiedades de la PhFT y el 3DOOM e introducido el LACIF para la superficie esférica, la aplicación de éste último es inmediata. Pero a nivel de implementación se plantean una serie de dificultades que han de ser abordadas:

- El problema completo tiene 4 grados de libertad, uno para la escala y 3 para las rotaciones. Esto hace difícil el barrer todas las posibles opciones y el representar los resultados obtenidos.
- Como en el capítulo anterior, el 3DOOM de un objeto con un factor de escala menor es una versión pasa-baja del original. Esto ha de ser tenido en cuenta en la implementación del método.
- Por último, el cálculo de las correlaciones en SO(3) se lleva a cabo mediante la expresión (3.35) $f(\theta, \varphi) * g(\theta, \varphi) =$

$$\sum_{j,m,m'} d_{m,m'}^j(\beta) \cdot F_{j,m'} G_{j,m} e^{im\gamma} e^{im\alpha}, \text{ lo cual implica escoger un } j_{\max} \text{ de}$$

desarrollo en términos de los armónicos esféricos.

Con respecto al primer punto, en este capítulo abordaremos la presentación de los resultados en dos etapas, una parcial donde sólo consideraremos un ángulo de Euler y la variación de escala y otra total donde tendremos en cuenta el problema completo. Por su volumen, los datos que ofrecemos son más exhaustivos para el problema reducido que para el completo.

Con respecto a los otros dos puntos, en esta implementación se tendrán en cuenta de forma conjunta. Se trabajará con $j_{\max} = 80$, que, como hemos visto anteriormente, ofrece un buen equilibrio entre tiempo de cálculo y exactitud. El hecho de limitar el orden del desarrollo también se justifica por el efecto pasa-baja del cambio de escala en el 3DOOM. Como se ha explicado en el capítulo 4, el escalar una PhFT provoca un efecto de filtrado pasa baja, que se transmite al 3DOOM. Por ello, si sólo usamos hasta un cierto orden en el desarrollo, éste también estará filtrado en bajas frecuencias, como muestran las Figuras 3.20 y 3.21, con lo que esperamos mejorar el resultado de esta forma.

Así mismo, este efecto se tendrá en cuenta mediante el uso de una referencia compuesta por la función soporte $v_0(\theta, \varphi)$ y una versión promedio de los 3DOOMs a diferentes escalas. A diferencia del capítulo 4, no se usará ningún umbral en la función soporte, $v_0(\theta, \varphi)$, para filtrar información, definiendo $v_0(\theta, \varphi)$ como un cuadrado sobre S^2 tal que sus aristas opuestas están separadas 180° , según (θ, φ) .

5.3.1 Resultados con el 3DOOM parcial

Comencemos planteando una versión limitada del problema de invariancia a escala y rotación. Si sólo consideramos una rotación, las PhFTs sólo sufrirán una traslación con respecto a un eje, tal y como se ve en la siguiente figura:

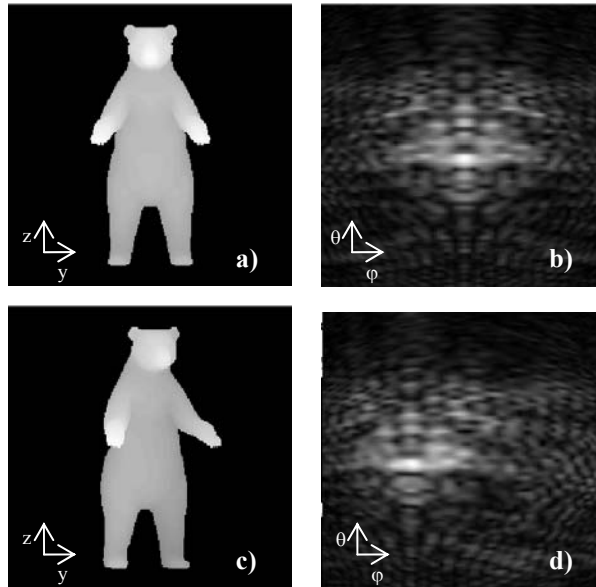


Figura 5.2 En a), imagen de rango del objeto “oso” sin rotar, en b) PhFT correspondiente a la imagen a), idem en c) y d) pero con el objeto rotado 17° alrededor del eje Z

De acuerdo con Esteve-Taboada, [Esteve-Taboada-03], se puede usar este hecho para generar una versión parcial del 3DOOM, que sólo tenga en cuenta una única rotación. Este 3DOOM parcial se muestra en la siguiente figura:

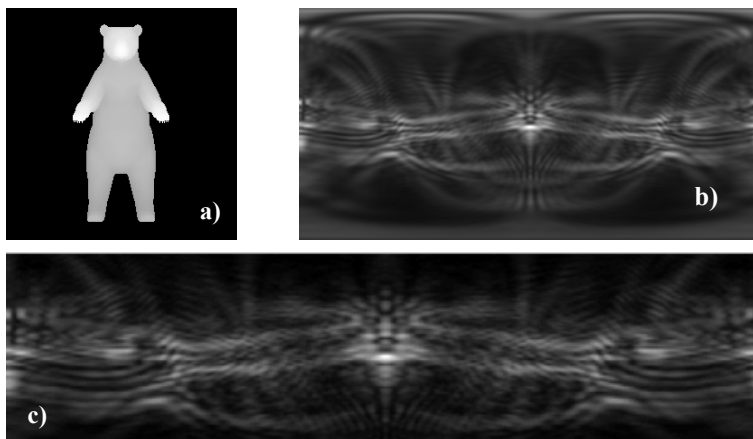


Figura 5.3 En a) mostramos el objeto “oso”, en b) el 3DOOM completo correspondiente y en c) mostramos el 3DOOM parcial

Nótese como la Figura 5.3 c) es muy similar a la banda central del 3DOOM completo, encontrándose las diferencias en las zonas superior e inferior. Esto se debe a que no han sido tenidas en cuenta las rotaciones respecto al eje Y, que sí permitirían ver las zonas superior e inferior del objeto.

El 3DOOM parcial está generado de forma totalmente similar al completo, mediante la rotación del objeto (pero sólo en torno a un eje), el cálculo de PhFT y el pegado de esa distribución en una versión plana del 3DOOM. Este proceso se muestra en la siguiente figura:

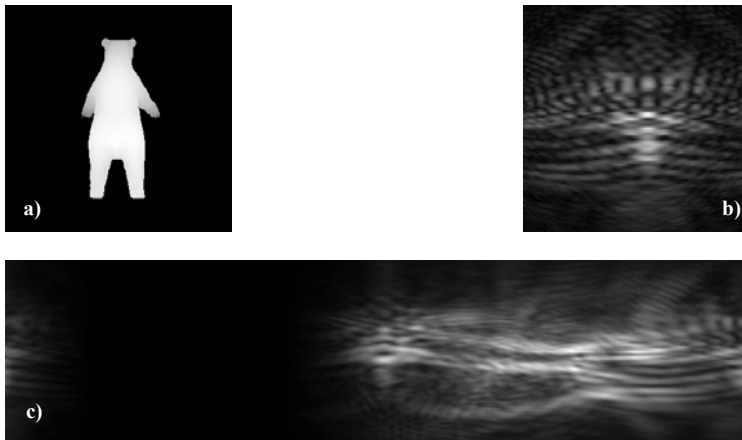


Figura 5.4 Proceso de generación del 3DOOM parcial en una etapa intermedia, en a) mostramos el objeto rotado 174° según el eje Z, en b) la PhFT correspondiente y en c) mostramos el 3DOOM parcial obtenido hasta ese ángulo

Una vez generado este 3DOOM parcial se puede adaptar el LACIF para obtener reconocimiento invariante a escala y a una rotación en torno al eje Z

$$LACIF_{3DOOM, \phi}(u, v) = \frac{(3DOOM_P * \phi_1)^2(u, v)}{\sqrt{N} (3DOOM_P^2 * \phi_0)(u, v) - (3DOOM_P * \phi_0)^2(u, v)} \quad (5.6)$$

donde hay tres factores a tener en cuenta. El primero es que el 3DOOM parcial contiene el conjunto de PhFTs, por tanto tomaremos éste como escena y no como venimos haciendo hasta ahora. Esto no es más que una sutileza matemática para facilitar la implementación y no tiene mayores implicaciones.

El segundo es que, también a nivel de implementación, trabajaremos con el promedio de los 3DOOMs parciales a diferentes escalas, por las razones expuestas arriba. Y tercero, por $\{\phi_0(u,v), \phi_1(u,v)\}$ notamos la base vectorial de referencia que usamos para aplicar el LACIF, que en este caso es el resultado de ortonormalizar el conjunto $\{w(u,v), PhFT(u,v)\}$, siendo $PhFT(u,v)$ la correspondiente a una determinada imagen de rango con una escala y orientación y que se muestra en un caso particular en la siguiente figura:

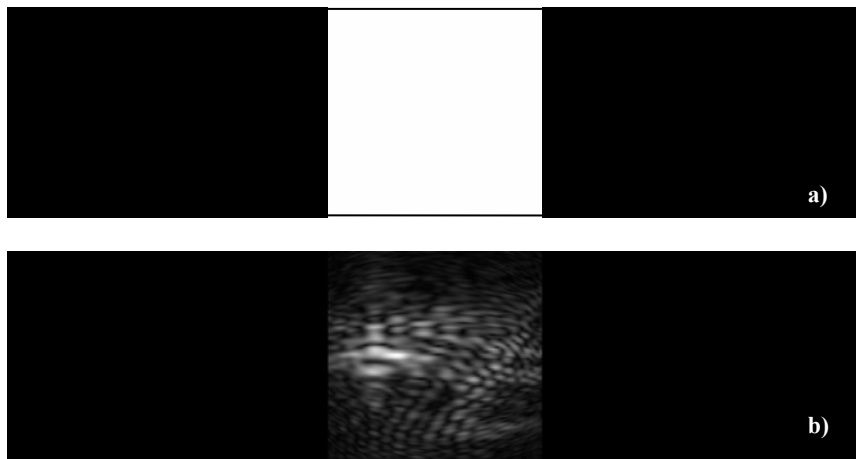


Figura 5.5 En a) mostramos la función soporte $w(u,v)$, en b) la Phft correspondiente al objeto “oso” rotado 30° y con escala 0.75

Con todo esto ya podemos aplicar nuestro método a pruebas de reconocimiento y discriminación entre diferentes objetos 3D, en este caso los objetos “oso”, “test 2” y “test 3”, tomando como objeto correcto el “oso”.

El rango de variación del factor de escala que consideraremos será [0.6-1.2] con una variación de 0.05 entre escalas y el rango de variación angular alrededor del eje Z estará entre [0°-360°], con una variación de 10°.

Debido a la complejidad para representar los datos hemos seguido el siguiente convenio; se considera que el objeto correcto ha sido detectado de forma satisfactoria si obtenemos un valor de pico igual o superior a 0.7 en la salida del LACIF para un valor de ángulo y escalado dados; en caso contrario consideramos que el objeto no ha sido detectado. Recordaremos que el valor máximo proporcionado por el LACIF es la unidad.

Si el objeto, para una escala y rotación dada, ha sido detectado la celda se pinta de color blanco, en caso contrario se pinta de negro. Las celdas se han distribuido de forma polar, siendo la coordenada angular φ , muestreada en intervalos de 10°, el ángulo de rotación del objeto según el eje Z, y la radial la escala, muestreada en intervalos de 0.05. En las siguientes figuras mostramos los resultados de acuerdo con esta convención:

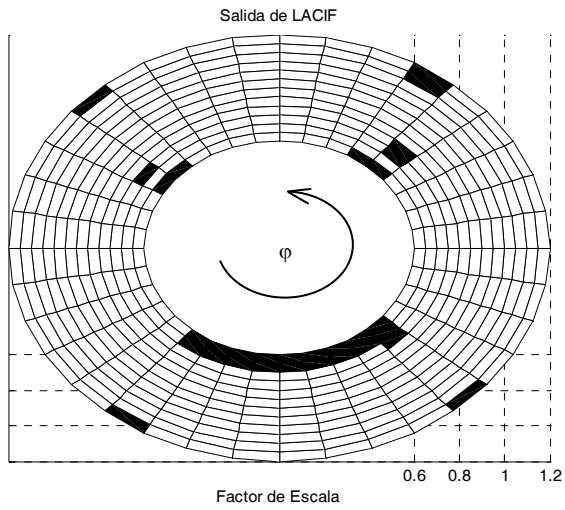


Figura 5.6 Diagrama circular de reconocimiento para el objeto “oso”, en la dirección radial se representa la variación del factor de escala, en la angular la rotación con respecto al eje Z, la tasa de detección es del 96%

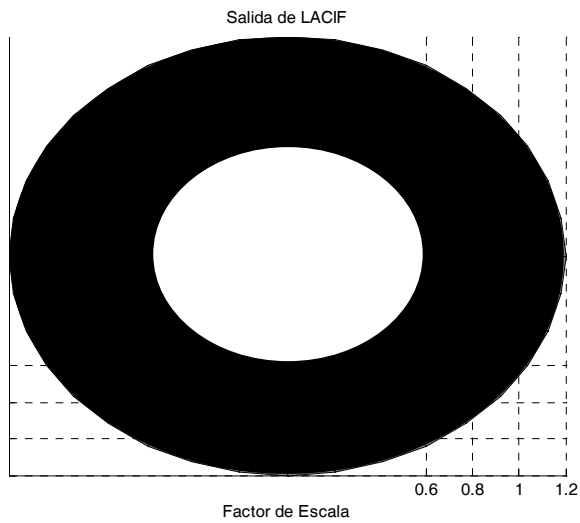


Figura 5.7 Diagrama circular de reconocimiento entre el objeto “oso” y el objeto “test 2”, en la dirección radial se representa la variación del factor de escala, en la angular la rotación con respecto al eje Z

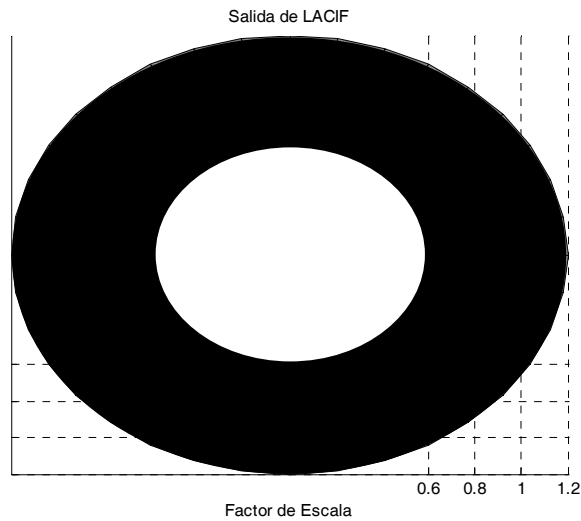


Figura 5.8 Diagrama circular de reconocimiento entre el objeto “oso” y el objeto “test 3”, en la dirección radial se representa la variación del factor de escala, en la angular la rotación con respecto al eje Z

Se observa como el método es capaz de reconocer el objeto verdadero y discriminar con otros objetos de forma robusta.

En las siguientes figuras mostramos el plano de salida para tres casos concretos:

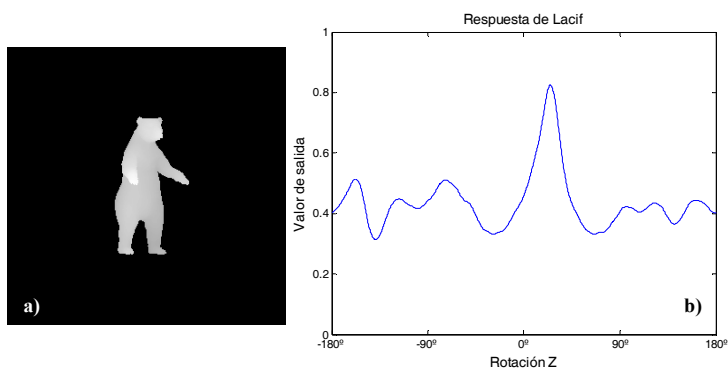


Figura 5.9 En a) mostramos el objeto “oso” rotado 30 grados alrededor del eje Z y con escala 0.7, en b) mostramos el perfil horizontal del plano de salida que contiene al pico del LACIF

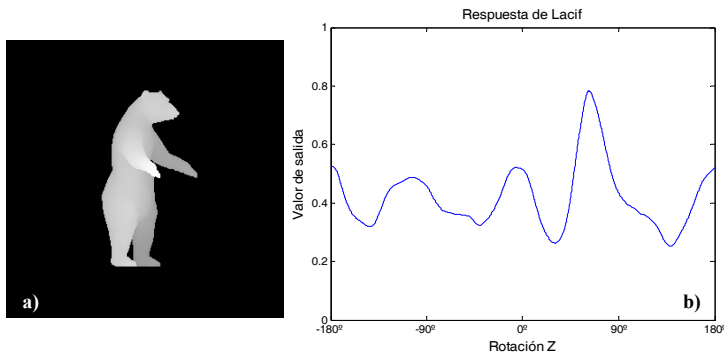


Figura 5.10 En a) mostramos el objeto “oso” rotado 60 grados alrededor del eje Z y con escala 0.9, en b) mostramos el perfil horizontal del plano de salida que contiene al pico del LACIF

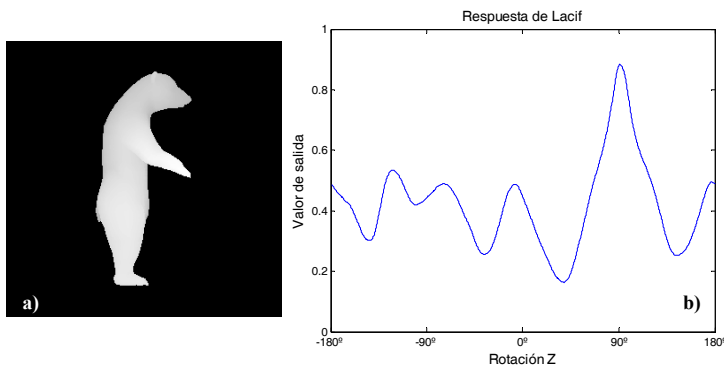


Figura 5.11 En a) mostramos el objeto “oso” rotado 90 grados alrededor del eje Z y con escala 1.1, en b) mostramos el perfil horizontal del plano de salida que contiene al pico del LACIF

Se observa como el LACIF da un pico de correlación en torno al ángulo correcto, siendo inferior para el resto de ángulos

5.3.2 Resultados con el 3DOOM completo

Una vez analizados los resultados de la aplicación del LACIF al variar la escala y rotar el objeto alrededor del eje Z, vamos a ver qué resultados ofrece

ante el problema completo, es decir considerando además la rotación alrededor del eje Y. En este caso el LACIF lo aplicaremos sobre la superficie de la esfera del 3DOOM.

Para aplicar el LACIF en la superficie de una esfera, necesitamos implementar la ecuación (5.2), teniendo en cuenta que nuestro conjunto, a partir del cual generaremos la base vectorial $\{\phi_0(u,v), \phi_1(u,v)\}$, será el formado por el soporte y la PhFT del objeto en una escala y rotación dada $\{w(\theta, \varphi), PhFT(\theta, \varphi)\}$, tal y como se muestra en la siguiente figura:



Figura 5.12 En a) mostramos el objeto “oso” rotado (45,90-30) a escala 0.6, en b) mostramos la función soporte $v_0(\theta, \varphi)$, en c) la PhFT correspondiente al objeto a)

Hacemos notar la diferencia entre este apartado y el anterior. En éste, el espacio de definición de las funciones está dado por las coordenadas (θ, φ) , es decir, estamos trabajando en la superficie de una esfera de radio unidad, S^2 . En el apartado anterior la función soporte estaba dada en términos de (u, v) , es decir estábamos en el espacio recíproco de \mathbb{R}^2 .

Una vez obtenida la base, aplicamos el LACIF, tomando como escena el $3DOOM(\theta, \varphi)$ promediado a diferentes escalas, para el objeto correcto y otros incorrectos para el rango de escalas [0.6-1.2] y ciertas rotaciones. En

todos los casos el grado de desarrollo en armónicos esféricos es $j_{\max} = 80$.

En las siguientes figuras mostramos los resultados obtenidos:

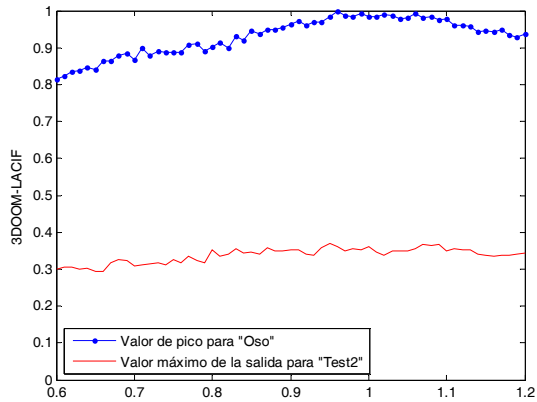


Figura 5.13 Resultados reconocimiento del objeto “oso” sin rotar, $(\alpha, \beta, \gamma) = (0, 0, 0)$, y discriminación del LACIF para el objeto “test 2”

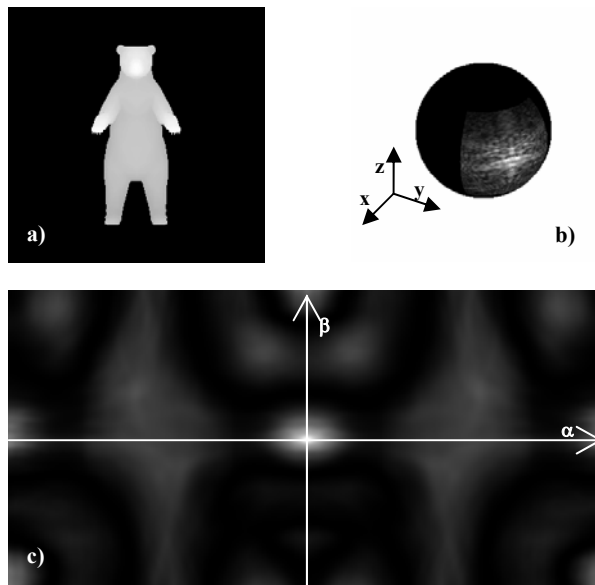


Figura 5.14 En a) mostramos el objeto “oso” sin rotar y a escala 1, en b) la PhFT correspondiente, en c) mostramos el plano de salida del 3DOOM-LACIF ($\gamma=0$)

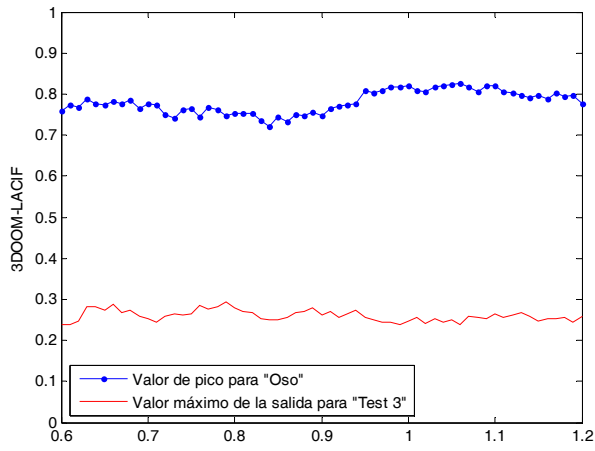


Figura 5.15 Resultados reconocimiento del objeto “oso” rotado (120,30,0), y discriminación del LACIF para el objeto “test 3”

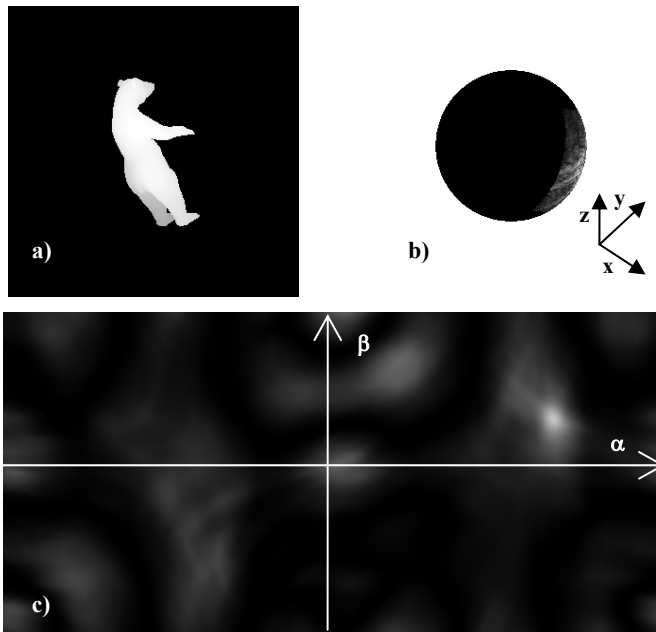


Figura 5.16 En a) mostramos el objeto “oso” rotado (120,30,0), y a escala 0.75, en b) la PhFT correspondiente, en c) mostramos el plano de salida del 3DOOM-LACIF ($\gamma=0$)

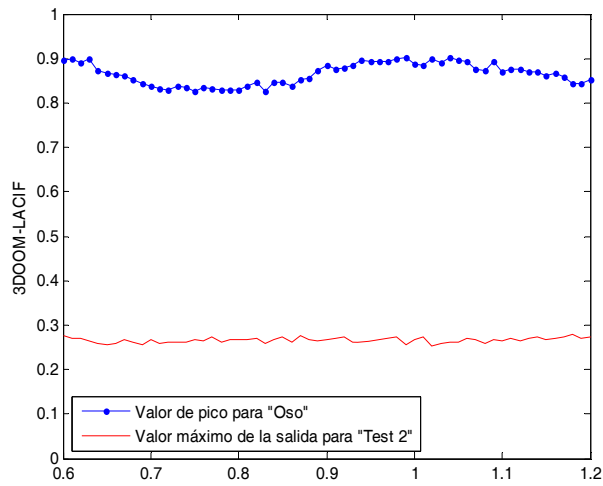


Figura 5.17 Resultados reconocimiento del objeto “oso” rotado (90,-30,0), y discriminación del LACIF para el objeto “test 2”

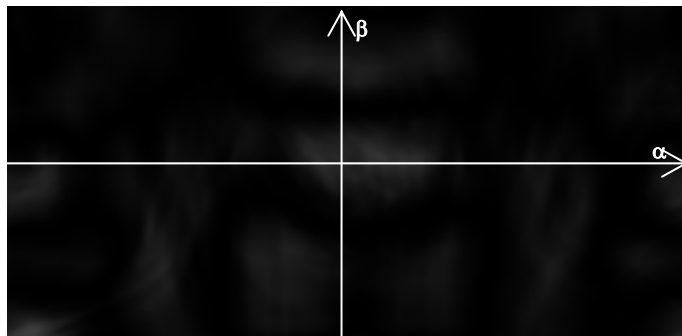


Figura 5.18 Plano de salida del 3DOOM-LACIF ($\gamma=0$) obtenido con el objeto “test 2” a escala 1.15 al buscar el objeto “oso” rotado (90,-30,0)

Se observa que el método propuesto es capaz de reconocer y discriminar objetos, de forma invariante a su factor de escala, en el rango considerado, y a su orientación.

5.4 Conclusiones

En este capítulo se ha mostrado cómo la codificación 3DOOM comparte las propiedades de la PhFT que convierten un cambio de escala en una variación de intensidad de la imagen.

Usando este hecho se ha propuesto el uso del LACIF para obtener reconocimiento de objetos 3D con invariancia a escala y a rotación. Mostrándose tanto un caso parcial, con una sola rotación, como el problema completo, con las tres rotaciones respecto de los ángulos de Euler, habiendo implementado el equivalente del LACIF en 2D en el espacio de funciones definidas en la superficie de una esfera.

De forma adicional, la codificación propuesta es invariante a la posición del objeto, siendo, por tanto, invariante a traslaciones de éste. Es decir, el método en su conjunto es invariante a traslaciones, escala y rotaciones.

Los resultados de este trabajo han sido publicados en [Vallés-08].

Capítulo 6

Conclusiones

A lo largo de los distintos capítulos de esta memoria se han propuesto métodos de reconocimiento invariante de objetos tridimensionales frente a diversas alteraciones de los mismos.

Como resumen, se presentan las aportaciones de este trabajo en los siguientes puntos:

- De forma general, se ha empleado una estructura de espacio vectorial para el conjunto de imágenes 2D. Dicho espacio está constituido por las imágenes $f(x, y)$ de cuadrado integrable, definiéndose el producto escalar como $\langle f | g \rangle \equiv \int_{\mathbb{R}^2} f^*(x, y) g(x, y) dx dy$, coincidiendo con la correlación entre funciones. Gracias a este planteamiento, a lo largo de todo el trabajo se han podido aplicar las propiedades algebraicas de los espacios vectoriales al conjunto de las imágenes 2D.
- Aplicando la estructura de espacio vectorial, se ha podido reducir sustancialmente la dificultad del reconocimiento de objetos 3D con invariancia a iluminaciones. Si en un planteamiento directo con imágenes digitales de $N_1 \times N_2$ píxeles, la dimensionalidad del problema viene dada por el producto $N_1 \times N_2$, se ha demostrado que es posible abordarlo, aproximadamente, como un problema con sólo tres grados de libertad.

- Se ha propuesto un método de reconocimiento para objetos tridimensionales, el LADC, (Local Angular Distance Correlation) que permite obtener invariancia a la posición e intensidad de la fuente que ilumina a los objetos 3D y es implementable en forma rápida por medio de FFTs. Se han comparado los resultados obtenidos frente a otros métodos similares.
- De forma adicional, mediante la aplicación de la transformación de Karhunen-Loeve, o de componentes principales, se ha mostrado que existe una forma matemáticamente óptima de mejorar los resultados obtenidos con el LADC, a costa de aumentar la dimensionalidad del problema y, por ello, el tiempo de cálculo necesario.
- Se ha usado una codificación, previamente existente, de la información geométrica de los objetos 3D. Esta codificación, PhFT, (Phase Fourier Transform), toma como entrada imágenes de rango, una técnica de representación 3D bien conocida en la que el tono de gris es proporcional a la distancia entre el objeto 3D y un plano de referencia. Con estas imágenes de rango, la PhFT representa en un plano las diferentes orientaciones de la superficie visible del objeto.
- Se ha mostrado como la PhFT permite una representación de la información que es invariante a traslaciones y transforma un cambio de escala del objeto 3D en una variación de intensidad y una rotación en algo aproximadamente igual a una traslación.
- A partir de esta codificación básica se ha propuesto un objeto matemático, 3DOOM, (3 Dimensional Object Orientation Map), definido en la superficie de una esfera, tal que engloba a todas las posibles PhFTs de un objeto. Gracias a ello, el 3DOOM contiene toda la información geométrica de la superficie externa de un objeto.
- Usando esta información codificada en el 3DOOM, se ha propuesto una operación equivalente a la correlación en \mathbb{R}^2 , pero definida en la

superficie de la esfera, S^2 . De la misma forma que la correlación en \mathbb{R}^2 permite reconocer objetos 2D e indica cual es su posición, el método propuesto para S^2 permite el reconocimiento de objetos tridimensionales y estima la rotación a que han sido sometidos.

- Usando otra de las propiedades de la PhFT se ha reducido el problema del reconocimiento de objetos 3D con invariancia a escala, al reconocimiento bidimensional con invariancia a cambios multiplicativos de intensidad, aplicándose con éxito el LACIF, (Locally Adaptive Contrast Invariant Filter).
- Se ha generalizado el LACIF en el espacio de funciones definido en la superficie de una esfera, obteniendo invariancia a escala y rotación de forma simultánea. Esto ha sido posible ya que un cambio de escala de un objeto 3D se traduce en un cambio de intensidad en su PhFT y la codificación 3DOOM está definida en términos de las PhFTs de un objeto, haciendo posible la aplicación del filtraje LACIF.
- En todos los métodos propuestos, tanto en el caso del LADC aplicado a cambios de iluminación como en los basados en PhFTs, la invariancia a traslaciones característica de la correlación en \mathbb{R}^2 se ha mantenido.

Bibliografía

- [Arsenault-85] H. H. Arsenault and C. Delisle, “Contrast-invariant pattern recognition using circular harmonic components”, *Appl. Opt.* 24, 2072-2075 (1985)
- [Arsenault-89] H. H. Arsenault, T. Szoplik and B. Macukow, *Optical Processing and Computing*, Academic Press Inc., San Diego (1989)
- [Arsenault-00] H. H. Arsenault and D. Lefebvre, “Homomorphic cameo filter for pattern recognition that is invariant with changes in illumination”, *Opt. Lett.* 25, 1567-1569 (2000)
- [Arsenault-05] H. H. Arsenault and P. García-Martínez, “Intensity-invariant nonlinear filtering for detection in camouflage”, *Appl. Opt.* 44, 5483-5490 (2005)
- [Bamler-82] R. Bamler and J. Hofer-Alfeis, “Three- and four-dimensional filter operations by coherent optics,” *Opt. Acta* 29, 747–757 (1982).
- [Basri-03] R. Basri and D. W. Jacobs, “Lambertian Reflectance and Linear Subspaces”, *IEEE Trans. Pattern Analysis and Machine Intelligence* 25, 218-233, (2003)
- [Belhumeur-96] P. N. Belhumeur and D. J. Kriegman, “What Is the Set of Images of an Object Under All Possible Lighting Conditions?”, *International Journal of Computer Vision*, Vol. 28, 1–16 (1998)
- [Belhumeur-97] P. N. Belhumeur, J. P. Hespanha and D. J. Kriegman, “Eigenfaces vs. Fisherfaces:

- recognition using class specific linear projection”,
IEEE Transactions on Pattern Analysis and
Machine Intelligence 19, 711-720 (1997)
- [Blais-02] J. A .R. Blais and D. A. Provins, “Spherical
Harmonic Analysis and Synthesis for Global
Multiresolution Applications”, Journal of Geodesy
76, 29-35 (2002)
- [Blanco-97] M. A. Blanco, M. Flórez, and M. Bermejo,
“Evaluation of the rotation matrices in the basis of
real spherical harmonics”, J. Mol. Struct.
(Theochem) 419, 19-27 (1997)
- [Boyer-02] K. Boyer, R. Srikantiah, and P. Flynn, “Saliency
Sequential Surface Organization for Free Form
Object Recognition”, Computer Vision and Image
Understanding 88, 152-188 (2002)
- [Bracewell-63] R. Bracewell, *The Fourier transform and its
applications*, Ed. McGraw-Hill (1963)
- [Brezis-84] H. Brezis, *Análisis funcional: teoría y
aplicaciones*, Alianza, Madrid (1984)
- [de Burgos-94] J.de Burgos, *Curso de álgebra y geometría*,
Alambra, Madrid (1994)
- [Campbell-01] R. Campbell and P. Flynn, “A survey of free-form
object representation and recognition techniques,”
Computer Vision and Image Understanding 81 2,
166–210 (2001)
- [Campbell-03] R. Campbell and P. Flynn, “Free-Form Object
Recognition in Range Data Using Weak
Correspondence Between Local Features”, Int. J.
Pattern Recognition and Artificial Intelligence 17,
1245-1277 (2003)

- [Casasent-76a] D. Casasent and D. P. Psaltis, "Position, rotation, and scale invariant optical correlation", *App. Opt.* 15, 1795-1799 (1975)
- [Casasent-76b] D. Casasent and D. P. Psaltis, "Scale invariant optical correlation using Mellin transforms", *Opt. Comm.* 17, 59-63 (1976)
- [Cendrillon-00] R. Cendrillon and B. Lovell, "Real-time face recognition using Eigenfaces", *Proc. of the SPIE International Conference on Visual Communications and Image Processing 4067*, 269-276 (2000)
- [Chai-07] X. Chai, S. Shan, X. Chen, "Locally Linear Regression for Pose-Invariant Face Recognition" *IEEE Transaction on Image Processing* 16, 1716-1725 (2007)
- [Chalasincka-Macukow-93] K. Chalasincka-Macukow, F. Turon, M. J. Yzuel, and J. Campos, "Contrast performance of pure phase correlation," *J. Opt.* 24, 71-75 (1993).
- [Chang-97] S. D. Chang, M. Rioux, J. Domey, "Face recognition with range images and intensity images", *Opt. Eng.* 36, 1106-1112 (1997)
- [Cojoc-97] D. Cojoc, M. T. Molina, J. García, and C. Ferreira, "Coordinate-transformed filter for shift-invariant and scale-invariant pattern recognition", *Appl. Opt.* 36, 4812-4815 (1997)
- [Cooley-65] J. W. Cooley, and J. W. Tukey, "An Algorithm for the Machine Computation of the Complex Fourier Series", *Mathematics of Computation* 19, 297-301 (1965)

- [Denis-94] D.C. Ghiglia y L.A.Romero, “Robust two-dimensional weighted phase unwrapping that uses fast transforms and iterative methods”, *J. Opt. Soc. Am. A* 11, 107-117 (1994)
- [Dickey-91] F. M. Dickey and L. A. Romero, “Normalized correlation for pattern recognition,” *Opt. Lett.* 16, 1186–1188 (1991)
- [Dirac-58] P. A. M. Dirac, *The principles of quantum mechanics*, Clarendon press, Oxford (1958).
- [Dotsenko-88] V. S. Dotsenko, “Neural networks: translation-, rotation- and scale-invariant pattern recognition”, *J. Phys. A: Math. Gen.* 21 L783-L787 (1988)
- [Duda-73] R. O. Duda and P. E. Hart, *Pattern classification and scene analysis*, John Wiley & Sons (1973)
- [Duvernoy-84] J. Duvernoy, “Optical-digital processing of directional terrain textures invariant under translation, rotation, and change of scale,” *Appl. Opt.* 23, 828–837 (1984)
- [Esteve-Taboada-00] J. J. Esteve-Taboada, J. García and C. Ferreira “Extended scale invariance pattern recognition using white light illumination”, *Appl. Opt.* 39, 1268-1271 (2000)
- [Esteve-Taboada-02] J. J. Esteve-Taboada, “Procesado de imágenes por técnicas de multiplexado. Aplicación al reconocimiento de objetos tridimensionales”, Tesis Doctoral, Universitat de València, Abril 2002.
- [Esteve-Taboada-03] J. J. Esteve-Taboada and J. García, “Detection and orientation evaluation for three-dimensional objects”, *Opt. Comm.* 217, 123-131 (2003)

- [Fang-90] M. Fang and G. Häusler, “Class of transforms invariant under shift, rotation, and scaling” *Appl. Opt.* 29, 704-708 (1990)
- [Ferreira-89] C. Ferreira, M. J. Buades and A. Moya, “Anamorphic correlator for character recognition. Detection of characters of different size”, *J. Optics* 20, 181-185 (1989)
- [Foley-96] J. Foley, A. van Dam, S. Feiner and J. Hughes, *Computer graphics, principle and practices*, 2nd edition, Addison-Wesley (1996)
- [Fortuna-04] J. Fortuna, D. Capson, “ICA filters for lighting invariant face recognition” *Pattern Recognition, Proceeding on ICPR 2004* 1, 334-337 (2004)
- [FFTW-99] <http://www.fftw.org/>
- [Galindo-89] A. Galindo y P. Pascual, *Mecánica cuántica I y II*, Eudema, Madrid (1989)
- [García-92] J. García, T- Szoplik and C. Ferreira, “Shift and scale invariant pattern recognition using an elliptic coordinate transformed phase only filter”, *Appl. Opt.* 31, 4823-4828 (1992)
- [García-03] J. García, J. Valles and C. Ferreira, “Detection of three-dimensional objects under arbitrary rotations based on range images”, *Opt. Express* 11, 3352-3358 (2003)
- [García-Martínez-95] P. García-Martínez, J. García and C. Ferreira, “A new criterion for determining the expansion center for circular-harmonic filters”, *Opt. Commun.* 117, 399-405 (1995)
- [García-Martínez-98] P. Garcia-Martinex, D. Mas, J. Garcia, and C. Ferreira, “Nonlinear morphological correlation:

- optoelectronic implementation” *Appl. Opt.* 37, 2112–2118 (1998)
- [García-Martínez-99] P. Garcia-Martinez and H. H. Arsenault, “A correlation matrix representation using sliced orthogonal nonlinear generalized decomposition” *Opt. Commun.* 172, 181–192 (1999).
- [García-Martínez-00] P. Garcia-Martinez, H. H. Arsenault, and C. Ferreira, “Binary image decomposition for intensity-invariant optical nonlinear correlation”, in *Optics in Computing 2000*, R. A. Lessard and T. V. Galstian, ed., *Proc. Soc. Photo.-Opt. Instrum. Eng.* 4089, 433–438 (2000).
- [Ge-01] Z. Ge, F. Kobayashi, S. Matsuda and M. Takeda, “Coordinate-transform technique for closed fringe analysis by the Fourier-transform method”, *Appl. Opt.* 40, 1649-1657 (2001)
- [Goldstein-02] H. Goldstein, C. Poole and J. Safko, *Classical mechanics 3rd edition*, pp. 150-161, Addison-Wesley (2002)
- [Goodman-96] J. W. Goodman, *Introduction to Fourier Optics, second Edition*, McGraw-Hill (1996)
- [Gross-04] R. Gross, S. Baker, I. Matthews, and T. Kanade, “Face Recognition Across Pose and Illumination” *Handbook of Face Recognition*, Stan Z. Li and Anil K. Jain, ed., Springer-Verlag, June, 2004.
- [Hallinan-94] P. Hallinan, “A Low-Dimensional Representation of Human Faces for Arbitrary Lighting Conditions,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition* 21, 995-999 (1994)

- [Hansen-81a] E. W. Hansen, "Circular harmonic image reconstruction: experiments", *Appl. Opt.* 20, 2266-2274 (1981)
- [Hansen-81b] E. W. Hansen, "Theory of circular harmonic image reconstruction", *J. Opt. Soc. Am.* 71, 304-308 (1981)
- [Hansen-82] E. W. Hansen, J. G. Verly y E. B. Keirstead, "Rotation invariant optical processing", *J. Opt. Soc. Am.* 72, 1670-1676 (1982)
- [Healy-95] G. Healey and L. Wang, "Illumination-invariant recognition of texture in color images", *J. Opt. Soc. Am. A* 12, 1877-1883 (1995)
- [Healy-98] D. M. Healy Jr., D. Rockmore and P. J. Kostelec, "FFTs for the 2-sphere – Improvements and Variations", *Journal of Fourier analysis and applications* 9, 341-385 (2003).
- [Hear-04] D. Hear, M. P. Baker, *Computer Graphics with OpenGL, 3rd edition*, Prentice Hall (2004)
- [Hecht-01] E. Hecht, *Optics 4th edition*, Addison-Wesley (2001)
- [Hong-06] S. H. Hong and B. Javidi, "Distortion-tolerant 3D recognition of occluded objects using computational integral imaging" *Opt. Express* 14, 12085-12095 (2006)
- [Horner-84] J. L. Horner and P. D. Gianino, "Phase-Only Matched Filtering", *Appl. Opt.* 23, 812 (1984).
- [Hotta-98] K. Hotta, T. Kurita, T. Mishima, "Scale Invariant Face Detection Method Using Higher-Order Local Autocorrelation Features Extracted from Log-Polar Image", *Proceedings of Third IEEE*

- International Conference on Face & Gesture Recognition, 70-75 (1998)
- [Hyvärinen-01] A. Hyvärinen, J. Karhunen, E. Oja, *Independent Component Analysis*, John Wiley & Sons (2001)
- [Hsu-82a] Y. N. Hsu, H. H. Arsenault, “Optical recognition using circular harmonic expansion”, *Appl. Opt.* 21, 4016-4019 (1982)
- [Hsu-82b] Y. N. Hsu, H. H. Arsenault and G. Abril, “Rotational-Invariant digital pattern recognition using circular harmonic expansion”, *Appl. Opt.* 21, 4012-4015 (1982)
- [Ishiyama-07] R. Ishiyama, M. Hamanaka and S. Sakamoto, “Face recognition under variable pose and illumination conditions using 3D facial appearance models”, *Systems and Computers in Japan* 38, 57-70 (2007).
- [Jackson-99] J. D. Jackson, *Classical Electrodynamics*, Wiley (1999)
- [Jacobs-97] D. W. Jacobs, “Matching 3-D Models to 2-D Images”, *International Journal of Computer Vision* 21, 123–153 (1997)
- [Javidi-00] B. Javidi and E. Tajahuerce, “Three-dimensional object recognition by use of digital holography”, *Opt. Lett.* 25, 610–612 (2000).
- [Javidi-02] B. Javidi, ed., *Image Recognition and Classification: Algorithms, Systems, and Applications*, Marcel Dekker, New York (2002)
- [Javidi-05] B. Javidi, I. Moon, S. Yeom, and E. Carapezza, “Three-dimensional imaging and recognition of microorganism using single-exposure on-line

- (SEOL) digital holography”, *Opt. Express*, 13, 4492-4506 (2005).
- [Javidi-06] B. Javidi, S. Yeom, I. Moon, and M. Daneshpanah “Real-time automated 3D sensing, detection, and recognition of dynamic biological micro-organic events”, *Opt. Express* 14, 3806-3829 (2006)
- [Javidi-07] F. Sadjadi, B. Javidi, (eds.), *Physics of Automatic Target Recognition*, Springer (2007)
- [Karhunen-47] K. Karhunen, Kari, “Über lineare Methoden in der Wahrscheinlichkeitsrechnung” *Ann. Acad. Sci. Fennicae. Ser. A. I. Math.-Phys.* 37, 1-79 (1947)
- [Khazdan-03] M. Kazhdan, T. Funkhouser and S. Rusinkiewicz, “Rotation invariant spherical harmonic representation of 3D shape descriptors”, *Symposium on Geometry Processing*, 167-175 (2003)
- [Kim-04] D. Kim and B. Javidi, “Distortion-tolerant 3-D object recognition by using single exposure on-axis digital holography”, *Opt. Express* 12, 5539-5547 (2004)
- [Kittler-05] J. Kittler, A. Hilton, M. Hamouz and J. Illingworth, “3D Assisted Face Recognition: A Survey of 3D Imaging, Modelling and Recognition Approaches”, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPRW'05), Computer Vision and Pattern Recognition Workshop 0*, 114 (2005).
- [Konforti-90] N. Konforti, D. Mendlovic and E. Marom, “Real time optical generation of circular or Mellin

- harmonics filters”, J. Opt. Soc. Am. 7, 225-300 (1990)
- [Kostelec-03] Peter J. Kostelec and Daniel N. Rockmore, SOFT: SO(3) Fourier Transforms, <http://www.cs.dartmouth.edu/~geelong/soft/>
- [Kostrikin-78] A. I. Kostrikin, *Introducción al algebra*, Editorial Mir (1978)
- [Kumar-90] B.V. K. V. Kumar and L. Hassebrook, “Performance measures for correlation filters”, Appl. Opt. 29, 2997-3006 (1990)
- [Laurin-95] Laurin D.G., Rioux M., “3-Dimensional Object tracking using a Fourier-Transform of sine-coded range images”, Opt. Eng. 34, 1789-1798 (1995).
- [Lefebvre-02] D. Lefebvre, H. H. Arsenault, P. Garcia-Martinez, and C. Ferreira, “Recognition of unsegmented targets invariant under transformations of intensity,” Appl. Opt. 41, 6135–6142 (2002).
- [Lefebvre-03] D. Lefebvre, H. H. Arsenault, and S. Roy, “Nonlinear filter for pattern recognition invariant to illumination and to out-of plane rotations,” Appl. Opt. 42, 4658–4662 (2003).
- [Li-00] Y. Li and J. Rosen, “Three-Dimensional Pattern Recognition with a Single Two-Dimensional Synthetic Reference Function”, Appl. Opt. 39, 1251-1259 (2000)
- [Li-02] Y. Li and J. Rosen, “Object recognition using three-dimensional optical quasi-correlation,” J. Opt. Soc. Am. A 19, 1755–1762 (2002).

- [Li-03] Y. Li and J. Rosen, "Scale invariant recognition of three-dimensional objects by use of a quasi-correlator," *Appl. Opt.* 42, 811-819 (2003).
- [Li-04] W. J. Li, C. J. Wang, D. X. Xu and S. F. Chen, "Illumination Invariant Face Recognition Based on Neural Network Ensemble", *Proceedings of the 16th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2004)* 15, 486-490 (2004)
- [Li-07] S. Z. Li, R. Chu, S. Liao and L. Zhang, "Illumination Invariant Face Recognition Using Near-Infrared Images", *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29, 627-639 (2007)
- [Loeve-78] M. Loève, *Probability theory. Vol. II, 4th ed., Graduate Texts in Mathematics, Vol. 46*, Springer-Verlag (1978)
- [Lohmann-67] A. W. Lohmann and D. P. Paris, "Binary Fraunhofer holograms generated by computer", *Appl. Opt.* 6, 1739-1748 (1967)
- [Makadia-03] A. Makadia, K. Daniilidis, "Direct3D-Rotation Estimation from Spherical Images via a Generalized Shift Theorem", *Proceedings of 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 2, 217-226 (2003)
- [Maloney-86] L. T. Maloney and B. A. Wandell, "Color constancy: a method for recovering surface spectral reflectance", *J. Opt. Soc. Am. A* 3, 29- (1986)

- [Marion-75] J. B. Marion, *Dinámica clásica de las partículas y sistemas*, Academia Press, New York (1975)
- [Mathworld-99a] <http://mathworld.wolfram.com/DiscreteFourierTransform.html>
- [Mathworld-99b] <http://mathworld.wolfram.com/LegendrePolynomial.html>
- [Mathworld-02a] <http://mathworld.wolfram.com/VectorSpaceProjection.html>
- [Mathworld-02b] <http://mathworld.wolfram.com/IntegralKernel.html>
- [Mathworld-06] <http://mathworld.wolfram.com/GramSchmidtOrthogonalization.html>
- [Matoba-01] O. Matoba, E. Tajahuerce, and B. Javidi, “Real-time threedimensional object recognition with multiple perspectives imaging,” *Appl. Opt.* 40, 3318–3325 (2001).
- [Mendlovic-88] D. Mendlovic, E. Marom and N. Konforti, “Shift and scale invariant pattern recognition using Mellin radial harmonics”, *Opt. Comm.* 67, 172-176. (1988)
- [Mendlovic-89] D. Mendlovic, N. Konforti, and E. Marom, “Scale and projection invariant pattern recognition” , *Appl. Opt.* 28, 4982-4986 (1989)
- [Mendlovic-95] D. Mendlovic, Z. Zalevsky, I. Kiryuschev, and G. Lebreton, “Composite harmonic filters for scale-, projection-, and shift-invariant pattern recognition”, *App. Opt.* 34, 310-316 (1995)
- [Mersereau-86] K. Mersereau and G. M. Morris, “Scale, rotation, and shift invariant image recognition”, *Appl. Opt.* 25, 2338-2342 (1986)

- [Moya-93] A. Moya, E. Tajahuerce, J. García, D. Mendlovic and C. Ferreira, "Method for determining the proper expansion center and order for Mellin radial harmonic filters", *Opt. Commun.* 103, 39-45 (1993)
- [Moya-97] A. Moya, "Aportaciones en reconocimiento de formas invariante a escala y proyecciones", Tesis Doctoral, Universitat de València, Diciembre 1997
- [Moya-00] A. Moya, J. J. Esteve-Taboada, J. García, and C. Ferreira, "Shift- and Scale-Invariant Recognition of Contour Objects with Logarithmic Radial Harmonic Filters", *Appl. Opt.* 39, 5347-5352 (2000)
- [Nayar-96] S. K. Nayar, H. Murase, H., "Dimensionality of illumination in appearance matching", *Proceedings of 1996 IEEE International Conference on Robotics and Automation 2*, 1326-1332 (1996)
- [OpenGL-05] OpenGL Architecture Review Board, *OpenGL(R) Programming Guide: The Official Guide to Learning OpenGL(R), Version 2*, Addison-Wesley Professional (2005)
- [Paquet-95] E. Paquet, M. Rioux and H. H. Arsenault, "Invariant pattern recognition for range images using the phase Fourier transform and a neural network," *Opt. Eng.* 34, 1178-1183 (1995).
- [Paulpandian-93] T. Paulpandian, V. Ganapathy, "Translation and scale invariant recognition of handwritten Tamil characters using a hierarchical neural network", *Proceedings of 1993 IEEE International*

- Symposium on Circuits and Systems 3, 2439-2441 (1993)
- [Pentland-94] A. Pentland, B. Moghaddam, T. Starner, "View-based and modular eigenspaces for face recognition", Proceedings of 1994 IEEE Computer Society Conference on Computer Vision and Pattern Recognition 84, 21-23 (1994)
- [Perantonis-92] S. J. Perantonis and P. J. G. Lisboa, "Translation, rotation, and scale invariant pattern recognition by high-order neural networks and moment classifiers," IEEE Trans. Neural Networks 3, pp. 241–251 (1992).
- [Petrich-00] W. Petrich, B. Dolenko, J. Frh, M. Ganz, H. Greger, S. Jacob, F. Keller, Alexander E. Nikulin, M. Otto, O. Quarder, Ray L. Somorjai, A. Staib, Gerhard Werner, and Hans Wielinger "Disease Pattern Recognition in Infrared Spectra of Human Sera with Diabetes Mellitus as an Example", Appl. Opt. 39, 3372-3379 (2000)
- [Poon-99] T. Poon and T. Kim, "Optical image recognition of three-dimensional objects" Appl. Opt. 38, 370-381 (1999).
- [Pu-99] A. Pu, R. Denkewalter and D. Psaltis "Real-time vehicle navigation using a holographic memory", Opt. Eng. 36, 2737-2746 (1997)
- [Sheng-87] Y. Sheng y H. H. Arsenault, "Method for determining expansion centers and predicting sidelobe levels for circular harmonic filters", J. Opt. Soc. Am. A 4, 1793-1799 (1987)

- [Rau-66] J. E. Rau, "Detection of differences in real distributions", *J. Opt. Soc. Am.* 56, 1490-1494 (1966)
- [Réfrégier-04] P. Réfrégier, *Noise Theory and Application to Physics*, Springer (2004)
- [Rioux-84] M. Rioux, "Laser range finder based on synchronized scanners," *Appl. Opt.* 23, 3837-3844 (1984).
- [Rosen-89] J. Rosen and J. Shamir, "Scale-invariant pattern recognition with logarithmic radial harmonic filters", *App. Opt.* 28, 240-244 (1989)
- [Rosen-98a] J. Rosen, "Three-dimensional electro-optical correlation," *J. Opt. Soc. Am. A* 15, 430-436 (1998)
- [Rosen-98b] J. Rosen, "Three-dimensional joint transform correlator," *Appl. Opt.* 37, 7538-7544 (1998)
- [Roy-04] S. Roy, D. Lefebvre, and H. H. Arsenault, "Recognition invariant under unknown affine transformations of intensity," *Opt. Commun.* 238, 69-77 (2004)
- [Sakurai-85] J. J. Sakurai, *Modern Quantum Mechanics*, Addison-Wesley (1985).
- [Sheng-86a] Y. Sheng and H. H. Arsenault, "Experiments on pattern recognition using invariant Fourier-Mellin descriptors", *J. Opt. Soc. Am. A* 3, 771-776 (1986)
- [Sheng-86b] Y. Sheng and J. Duvernoy, "Circular-Fourier-radial-Mellin transform descriptors for pattern recognition", *J. Opt. Soc. Am. A* 3, 885-888 (1986)

- [Sheng-87] Y. Shen and H. H. Arsenault, "Method for determining expansion centers and predicting sidelobe levels for circular-harmonic filters", J. Opt. Soc. Am. A 4, 1793-1797 (1987)
- [Sutherland-92] J. G. Sutherland, *The holographic neural method in fuzzy, holographic and parallel intelligence*, Wiley (1992)
- [Szoplik-85a] T. Szoplik, "Shift and scale invariant anamorphic correlator", J. Opt. Soc. Am. A 2, 1419-1423 (1985)
- [Szoplik-85b] T. Szoplik and H. H. Arsenault, "Shift and scale invariant anamorphic correlator using multiple circular harmonic filters", Appl. Opt. 24, 3179-3183 (1985)
- [Tadeusz-98] D. Tadeusz and D. D. Ewa, "Application of Neural Networks in 3D Object Recognition System", Pattern Recognition and Artificial Intelligence 12, 491-504 (1998)
- [Tajahuerce-94] E. Tajahuerce, A. Moya, J. García and C. Ferreira, "Real filter based on Mellin radial harmonics for scale-invariant pattern-recognition", Appl. Opt. 33, 3086-3093 (1994)
- [Tajahuerce-01] E. Tajahuerce, O. Matoba, and B. Javidi, "Shift-invariant three-dimensional object recognition by means of digital holography," Appl. Opt. 40, 3877-3886 (2001).
- [Tajahuerce-94] E. Tajahuerce, A. Moya, J. García and C. Ferreira, "Real filter based on Mellin radial harmonics for scale-invariant pattern recognition", Appl. Opt. 33, 3086-3093 (1994)

- [Takatsuka-98] M. Takatsuka, R.A. Jarvis, “Hierarchical Neural Networks for Learning 3-Dimensional Objects from Range Images”, *Journal of Electronic Imaging* 7, 16-28 (1998)
- [Takeda-83] M. Takeda and K. Mutoh, “Fourier transform profilometry for the automatic measurement of 3-D object shapes,” *Appl. Opt.* 22, 3977-3982 (1983)
- [Tan-98] T. N. Tan, “Locating and recognizing road vehicles”, *Opt. Eng.* 37, 202-207 (1998)
- [Tian-99] Y. Tian, H. T. Tsui, S. Y. Yeung, and S. Ma, “Shape from shading for multiple light sources,” *J. Opt. Soc. Am. A.* 16, 36–52 (1999)
- [Torres-Mendez-00] L. A. Torres-Mendez, J. C. Ruiz-Suarez, L. E. Sucar, and G. Gomez, “Translation, rotation, and scale-invariant object recognition”, *IEEE Trans. Systems, Man, and Cybernetics, Part C*, 30, 125-130 (2000)
- [Turk-91a] M. Turk and A. Pentland, “Face recognition using eigenfaces”. *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 3, 586–591 (1991)
- [Turk-91b] M. Turk and A. Pentland, “Eigenfaces for recognition”, *Journal of Cognitive Neuroscience* 3, 71–86 (1991).
- [VanderLugt-64] A. VanderLugt, “Signal detection by complex spatial filtering” *IEEE Trans. Inform. Theory* 2, 139-45 (1964)
- [Van Etten-05] Wim C. Van Etten, *Introduction to Random Signals and Noise*, Wiley (2005)

- [Vallés-04] J. J. Vallés, “Estimación de rotaciones de objetos tridimensionales”, Trabajo de investigación, Universitat de València (2004)
- [Vallés-06] J. J. Vallés, J. García, P. García-Martínez, and Henri H. Arsenault, “Three-dimensional object detection under arbitrary lighting conditions”, *Appl. Opt.* 45, 5237-5247 (2006)
- [Vallés-07] J. J. Vallés, P. Garcia-Martinez, J. García, and C. Ferreira, “Phase Fourier vector model for scale invariant three-dimensional image detection”, *Opt. Express* 15, 7818-7825 (2007)
- [Vallés-08] J. J. Vallés, P. García-Martínez, and C. Ferreira, “Spherical nonlinear correlations for global invariant three-dimensional object recognition”, *Appl. Opt.* 47, A43-A51 (2007)
- [Walker-96] J. S. Walker, *Fast Fourier Transform, 2nd ed*, CRC Press, Boca Raton, FL (1996).
- [Wandelt-01] B. Wandelt and K. Gorski, “Fast convolution on the sphere”, *Physical reviews D* 63, 123002 (2001)
- [Watanabe-85] S. Watanabe, *Pattern recognition: Human and mechanical*, Wiley (1985)
- [Weaver-66] C. S. Weaver and J. W. Goodman, “A technique for optically convolving two functions”, *Appl. Opt.* 5, 1248-1249 (1966)
- [Wikipedia-07] http://en.wikipedia.org/wiki/Principal_component_analysis
- [Yang-05] J. Yang and M. Al-Rawi, “Illumination invariant recognition of three-dimensional texture in color images”, *J. Comput. Sci. Technol.* 20, 378-388. (2005)

- [Zalevsky-97] Z. Zalevsky, D. Mendlovic, and J. García, “Invariant pattern recognition by use of wavelength multiplexing”, *Appl. Opt.* 36, 1059-1063 (1997)
- [Zhang-99] S. Zhang and M. A. Karim, “Illumination-Invariant Pattern Recognition with Joint-Transform-Correlator-Based Morphological Correlation”, *Appl. Opt.* 38, 7228-7237 (1999)
- [Zhang-03] L. Zhang and D. Samaras, “Face Recognition Under Variable Lighting using Harmonic Image Exemplars”, *Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 3, 351-363 (2003)
- [Zhou-05] S. K. Zhou and R. Chellappa, “Image-based face recognition under illumination and pose variations,” *J. Opt. Soc. Am. A.* 22, 217–229 (2005).
- [Zwillinger-96] D. Zwillinger, *Standard Mathematical Tables and Formulae*, CRC press (1996)

Apéndice I

Abreviaturas usadas en la memoria

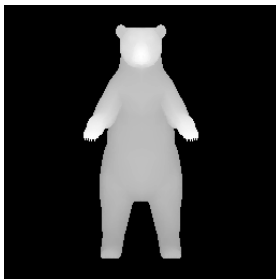
A continuación, indicamos las abreviaturas correspondientes a distintos aspectos presentes en esta memoria, y que, a lo largo de la misma, han sido usadas en mayor o menor medida:

2D:	Bidimensional
3D:	Tridimensional
3DOOM:	Three dimensional object orientation map
FFT:	Fast Fourier transform
LACIF:	Locally adaptive contrast invariant filter
LADC:	Local angular distance correlation
PCE:	Peak to correlation energy
PhFT:	Phase Fourier transform
POF:	Phase only filter
\mathbb{R}^2 :	Espacio de números reales de dimensión 2
S^2 :	Espacio sobre la superficie de una esfera
SNR:	Signal to noise ratio
SO(3):	Grupo especial ortogonal de grado 3

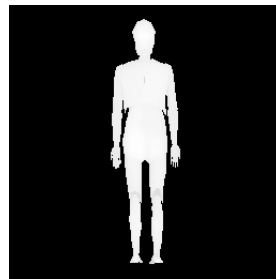
Apéndice II

Objetos 3D usados

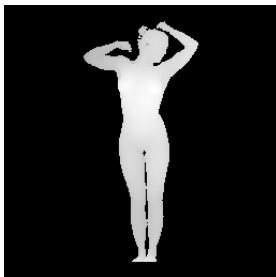
Todos los objetos usados en este trabajo se encuentran a la disposición del público en diferentes páginas web. A continuación mostramos los diferentes objetos tridimensionales que se han usado a lo largo de esta memoria:



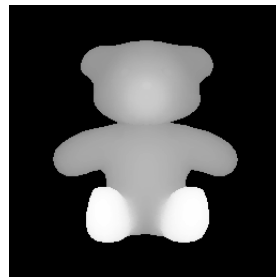
“Oso”



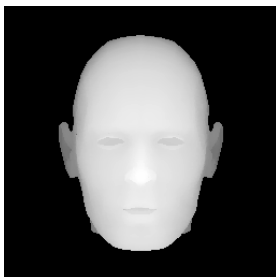
“Test 1”



“Test 2”



“Test 3”



“Cara 1”



“Cara 2”

Anexo

Recopilación de publicaciones

Publicación 1

Detection of three-dimensional objects under arbitrary rotations based on range images.

Javier García, José J Valles, Carlos Ferreira

Opt. Express, Vol. 11, Issue 25, pp. 3352-3358, (2003)

Publicación 2

Three-dimensional object detection under arbitrary lighting conditions.

José J. Vallés, Javier García, Pascuala García-Martínez and Henri H. Arsenault.

Appl. Opt. , Vol. 45, No. 21, pp. 5237-5247, (2006)

Publicación 3

Phase Fourier vector model for scale invariant three-dimensional image detection.

José J. Vallés, Pascuala Garcia-Martinez, Javier García, and Carlos Ferreira.

Opt. Express, Vol. 15, Issue 12, pp. 7818-7825, (2007)

Publicación 4

Spherical nonlinear correlations for global invariant three-dimensional object recognition.

José J. Vallés, Pascuala Garcia-Martinez, and Carlos Ferreira.

Appl. Opt., Vol. 47, No. 41, pp. A43-A51, (2008)

Detection of three-dimensional objects under arbitrary rotations based on range images

Javier García, Jose J. Valles, and Carlos Ferreira

Departamento de Optica. Universitat de Valencia. Calle Dr. Moliner 50 Burjassot Spain
javier.garcia.monreal@uv.es

Abstract: In this paper a unique map or signature of three dimensional objects is defined. The map is obtained locally, for every possible rotation of the object, by the Fourier transform of the phase-encoded range-image at each specific rotation. From these local maps, a global map of orientations is built that contains the information about the surface normals of the object. The map is defined on a unit radius sphere and permits, by correlation techniques, the detection and orientation evaluation of three dimensional objects with three axis translation invariance from a single range image.

©2003 Optical Society of America

OCIS codes: (100.6890) Three-dimensional image processing; (070.5010) Pattern recognition and feature extraction; (070.2590) Fourier transforms

References and links

1. R. Campbell and P. Flynn, "A survey of free-form object representation and recognition techniques," *Computer Vision and Image Understanding* **81** 2 (2001), pp. 166–210.
2. M. Rioux, "Laser range finder based on synchronized scanners," *Appl. Opt.* **23**, 3837-3844 (1984).
3. M. Takeda and K. Mutoh, "Fourier transform profilometry for the automatic measurement of 3-D object shapes," *Appl. Opt.* **22**, 3977-3982 (1983).
4. J. Rosen, "Three-dimensional electro-optical correlation," *J. Opt. Soc. Am.* **A 15**, 430-436 (1998).
5. T. Poon and T. Kim, "Optical image recognition of three-dimensional objects," *Appl. Opt.* **38**, 370-381 (1999).
6. E. Tajahuerce, O. Matoba, and B. Javidi, "Shift-Invariant Three-Dimensional Object Recognition by Means of Digital Holography," *App. Opt.* **40**, 3877-3886 (2001)
7. E Paquet, H H Arsenault and M Rioux , "Recognition of faces from range images by means of the phase Fourier transform," *Pure Appl. Opt.* **4**, 709-721 (1995).
8. P. Parrein, J. Taboury, P. Chavel, "Evaluation of the shape conformity using correlation of range images," *Opt. Commun.* **195** (5-6), 393-397 (2001).
9. Huber DF, Hebert M, "Fully automatic registration of multiple 3D data sets," *Image And Vision Computing* **21** (7): 637-650 (2003)
10. M. Rioux, P. Boulanger and T. Kasvand , "Segmentation of range images using sine wave coding and Fourier transformation," *App. Opt.* **26**, 287-292 (1987).
11. E. Paquet, M. Rioux and H. H. Arsenault, "Range image segmentation using the Fourier transform," *Opt. Eng.* **32**, 2173-2180 (1994)
12. J. J. Esteve-Taboada, D. Mas, and J. García, "Three-dimensional object recognition by Fourier transform profilometry," *App. Opt.* **38**, 4760-4765 (1999).
13. Esteve-Taboada JJ, García J, Ferreira C, "Rotation-invariant optical recognition of three-dimensional objects," *App. Opt.* **39**, 5998-6005 (2000)
14. Hsu Y., Arsenault HH, "Optical-pattern recognition using circular harmonic expansion," *App. Opt.* **21**, 4016-4019 (1982)
15. Chang S, Rioux M, Grover CP, "Range face recognition based on the phase Fourier transform," *Opt. Commun.* **222**, 143-153 (2003)
16. Hassebrook LG, Lhamon ME, Wang M, Chatterjee JP, "Postprocessing of correlation for orientation estimation," *Opt. Eng.* **36**, 2710-2718 (1997)
17. J. J. Esteve-Taboada and J. García, "Detection and orientation evaluation for three-dimensional objects," *Opt. Com.* **217**, 123-131 (2002).
18. <http://scienceworld.wolfram.com/astronomy/EquatorialCoordinates.html>
19. B. D. Wandelt and K. M. Görski, "Fast convolution on the sphere," *Phys. Rev. D* **63**, 123002 (2001).
20. J. R. Driscoll and D. M. Healy, "Computing Fourier transforms and convolutions on the 2-Sphere," *Adv. In App. Math.* **15**, 202-250 (1994).

1. Introduction

A considerable effort has been devoted in the last years to optical digitizing, processing and recognition of three dimensional (3D) objects [1-8]. In general, the information of a 3D object is contained in a set of triplets with the (x, y, z) coordinates of the object surface. Detecting an object means scanning over the 6 degrees of freedom (3 rotations and 3 translations) of the object for the best matching with the reference set of points. A significant reduction of the complexity of the problem can be achieved storing a few views of the object as planar images. For instance, a range image can contain the full 3D geometrical information, although only from a given point of view. With a sufficient amount of overlapping between views the full 3D object information can be stored in a few images [9]. Also, range images can be obtained from most of the 3D digitizing techniques [2, 3]. A main limitation of most 3D matching methods is the limited tolerance to rotation in the object. This problem can only be addressed by adding a significant simplification to the problem or by very heavy calculation (see reference [1] for a survey of some digital techniques).

In this paper we propose and demonstrate a method for detection and orientation evaluation of three dimensional objects that have undergone an arbitrary rotation in space. The method is based on computing a map or signature of the three dimensional object based on the local orientation of its surfaces, by means of range images. A matching between this map and the information extracted from a range image from an arbitrary point of view provides a means for detection and, also, for estimation of the object orientation in space.

2. Fourier transforms of phase encoded range images

A range image, $z=f(x, y)$, contains the depth information of an object from a given view line, that defines the z axis. Note that the range image is a single valued function, therefore only the part of the surface closer to the positive z axis is contained on it. The encoding of the depth information has been used in the literature to extend the possibilities of range images [7, 10, 11]. Following this approach, we encode the range image as phase as follows [7]:

$$P(x, y) = \exp[iwz(x, y)] \quad (1)$$

where w is a constant that permits the adjustment of the phase slope of the object.

A way to deal with range images keeping translation invariance is to use their Fourier transform. The Fourier transform of the phase encoded range image (*PhFT*) is then:

$$PhFT(u, v) = F_{2D} \{ \exp[iwz(x, y)] \} \quad (2)$$

where F_{2D} stands for two dimensional Fourier transform. Note that w determines the scaling of the Fourier transform frequencies. From now, we will assume $w=1$, that correspond to the PhFT mapping detailed in the following.

A planar surface of the object, after phase encoding, will become a linear phase factor. Thus its Fourier transform will be peaked around a well defined location. Referring to figure 1(a), the z axis defines the view line, the polar axis is the y axis and the azimuth angle is measured with respect to the z axis. A facet of the object is determined by the angles (α_x, α_y) , that are the angles of the z axis with the projection of the normal to the planar surface. The tangents of these angles are proportional to the location of the peak in the Fourier domain:

$$(u, v) \equiv (\tan(\alpha_x)/2\pi, \tan(\alpha_y)/2\pi) \quad (3)$$

Based on this expression we can relate the peak location with the orientation of the normal in conventional spherical angles, (θ, φ) as:

$$(u, v) = \left(\frac{\tan(\varphi)}{2\pi} \quad \frac{\tan(\theta)}{2\pi \cos(\varphi)} \right). \quad (4)$$

These relations correspond to a value of $w=1$. A lower value will concentrate the information in low frequencies, while a higher one will expand the scale of the PhFT. From Eq. (4), the PhFT can be coordinate transformed to obtain a distribution $PhFT_{sph}(\theta, \varphi)$, expressed in spherical coordinates, where the location of a peak gives directly the orientation of the corresponding facet. In general, for non-planar surfaces, the PhFT will contain the information of the orientations of the surfaces in the object. Figs. 1(b) and (c) show a sample object and the corresponding PhFT expressed in angular coordinates. It is worth noting that the angular variation for θ and φ has maximum span of ± 45 degrees. For higher angles the energy contents of a surface will be small because the apparent size of the facet will be reduced.

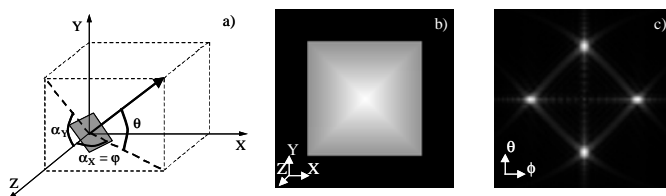


Fig. 1. (a) Definition of angular coordinates. (b) Range image of a pyramid shaped object. (c) PhFT intensity.

The intensity of the PhFT exhibit a crucial property: it is invariant to arbitrary translations of the object. This property is obvious for translations in the (x,y) plane, as they will produce just a linear phase factor in the PhFT. On the other hand, from the definition of the phase encoded range image (Eq. 1), a shift along the view line (z axis) will influence just as a constant phase factor. In both cases the PhFT is just altered by a phase that is irrelevant in intensity. This property makes the PhFT advantageous for 3D object matching, because the translation invariance is automatically obtained, when performing the correlation between the intensities of the PhFTs [8, 12].

3. Correlation under limited rotations

The problem of rotations is by far more complex than the shifts. Two particular cases have been addressed in the bibliography. In [13] circular harmonic components [14] are used to achieve full rotation invariance under rotations around the view line (z axis). This result is extended in [15] to partial rotations that permits the increase of the information content of the impulse response. The second case involves rotations around an axis perpendicular to the view line [i.e., contained in the (x,y) plane].

In order to analyze this situation we consider two coordinates system. The first one (x,y,z) is attached to the object and rotates with it, whilst the second one (x',y',z') is fixed in space, the z' axis defined by the fixed view line. Assuming a rotation of the object (and its attached coordinate system) around the y' axis of angle ω from the definition of angular coordinates, it is obvious that a normal defined by (θ, φ) in the rotated system will be represented by the angles $(\theta', \varphi') = (\theta, \varphi + \omega)$ in the fixed system [See Fig. 1(a)]. Therefore, $PhFT(\theta', \varphi')$, expressed in angular coordinates, will undergo just a displacement. Even in spatial frequencies, (u, v) , at first order approximation of ω the variations are linear with the rotation angle ω , as can be checked using Eq. (4). A rotation around any axis contained in the (x,y) plane can be reduced to this case simply by choosing the y axis as the rotation axis. This property has been used to obtain detection of 3D objects by correlation of the PhFT intensity with limited tolerance to rotations with respect to an axis perpendicular to the view line [15].

If the full 3D information of the object is known, it is possible to calculate in advance the intensity of the PhFT for any possible orientation. This makes it possible to build a map (3D object orientation map or 3DOOM) that contains the information of all view points by displacing and pasting on a large image the PhFT expressed in spherical coordinates. This 3DOOM contains the information about the normals for all possible angles in a single image, not only around a given view line. This approach is similar to that of a bank of filters [16], although there is a partial overlapping between the impulse responses at different angles. This approach was followed in [17] to achieve detection and orientation estimation with rotations around an axis perpendicular to the view line. In this work the 3DOOM was limited to a region around the equator (θ smaller than 45 degrees), because larger nodding of the object was not permitted and thus no information of the facet pointing to the poles was available.

3. Object orientations map on the unit sphere

A first step to extend the detection capability to any arbitrary rotation is to obtain a 3DOOM that covers the full angular range. This step involves the rotation of the model object scanning the possible orientations in such a way that any facet is facing the view line for at least one orientation. Alternatively we can consider the scan on the view line keeping the object fixed. An arbitrary rotation is usually described by the Euler angles. Figure 2(a) shows the convention that we are assuming in the following. With respect to conventional definition, the axes have been permuted cyclically, for a better correspondence with the previous spherical coordinates definition, which assumes that the y axis is the polar one. The arbitrary rotation is decomposed in (a) a rotation of angle α around the y axis (b) a rotation of angle β around the rotated x axis (c) a rotation of angle γ around the new z axis.

In matrix notation this decomposition is expressed as

$$R(\alpha, \beta, \gamma) \equiv R_Y(\gamma)R_{X'}(\beta)R_Y(\alpha). \quad (5)$$

Note that the last rotation (of angle γ) is not needed for scanning over all (θ, ϕ) range. The procedure for building the 3DOOM starts by preparing an image for the full (θ, ϕ) range. Then we scan the view line for all angles (θ, ϕ) , by means of the two rotations around the y and x' axes. For every angle of the view line the following procedure is performed: (a) the range image is computed (b) the PhFT is obtained in the coordinates of the rotated system (θ', ϕ') (c) the Intensity of the PhFT is coordinate transformed into the object axes (θ, ϕ) (d) The resulting image is pasted (by averaging) on the full 3DOOM image. Figure 2 shows a sample 3D image and the corresponding map in spherical angles.

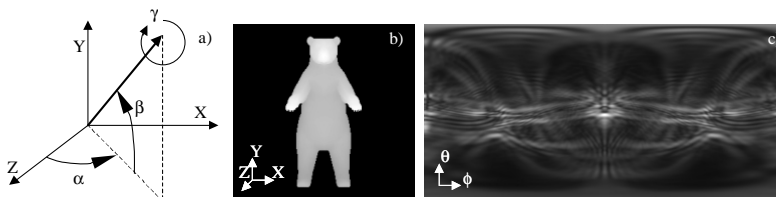


Fig. 2. (a) Definition of Euler angles. (b) Range image of an object. (c) 3DOOM on θ - ϕ Coordinates.

Once the 3DOOM is obtained the task of detection of an object at an arbitrary angular orientation consist on a matching of the intensity of the PhFT for this orientation with the 3DOOM. Unfortunately, although the 3DOOM has been up to now represented as a planar image, it correspond to angular variables, and problems of representation, well known in cartography, arise. In particular, the choice of the polar axis makes that the flat map will be

greatly distorted for orientations close to it. The problem derives from the coordinate transformation from local spherical coordinates (θ', φ') , where the PhFT is obtained, into object spherical coordinates (θ, φ) , that are the variables for the 3DOOM. This coordinate transformation is highly non linear [18], except for the case of rotations around y axis, as previously discussed. The distortion is indeed an artifact due to the representation of the 3DOOM as a planar map. As in the case of cartography there is no singular point in the representation if the 3DOOM is represented on a unit radius sphere.

3. Correlations on the unit sphere

In a conventional two dimensional correlation, the images to be correlated and the resulting correlation are 2D distributions. The correlation is performed by scanning the spatial coordinates for the matching of one of the two distributions on the other. The scanning is 2D, what defines the dimensionality of the output correlation. The problem is quite different for correlations on the unit sphere. The scan for matching has to be performed on the group of rotations on the sphere (SO3), as described by Euler angles, defining a 3D correlation output. Therefore, the problem is to obtain the correlation between two functions expressed in spherical coordinates. This is equivalent to define the matching between two functions defined on a unit radius sphere, namely $PhFT(\theta, \varphi')$ and the 3DOOM. Figure 3 shows an example of this statement of the problem of correlations on the unit sphere

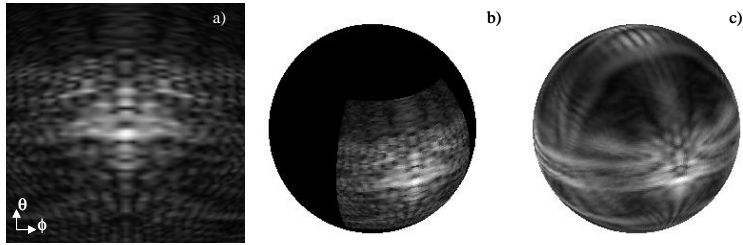


Fig. 3. (a) PhFT of the target shown in figure 2 (b). (b) The same PhFT depicted on the unit sphere. (c) 3DOOM on the unit sphere. The correlation consists on the matching of distributions on figures (b) and (c)

A mathematical definition of the correlation on the unit sphere is [19]:

$$T(\alpha, \beta, \gamma)_{\in SO(3)} = f(\theta, \varphi) * g(\theta, \varphi) \equiv \int [D(\alpha, \beta, \gamma) f](\theta, \varphi) g(\theta, \varphi) \sin(\theta) d\theta d\varphi. \quad (6)$$

where $D(\alpha, \beta, \gamma)$ is the rotation characterized by the proper Euler angles. Note that is not the only definition in the literature. In [20] the correlation is defined for objects with symmetry around the view line (or equivalently integrating on last rotation angle).

As in the planar case, it is possible to perform the correlation on the frequency domain, using Fourier transforms. On the unit sphere the Fourier expansion involves the use of spherical harmonics, instead of linear exponential functions. The Fourier transform of an object in terms of spherical harmonics expansion is defined as [19]:

$$f_{lm} = \int f(\theta, \varphi) Y_{lm}^*(\theta, \varphi) \sin(\theta) d\theta d\varphi. \quad (7)$$

where Y_{lm} are the spherical harmonics functions. Following Wendelt *et al.* [19], the correlation will be defined as:

$$T(\alpha, \beta, \gamma) = \sum_{m, m', m''} T_{mm'm''} e^{im\alpha + im'\beta + im''\gamma}. \quad (8)$$

$$T_{m'm''} = \sum_l g_{lm} d_{mm'}^l \left(\frac{\pi}{2} \right) d_{m''m'}^l \left(\frac{\pi}{2} \right) f_{lm''}^* \quad (9)$$

Where d_{mm}^l are the reduced rotation matrices [21] and f_{lm} and g_{lm} are the spherical Fourier transforms. In our case f and g are the PhFT from a given (unknown) point of view and the 3DOOM, respectively. Therefore, following this definition, the output of the correlation will be three dimensional, and the location of the correlation peak will give the (α, β, γ) triplet with the object orientation. This correlation admits a rapid implementation using the fast Fourier transform algorithm [20].

To test the performance of this method of 3D object matching we have implemented it digitally. In order to simplify the interpretation of the results, without losing generality, the rotations are limited to the two first rotations (the rotation around the final polar axis is not permitted). This way the correlation output is bidimensional and can be directly represented.

The target in our tests is the one depicted in Figure 2(b). Figures 4 (a-c) show a range image corresponding to an arbitrary orientation of the target, the corresponding intensity of the PhFT and its correlation with the 3DOOM [see Fig. 3(c)]. The correlation has been represented as a planar image. The correlation shows a distinct correlation peak from which the orientation of the object ($\alpha = -120^\circ$, $\beta = 53^\circ$) can be estimated. It is clear that the method provide a correlation that can sharply determine the presence and the orientation of the object.

The discrimination capabilities of the method have been tested by using another object shown on Figure 4(d). The PhFT for an arbitrary orientation [Fig. 4(e)] along with the correlation the correlation, with the same intensity normalization as the correlations for the true target, [Fig. 4(f)] are shown. The low output correlation value discriminates this object from the target. Additional non-shown simulations demonstrate that this discrimination happens for all values of rotations of the false target.

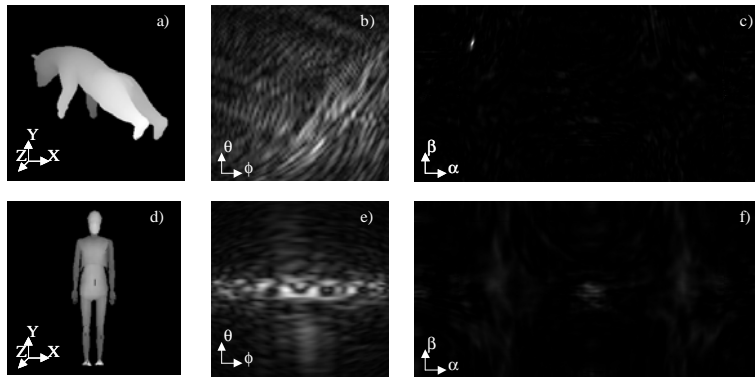


Fig. 4. (851KB) (a) Range image of the 3-D object rotated with $\alpha=-120^\circ$, $\beta=53.4^\circ$. (b) PhFT of the range image given in (a). (c) Output of the correlation. (d) Range image of another object. (e) PhFT of the range image given in (d). (f) Output of the correlation.

4. Conclusions

We have proposed and demonstrated a method for detection of three dimensional objects. The method is based on constructing a map of orientations of the object on a unit sphere using the intensity of the PhFT of range images. The results prove the capability for detection as well as to estimate the orientation of the detected object.

Acknowledgments

This work was supported by FEDER funds and the Spanish Ministerio de Ciencia y Tecnología under the project BFM2001-3004. Jose J. Vallés acknowledges the grant from the Spanish Ministerio de Educación, Cultura y Deporte.

Three-dimensional object detection under arbitrary lighting conditions

José J. Vallés, Javier García, Pascuala García-Martínez, and Henri H. Arsenault

A novel method of 3D object recognition independent of lighting conditions is presented. The recognition model is based on a vector space representation using an orthonormal basis generated by the Lambertian reflectance functions obtained with distant light sources. Changing the lighting conditions corresponds to multiplying the elementary images by a constant factor and because of that, all possible lighting views will be elements that belong to that vector space. The recognition method proposed is based on the calculation of the angle between the vector associated with a certain illuminated 3D object and that subspace. We define the angle in terms of linear correlations to get shift and illumination-invariant detection. © 2006 Optical Society of America

OCIS codes: 070.0070, 070.4550, 100.5010, 100.6890.

1. Introduction

Much effort has recently been devoted to optical digitizing, processing, and recognition of three-dimensional (3D) objects.^{1–8} For example, automatic measurements of 3D object shape can be accomplished using Fourier transform profilometry.³ Such surface measurement techniques have also been used for 3D pattern recognition.⁹ Other 3D recognition methods obtain a range image of the 3D object as an intermediate step of the recognition task.^{7,10,11} Another approach is to use digital holographic techniques,^{6,12} but such methods require a long time to calculate the 3D optical field. Some 3D recognition methods use optical correlations. Bamler and Hofer-Alfeis¹³ showed a method for performing 3D optical correlations by slice-by-slice mapping the 3D observed scene along its longitudinal axis. After mapping, they used conventional 2D optical correlations between each slice and all the others. However, the algorithms to reconstruct the 3D image require intensive computer resources. Rosen

has extended the correlation from 2D to 3D by introducing a 3D optical Fourier transform.⁴ By fusion of several projections of the tested scene, a 3D object function is first Fourier transformed, then filtered by some 3D reference filter, and finally inversely Fourier transformed into the correlation space, so a target can be detected and located in its 3D environment. This highly complex method requires extensive use of digital computation. To reduce the difficulty, the authors proposed an improvement of the 3D correlation space.¹⁴ Matoba *et al.*¹⁵ captured multiple perspectives of 3D objects by a microlens array, after which all the perspectives of the reference and of the input scene were cross correlated by a 2D joint transform correlator (JTC). The authors showed that the system can recognize 3D objects with slightly out-of-plane rotations. In fact, distortion, rotation, and scaling changes of 3D objects are a challenge in many detection systems. Some of the above 3D recognition methods have been extended and new ones were proposed to deal with that automatic target recognition issue.^{16–19} Recently, the addition of object color information in the 3D detection process has been considered.²⁰ The method is based on a multichannel correlation in various chromatic systems based on Fourier-transform profilometry.

Although previous methods dealt with important topics in correlation-based 3D pattern recognition, taking into account changes of illumination of objects has received less attention. It is true that topics related to illumination such as shape from shading, and

J. J. Vallés, J. García, and P. García-Martínez (pascuala.garcia@uv.es) are with the Departament d'Òptica, Universitat de València, C/Doctor Moliner, 50, 46100 Burjassot, Spain. H. H. Arsenault is with the Departement de Physique, Génie Physique et Optique, Université Laval, Quebec, G1K 7P4, Canada.

Received 21 November 2005; accepted 30 January 2006; posted 14 February 2006 (Doc. ID 66083).

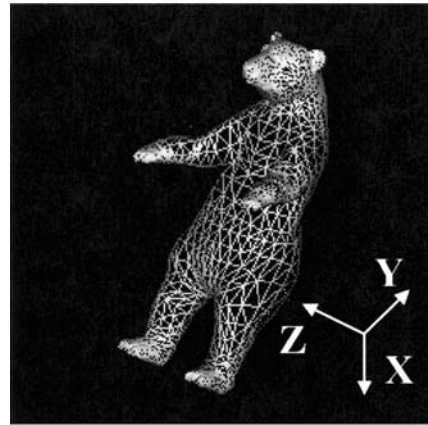
0003-6935/06/215237-11\$15.00/0

© 2006 Optical Society of America

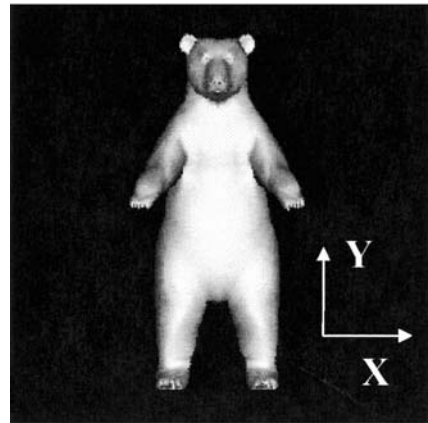
photometric stereo have been extensively studied in computer vision.²¹ We are particularly interested in detection methods based on linear subspaces created by an image database with multiple illumination changes of a given object.^{22,23} Those processes have been mainly applied to face recognition, where the recognition compared a new query image to each model in turn. To compare the models, a measure of similarity is required. The differences between 3D recognition methods depend on the selection of the subspace and on the measure of similarity applied. Other recognition methods are vision systems based on principal components analysis (PCA) and various Lambertian reflectance illumination models.²³ Although the above methods are considered as 3D recognition, in fact they are not, because no 3D image is utilized. Instead, a wide 2D database is created with all perspectives and changes in illumination for given objects, usually faces.

In this paper, we introduce a 3D recognition method based on the detection of 3D objects under changes of illumination conditions. The illumination model is defined in terms of Lambertian reflectance surfaces, surfaces that scatter light equally in all directions and appear equally bright from all viewing directions. We generate a subspace using an orthonormal basis. The basis is defined in terms of different directional light sources. The recognition process uses the calculation of the cosine of the angle between a target and the vector subspace. If the target belongs to that subspace, the vector that corresponds to the target will be contained in the subspace, the angle will be zero, and the cosine will be equal to unity. On the other hand, if the target does not belong to that space, the cosine will be smaller than 1. We will define the cosine measurements in terms of correlations. From the point of view of vector spaces, intensity invariant pattern recognition consists of recognizing vectors independently of their length, which can be viewed as an angle measurement between vectors in vector spaces. This angle provides a measure of the similarity between the object and the reference function.

Lefebvre *et al.*²⁴ defined a nonlinear filtering method called the locally adaptive contrast invariant filter (LACIF), which is invariant under any linear intensity transformation. This LACIF operation uses three correlation operations involving local statistics and nonlinearities. It was applied directly to scenes containing unsegmented targets. One of the advantages of the LACIF method is that no *a priori* information about the constant values involved in the linear illumination model is assumed. The LACIF method can be combined with synthetic discrimination filters to achieve both illumination invariance and out-of-plane rotation invariance.²⁵ The authors recently²⁶ generalized the LACIF filtering for situations where an additional linear intensity gradient across an object is present. It is interesting to consider the LACIF technique in the context of a vector space interpretation. In this paper, we have extended the LACIF idea to 3D



(a)



(b)

Fig. 1. (a) 3D object mesh. (b) Sample view of 3D object shaded.

object recognition, and we compare both illumination invariant detection methods.

The paper is organized as follows: In Section 2 the different illumination techniques for 3D objects are introduced. The 3D image representation in terms of vector spaces is in Section 3. The results comparing the new method with the LACIF are presented in Section 4, and in Section 5 we present our conclusion.

2. Illumination Model for Three-Dimensional Images

Variations in lighting can have a big effect on the appearance of 3D images. When the sources of light are far from the object, the lighting conditions may be described by specifying the intensity of light as a function of direction. Our illumination model is defined in terms of ambient illumination and Lambertian illumination. An ambient light is a diffuse, nondirectional source of light, typically the result of multiple reflections of light from many surfaces in the environment. It has a uniform intensity at any point

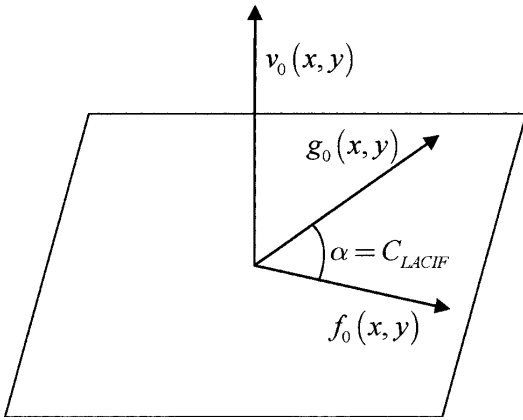


Fig. 2. Geometrical interpretation of LACIF.

in the environment. The influence of ambient illumination for a given surface can be expressed mathematically as

$$I_1(x, y, z) = I_a k_a(x, y, z), \quad (1)$$

where I_a is the intensity of ambient light, and $k_a(x, y, z)$ is the ambient-reflection coefficient of an object's surface that is determined by the material properties of the surface. Ambient light will generate a uniform intensity across the surface of an object. Note that Eq. (1) gives an intensity distribution for a 3D space. Moreover, we assume that, aside from the ambient light, the surface of the object reflects light according to Lambert's law²⁷ from localized light sources as

$$I_2(x, y, z) = I_p k_l(x, y, z) (\mathbf{L}(x, y, z) \cdot \mathbf{N}(x, y, z)), \quad (2)$$

$$I(x, y, z) = \begin{cases} I_a k_a(x, y, z) + I_p k_l(x, y, z) [L_x N_x(x, y, z) + L_y N_y(x, y, z) + L_z N_z(x, y, z)] & \text{if } (\mathbf{L} \cdot \mathbf{N}) \geq 0 \\ I_a k_a(x, y, z) & \text{otherwise} \end{cases}, \quad (4)$$

where I_p is the point light source's intensity, $k_l(x, y, z)$ is the material's diffuse reflection coefficient, $\mathbf{L}(x, y, z)$ is the unit vector of light direction, $\mathbf{N}(x, y, z)$ is the normal unit vector of the surface's object for a given point, and \cdot represents the inner or scalar product.

Finally our illumination model is given by the contribution of ambient illumination [Eq. (1)] and Lambertian illumination [Eq. (2)] as

$$I(x, y, z) = I_a k_a(x, y, z) + I_p k_l(x, y, z) \times (\mathbf{L}(x, y, z) \cdot \mathbf{N}(x, y, z)). \quad (3)$$

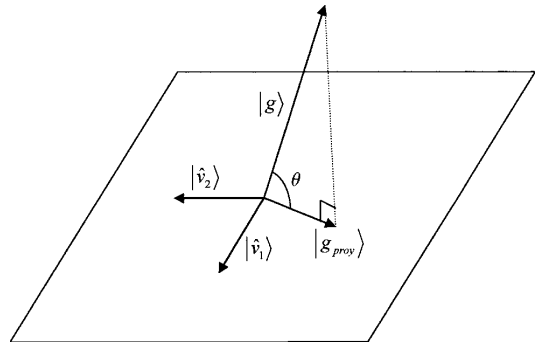


Fig. 3. Definition of the angle between a vector and the subspace.

From Eq. (3) we may obtain images that are not physically realizable, because the corresponding linear combination may contain negative values. These areas define the shadows, so a point P of the surface is a shadow if the angle between the surface normal and the direction of light is obtuse, i.e., $\cos(\mathbf{L} \cdot \mathbf{N}) \leq 0$. We shall assume that when the linear combination of Eq. (3) produces negative gray values, those values can be set to zero. This criterion is widely used for purposes of display or recognition.

Although the illumination model considered in this paper deals with the 3D information of an object, from now on we will consider the information of Eq. (3) as looking at the object from a particular point of view or perspective. It is similar to taking a picture of the 3D image, so the images will be two dimensional. Figure 1(a) shows a 3D object and Fig. 1(b) the corresponding 2D object for a given illumination and perspective. Figure 1(b) has been calculated as

where $\{N_i(x, y, z)\}_{i=x,y,z}$ and $\{L_i\}_{i=x,y,z}$ are the Cartesian coordinates of \mathbf{L} and $\mathbf{N}(x, y, z)$, respectively. In the literature, there are more complex illumination models²⁷ that take into account only matte surfaces, etc.

3. Vector Spaces for Images

Images may be considered as vectors in a Hilbert space. Then, any vector can be expressed in terms of an given basis. Correlation is a measure of the similarity between images, because one way to understand the correlation operation is to consider it as an inner product between two functions, the object and the reference functions.

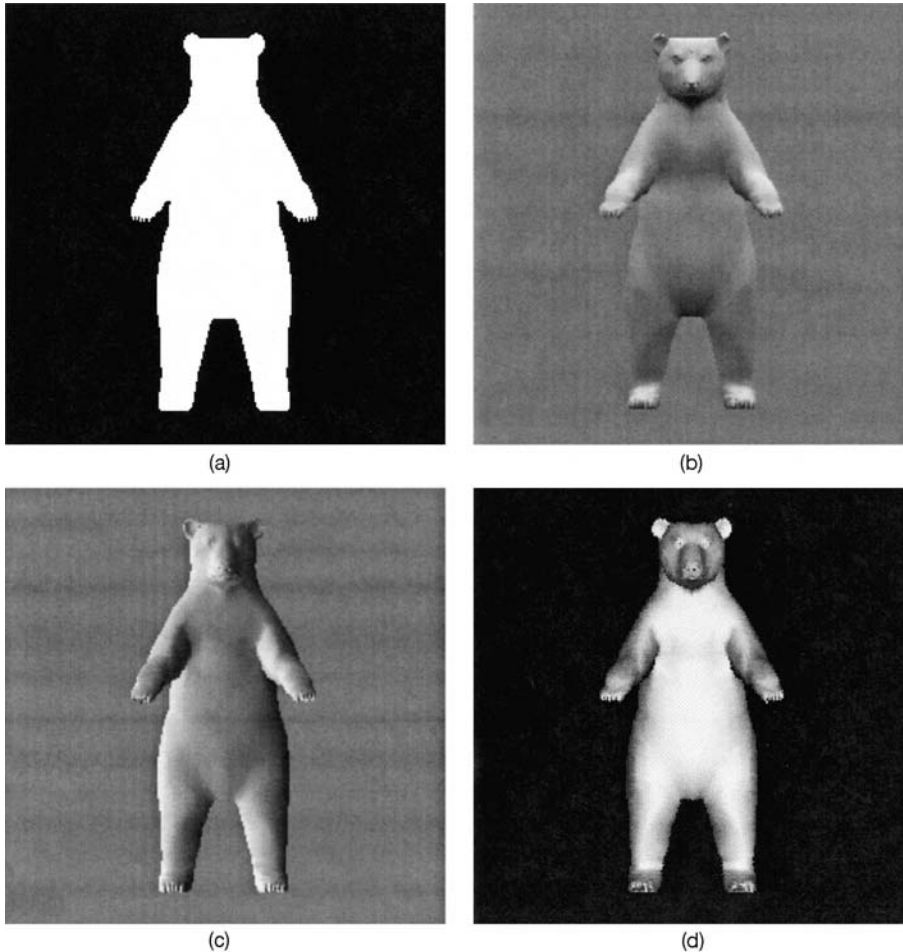


Fig. 4. Vector basis $\{k_a(x, y), [N_i(x, y)]\}$. (a) is $k_a(x, y)$ equal to $v_0(x, y)$, and (b), (c), and (d) are $\{N_i(x, y)\}_{i=X,Y,Z}$, respectively.

From the point of view of vector spaces, intensity invariant pattern recognition consists of recognizing vectors independently of their length, which can be viewed as an angle measurement between vectors in vector spaces. This angle provides a measure of the similarity between the object and the reference function.

For 2D images, Dickey and Romero²⁸ defined a normalized correlation as a normalization of the inner product represented by the correlation integral. This method yields correlation peak values that are invariant under a multiplicative factor. Arsenaault and Lefebvre²⁹ used a homomorphic transformation to change a multiplicative-intensity problem into an additive-intensity problem that can be addressed with the synthetic discriminant filter mentioned above. Lefebvre *et al.*²⁴ defined a nonlinear filtering method LACIF, which is invariant under any linear intensity transformation. This LACIF operation uses three correlations involving local statistics and nonlinearities.

It was applied directly to scenes containing unsegmented targets. One of the advantages of the LACIF method is that no *a priori* information about the constant values involved in the linear illumination model is assumed. The LACIF method can be combined with synthetic discrimination filters to achieve both illumination invariance and out-of-plane rotation invariance.²⁴ Recently, in Ref. 25 the authors generalized the LACIF filtering for situations where a linear intensity gradient across an object is present.

From our illumination model defined in Eq. (4), any 3D image can be expressed as a linear combination of four images $(k_a(x, y), \{N_i(x, y)\}_{i=x,y,z})$. In this paper, we define a new method based on a vector space representation in order to detect 3D objects under different lighting conditions. Our motivation is to develop an algorithm to recognize images that can be defined as a linear combination of other elementary images. That linear combination is mathematically defined as

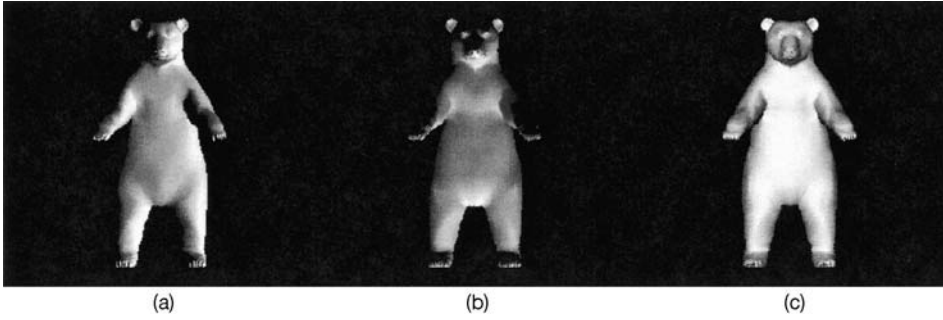
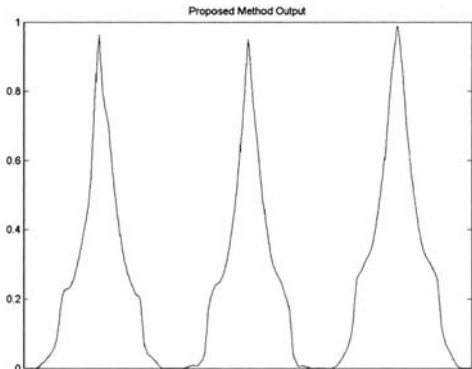


Fig. 5. Three versions of the target with different illuminations.

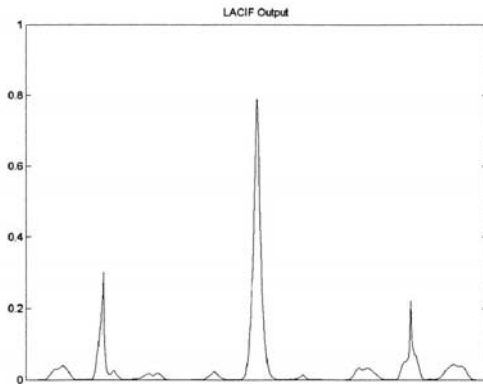
$$f(x, y) = c_0 v_0(x, y) + c_1 v_1(x, y) + c_2 v_2(x, y) + \dots, \quad (5)$$

where $f(x, y)$ is a certain image and $\{c_i\}_{i=0,1,\dots}$ are arbitrary coefficients. For convenience, we assume that $v_0(x, y)$ is the silhouette of the image defined as

$$v_0(x, y) = \begin{cases} 1 & \text{if } f(x, y) \neq 0 \\ 0 & \text{otherwise} \end{cases}. \quad (6)$$



(a)



(b)

Fig. 6. (a) Correlation peak profile for Fig. 5 using the proposed method. (b) Correlation peak profile for Fig. 5 using the LACIF.

Moreover, from Eq. (5), $\{v_i(x, y)\}_{i=1,2,\dots}$ are 2D images defined inside $v_0(x, y)$. From the point of view of vector spaces, $f(x, y)$ is a vector ($|f\rangle$) of a vector subspace generated by $\{|v_i\rangle\} = \{v_i(x, y)\}$. In addition, the inner product between 2D images, $a(x, y)$ and $b(x, y)$, is defined as

$$\langle a|b\rangle \equiv (a * b)(0, 0) = \int_{\mathbb{R}^2} a^*(x, y) b(x, y) dx dy, \quad (7)$$

where $*$ represents correlation. Note that we use as the inner product the value of the correlation between two functions at the origin. In the previous equations, we have supposed that images are continuous functions defined in \mathbb{R}^2 . Because of that, the canonical basis of the vector space is represented by $\{\delta(x - x', y - y'), \forall x', y' \in \mathbb{R}^2\}$. Although the basis has infinite dimensions, for computing the results we have used discrete images of $N_1 \times N_2$ pixels. For a certain $N_1 = N_2 = 256$ pixel image size, the theoretical dimension of our space is $256 \times 256 = 65,536$, which is difficult to handle from a practical point of view unless strong restrictions apply. However, from Eq. (5), we just operate with a vector subspace whose dimensions are defined by the number of images $\{v_i(x, y)\}$, so the complexity of the dimensionality problem has been reduced from infinity to a small finite number.

The basis defined by $\{v_i(x, y)\}$ is neither orthogonal nor orthonormal. We apply the Gram-Schmidt³⁰ orthonormalization method to the basis. Then the reference image, $f(x, y)$, can be defined as

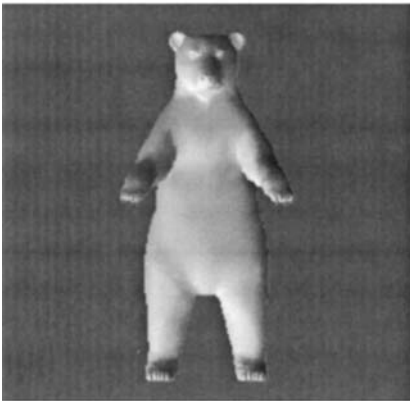
$$f(x, y) = d_0 \cdot \hat{v}_0(x, y) + d_1 \cdot \hat{v}_1(x, y) + d_2 \cdot \hat{v}_2(x, y) \dots, \quad (8)$$

where $\{\hat{v}_i(x, y)\}$ defines the orthonormal basis of the subspace, and $\{d_i\}$ are the components for the reference image. In vector space notation, the components can be defined as $\langle v_i|f\rangle$.

After those explanations, we can reformulate our original detection goal. The recognition method must



(a)



(b)



(c)

Fig. 7. (a) A 3D target. (b) Representation of (a) in the basis defined by $\{k_a(x, y), [N_i(x, y)]\}$ with no consideration of shadow effects. (c) Subtraction of (a) from (b).

distinguish between images that belong or not to a certain subspace or to a certain family of images defined in Eq. (5) or Eq. (8). For 2D images, Lefebvre *et*

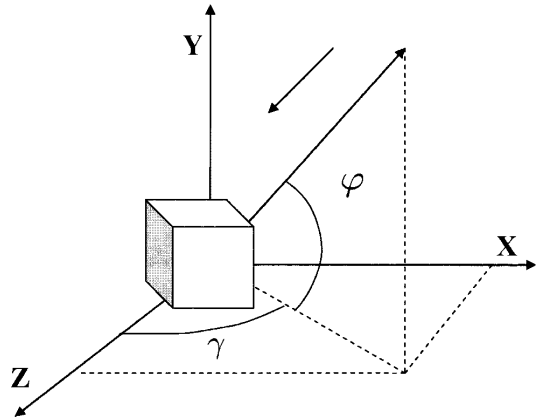


Fig. 8. Illumination variation in terms of the position of the point source.

*al.*²⁴ defined the LACIF operation as an operation able to perform this classification.

We now review the LACIF filtering and its extension to 3D object recognition. We shall also define a novel algorithm based on a local angular distance measurement.

A. Locally Adaptive Contrast Invariant Filter

This filtering operation was used to detect images independently of intensity changes. For an arbitrary target $(g(x, y))$ and a reference object $(f(x, y))$, the LACIF method is defined in terms of the correlation as

$$C_{\text{LACIF}}(g, f; x, y) = \frac{(f_0 * g)^2(x, y)}{\sqrt{N}(\hat{v}_0 * g^2)(x, y) - (\hat{v}_0 * g)^2(x, y)}, \quad (9)$$

where N is the number of pixels in the region of support, $\hat{v}_0(x, y)$ is the normalized version of $v_0(x, y)$, $f_0(x, y)$ is the zero-mean target defined as $f_0(x, y) = f(x, y) - \mu_f v_0(x, y)$, and μ_f is the mean of $f(x, y)$.

If the intensity transformation of the target is $af(x, y) + b$, where a and b are unknown parameters that are constant over the size of a single target, the LACIF output is equal to that obtained by the reference function, $f(x, y)$, i.e., the output is intensity invariant. In other words, if $g(x, y)$ is a linear combination of the orthonormal basis, $\{\hat{v}_i(x, y)\}$, the LACIF peak will be equal to 1, and be smaller than 1 if it is not. In terms of a vector interpretation, the LACIF technique is equivalent to projecting the target $(g(x, y))$ onto the subspace orthogonal to $v_0(x, y)$; that is, to project the zero-mean target onto this region of support, and then to calculating the cosine of the angle α between the target and the reference $(f(x, y))$.²⁴ The vector interpretation of the LACIF filter is illustrated in Fig. 2.

The LACIF method is used to determine whether an object belongs or not to the vector subspace gen-

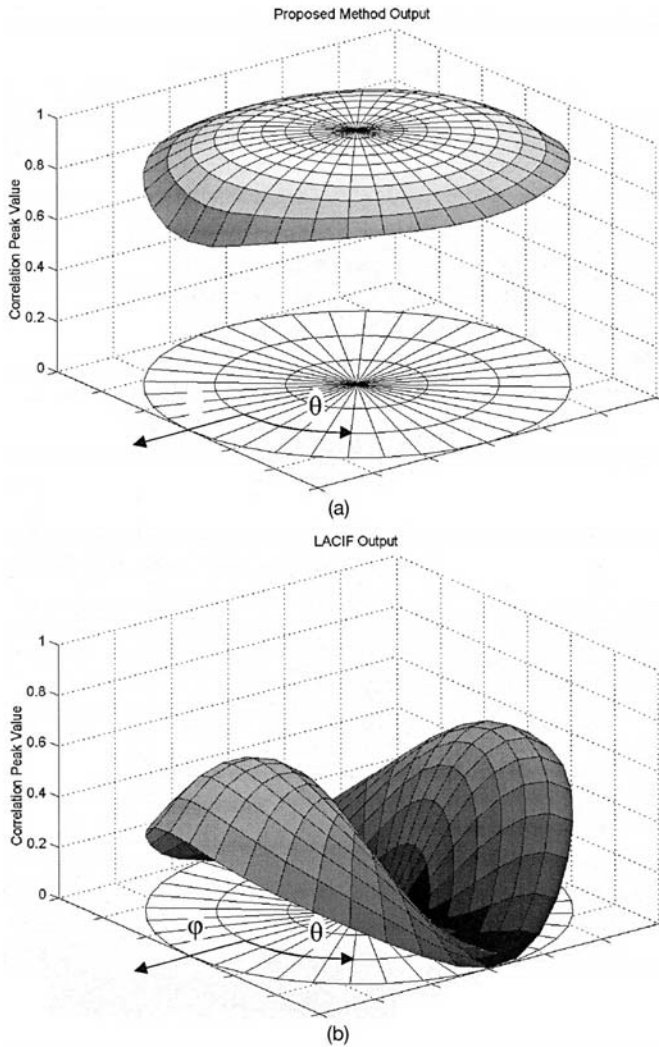


Fig. 9. Results for all illumination sampling for (a) the proposed LADC method and (b) the LACIF method.

erated by the vector basis $\{v_0(x, y), f(x, y)\}$. In addition, the LACIF can be generalized for the case of having more than two elements of the basis. Taking into account $\{\hat{v}_i(x, y)\}_{i=0,1,\dots}$ basis, LACIF can be defined as

$$C_{\text{LACIF}}(g, \hat{v}_i; x, y) = \frac{(\hat{v}_1 * g)^2(x, y)}{\sqrt{N}(\hat{v}_0 * g^2)(x, y) - \sum_{i \neq 1} (\hat{v}_i * g)^2(x, y)}. \quad (10)$$

Lefebvre *et al.*²⁵ applied Eq. (10) to affine transformations of intensity for 2D images.

According to our 3D illumination model, any 3D image can be expressed as a linear combination of four 2D images. Because of that, we propose the use

of the LACIF to detect 3D illuminated objects. The detection method will be based on knowing if a certain 3D illuminated object is similar or not to a 3D reference object. The reference object is defined in terms of a certain basis given by the normally illuminated images $\{N_i(x, y)\}$. Because the linear combination implies multiplication by constant factors, the 3D recognition result will be invariant for darker or lighter objects.

B. Proposed Method: Local Angular Distance Correlation

We now propose a method to detect 3D objects based on the vector subspace idea. Instead of using the LACIF, we propose a simpler and more robust method based on calculating the angle between a given 3D target (e.g., a vector) and the whole subspace defined by a base. For the sake of clarity, con-

consider the geometrical interpretation shown in Fig. 3. The vector space is generated only by two elements, $|\hat{v}_1\rangle$ and $|\hat{v}_2\rangle$. The vector $|g\rangle$ is the target. From Fig. 3, the angle, θ , between $|g\rangle$ and its projection, $|g_{\text{proj}}\rangle$ can be defined as

$$\cos^2(\theta) = \frac{\|g_{\text{proj}}\|^2}{\|g\|^2}, \quad (11)$$

where $\|g\|$ and $\|g_{\text{proj}}\|$ are the moduli of $|g\rangle$ and $|g_{\text{proj}}\rangle$. The projection of that vector onto the subspace defined by Eq. (8), can be expressed³¹ as

$$|g_{\text{proj}}\rangle \equiv \sum_i \langle \hat{v}_i | g \rangle |\hat{v}_i\rangle. \quad (12)$$

If vector $|g\rangle$ is contained in the subspace defined by Eq. (8), then $|g\rangle \equiv |g_{\text{proj}}\rangle$. In that case, the angle between the vector and the subspace will be equal to zero, and so the cosine is equal to unity. But if the vector is orthogonal to that subspace, it means that no information about the object is contained in that subspace, so $|g_{\text{proj}}\rangle = 0$ and the angle will be equal to $\pi/2$, and the cosine will vanish. Otherwise, the cosine value will be between zero and 1.

Note that Eq. (12) is a measurement of the similarity between two vectors, i.e., between the reference object (under varying illumination) and the target. However, the operation does not give the location of the target in a scene. It is not shift invariant because it is evaluated only at the origin. In order to extend the definition to the whole (x, y) space, the numerator of Eq. (11) is

$$\|g_{\text{proj}}(x, y)\|^2 = \sum_i |\langle \hat{v}_i | g \rangle|^2 = \sum_i |\hat{v}_i * g|^2(x, y), \quad (13)$$

and the denominator of Eq. (11) can be written as

$$\|g(x, y)\|^2 = \int_{\mathbb{R}^2} v_0(x, y) g^2(x, y) = \sqrt{N} (\hat{v}_0 * g^2)(x, y), \quad (14)$$

where

$$\hat{v}_0(x, y) = \frac{v_0(x, y)}{\sqrt{\langle v_0 | v_0 \rangle}} = \frac{v_0(x, y)}{\sqrt{N}}. \quad (15)$$

From Eq. (14), the energy of the target ($g^2(x, y)$) is evaluated inside the region of support for each point (x, y) of the image. This is the reason why our approach is a local angular distance measurement.

Taking into account previous equations, the final expression for our local angular distance correlation (LADC) is

$$\cos^2(\theta; g, \hat{v}_i; x, y) = \frac{\sum_i (\hat{v}_i * g)^2(x, y)}{\sqrt{N} (\hat{v}_0 * g^2)(x, y)}. \quad (16)$$

Equation (16) is defined in terms of correlations,

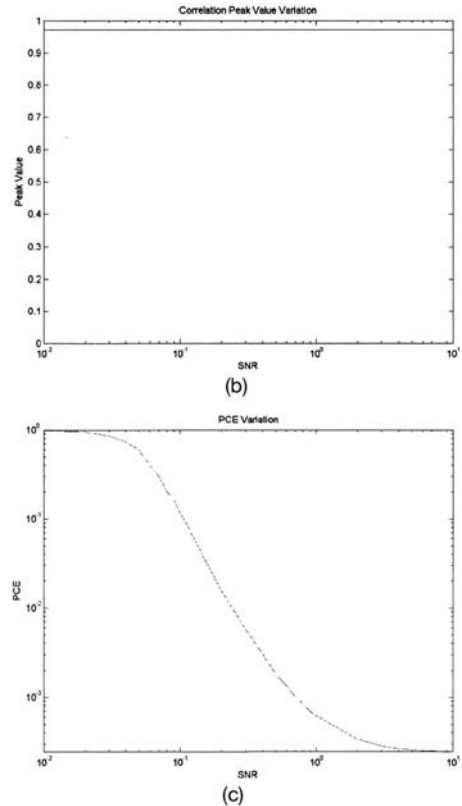
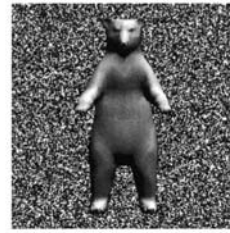


Fig. 10. (a) Sample noisy image (SNR = 5). (b) Correlation peak value variation. (c) PCE variation.

which can be an advantage if for optical implementations. Moreover, it gives a measure of the similarity between objects that belongs to a given vector subspace, and it localizes those targets in the scene because it is shift invariant. In contrast with the LACIF, which considers only the projection onto the silhouette, our method takes into account the projection onto all the subspace. This implies important differences for detecting images under different illuminations as we shall show in Section 4.

4. Illuminated Three-Dimensional Object Detection Results

In Section 2 we described our illumination model, which consists of expressing a 3D object as a linear

combination of normal surface vector basis images [see Eq. (4)]. Changing the lighting conditions implies changing the constants that multiply the vector basis. Figure 4 shows the vector basis for the reference 3D object of Fig. 1(b). Figure 4(a) is the region of support and Figs. 4(b), 4(c), and 4(d) are $\{N_i(x, y)\}_{i=X,Y,Z}$, respectively.

We now compare the results of 3D detection using the LACIF and the new correlation based on angular distance. Figure 5 shows three instances of the target with changes in illumination. The corresponding correlation peaks for the proposed method, and for the LACIF, are shown in Figs. 6(a) and 6(b), respectively. Note that the LACIF procedure yields missed detections for two of the targets (the output peaks are very low). This means that the operation is not intensity invariant for this case. On the other hand, the LADC output value is approximately the same for all illumination views. To illustrate the reason for those differences between the LACIF and the LADC, we must explain the choice of the vector basis. Because we are not taking into account shadow effects in our vector basis definition, if a certain illuminated 3D target is affected by shadows, the vector basis will not completely define the object, so the basis will not be complete. This is illustrated in Fig. 7, where we see differences between a certain illuminated target [Fig. 7(a)] and the projection of the target in the four-vector basis [Fig. 7(b)]. Figure 7(c) shows the result of subtracting Fig. 7(a) from Fig. 7(b). The difference is not zero because the basis is not complete. For this reason, a target cannot be expressed exactly in terms of a linear combination of our four-vector space basis, and the LACIF will fail in the detection. Because our LADC is a measurement of the distance between the target and the four-vector subspace, and because the targets are quite similar, the distance will be small and the correlation peak value will be close to 1. As a result, our LADC is more robust to changes in illumination than the LACIF.

To show the intensity invariance of the LADC method, we represent the correlation output for all possible illuminations as an angular map representation. The changes of illumination are generated by changing the position of the point source that illuminates the 3D object. Figure 8 shows the two angles that will define a specific position of the point source that illuminates the object.

The comparison between the LADC and LACIF methods are shown in Fig. 9. Note that the LADC values of Fig. 9(a) are quite stable around the value 1, while the LACIF values of Fig. 9(b) fails for most of the illuminations.

The next step is to study the LADC for 3D images in the presence of noise. We studied the robustness of the correlation peak for the LADC detection of targets under Gaussian disjoint noise. Figure 10(a) is a target corrupted with nonoverlapping noise. Figure 10(b) shows the variation of the correlation peak as the noise increases. A low signal-to-noise (SNR) ratio value implies a highly noisy image. Figure 10(b) shows the stability of the correlation peak, which means that the LADC method is not affected by this kind of noise. Another measure of the performance of the correlation peak is given by the peak to correlation energy (PCE). Figure 10(c) shows the values of the PCE as the noise is increased. The PCE measures the sharpness of the correlation peak. A high value of PCE means a sharp correlation peak. From Fig. 10(c) we see that instead of decreasing the correlation peak sharpness, the noise improves the results. The following paragraphs explain that result.

Consider an image

$$f'(x, y) \equiv f(x, y) + R(x, y), \quad (17)$$

where $f(x, y)$ is an arbitrary image, and $R(x, y)$ is a zero-mean Gaussian noise distribution with standard deviation σ that does not overlap the arbitrary image:

$$R(x, y) = \begin{cases} 0 & \text{if } \hat{v}_i(x, y) = 0 \\ \text{Gaussian noise elsewhere} & \end{cases} \quad (18)$$

The PCE value after applying our LADC method is

$$\text{PCE}(\sigma) = \frac{[\cos^2(\theta; f', \hat{v}_i, 0, 0)]^2}{\int_{\mathbb{R}^2} [\cos^2(\theta; f', \hat{v}_i, x, y)]^2 dx dy}. \quad (19)$$

Because the noise is nonoverlapping, the denominator of Eq. (19) can be split into two independent terms:

$$\text{PCE}(\sigma) = \frac{[\cos^2(\theta; f, \hat{v}_i, 0, 0)]^2}{[\cos^2(\theta; f, \hat{v}_i, 0, 0)]^2 + \int_{(x, y) \neq (0, 0)} [\cos^2(\theta; f', \hat{v}_i, x, y)]^2 dx dy}. \quad (20)$$

The second term of the denominator can be expressed as

$$\cos^2(\theta; f', \hat{v}_i, x, y) = \frac{\sum_i \left[\int_{\mathbb{R}^2} \hat{v}_i^*(x' - x, y' - y) [f(x, y) + R(x, y)] dx' dy' \right]^2}{\sqrt{N} \int_{\mathbb{R}^2} \hat{v}_0^*(x' - x, y' - y) [f^2(x', y') + f \cdot R(x', y') + R^2(x', y')] dx' dy'}. \quad (21)$$

Considering a zero-mean Gaussian noise, Eq. (21) can be rewritten as

$$\begin{aligned} \cos^2(\theta; f', \hat{v}_i, x, y) &= \frac{K}{\sqrt{N} \int_{\mathbb{R}^2} \hat{v}_0^*(x' - x, y' - y) [R^2(x', y')] dx' dy'} \\ &\propto K' \sigma^{-2}, \end{aligned} \quad (22)$$

where K and K' are constant terms. Then,

$$\text{PCE}(\sigma) = \frac{[\cos^2(\theta; f, \hat{v}_i, 0, 0)]^2}{[\cos^2(\theta; f, \hat{v}_i, 0, 0)]^2 + \mathcal{O}(\sigma^{-4})}. \quad (23)$$

If $\sigma \rightarrow \infty$, the PCE value will converge to 1.

Finally, in order to show the discrimination abilities of the new method, Fig. 11(a) shows an input

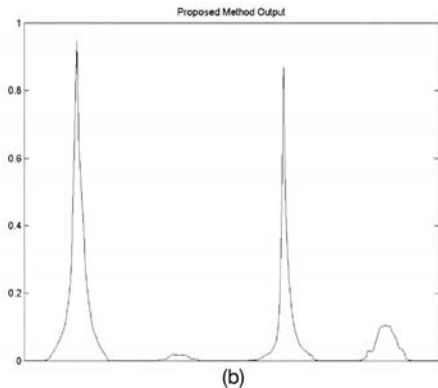
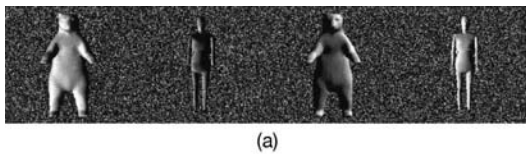


Fig. 11. (a) Input scene with true and false objects in a noisy environment (SNR = 1). (b) Output for the proposed method using (a).

image made of two reference objects with different illuminations and two false objects. The result for the

LADC is in Fig. 11(b). The correlation peaks for the true targets are equal to 1, whereas for the false target the output is lower, which demonstrates good discrimination even in the presence of nonoverlapping Gaussian noise. Note that we have used computer simulated objects in which 3D mesh is obtained digitally. For the case of a real 3D object, one needs to register different illumination views of the object in order to perform the vector basis.

5. Conclusion

We propose an illumination-invariant 3D recognition method based on vector spaces. The detection is done in terms of a local angular distance between a given target vector and a vector subspace defined by a 3D elementary image basis. We extended a previous intensity-invariant LACIF expression to 3D image recognition, and we have compared both methods. The results show the robustness of the new proposed method under changes of illumination. Experiments have validated the intensity invariant recognition of 3D images in the presence of nonoverlapping Gaussian noise. We have shown that the PCE value increases with the amount of noise in the image. We also successfully tested the method for composite images, when other false targets were in the scene.

This work was supported by FEDER funds and the Spanish "Ministerio de Educación y Ciencia" under the project FIS2004-06947-C02-01, the D.G Investigació i Transferència Tecnològica, project IIARC0/2004/217, and the Agencia Valenciana de Ciencia y Tecnología (AVCT), project GRUPOS03/117. J. J. Vallés acknowledges a FPU grant from the Ministry of Education and Science.

References

1. R. Campbell and P. Flynn, "A survey of free-form object representation and recognition techniques," *Comput. Vis. Image Underst.* **81**, 166–210 (2001).
2. M. Rioux, "Laser range finder based on synchronized scanners," *Appl. Opt.* **23**, 3837–3844 (1984).
3. M. Takeda and K. Mutoh, "Fourier transform profilometry for the automatic measurement of 3D object shapes," *Appl. Opt.* **22**, 3977–3982 (1983).
4. J. Rosen, "Three-dimensional electro-optical correlation," *J. Opt. Soc. Am. A* **15**, 430–436 (1998).
5. T. Poon and T. Kim, "Optical image recognition of three-dimensional objects," *Appl. Opt.* **38**, 370–381 (1999).
6. E. Tajahuerce, O. Matoba, and B. Javidi, "Shift-invariant

- three-dimensional object recognition by means of digital holography," *Appl. Opt.* **40**, 3877–3886 (2001).
7. E. Paquet, H. H. Arsenault, and M. Rioux, "Recognition of faces from range images by means of the phase Fourier transform," *Pure Appl. Opt.* **4**, 709–721 (1995).
 8. P. Parrein, J. Taboury, and P. Chavel, "Evaluation of the shape conformity using correlation of range images," *Opt. Commun.* **195**, 393–397 (2001).
 9. J. J. Esteve-Taboada, D. Mas, and J. Garcia, "Three-dimensional object recognition by Fourier transform profilometry," *Appl. Opt.* **38**, 4760–4765 (1999).
 10. E. Paquet, P. García-Martínez, and J. García, "Tridimensional invariant correlation based on phase-coded and sine-coded range images," *J. Opt.* **29**, 35–39 (1998).
 11. J. J. Esteve-Taboada and J. García, "Detection and orientation evaluation for three-dimensional objects," *Opt. Commun.* **217**, 123–131 (2003).
 12. B. Javidi and E. Tajahuerce, "Three-dimensional object recognition by use of digital holography," *Opt. Lett.* **25**, 610–612 (2000).
 13. R. Bamler and J. Hofer-Alfeis, "Three- and four-dimensional filter operations by coherent optics," *Opt. Acta* **29**, 747–757 (1982).
 14. Y. Li and J. Rosen, "Object recognition using three-dimensional optical quasi-correlation," *J. Opt. Soc. Am. A* **19**, 1755–1762 (2002).
 15. O. Matoba, E. Tajahuerce, and B. Javidi, "Real-time three-dimensional object recognition with multiple perspectives imaging," *Appl. Opt.* **40**, 3318–3325 (2001).
 16. J. J. Esteve-Taboada, J. García, and C. Ferreira, "Rotations-invariant optical recognition of three-dimensional objects," *Appl. Opt.* **39**, 5998–6005 (2000).
 17. J. J. Esteve-Taboada and J. García, "Detection and orientation evaluations for three-dimensional objects," *Opt. Commun.* **217**, 123–131 (2002).
 18. J. Garcia, J. J. Valles, and C. Ferreira, "Detection of three-dimensional objects under arbitrary rotations base don range images," *Opt. Express* **11**, 3352–3358 (2003).
 19. S. Yeom and B. Javidi, "Three-dimensional distortion-tolerant object recognition using integral imaging," *Opt. Express* **12**, 5795–5809 (2004).
 20. J. J. Esteve-Taboada, N. Palmer, J.-C. Giannessini, J. Garcia, and C. Ferreira, "Recognition of polychromatic three-dimensional objects," *Appl. Opt.* **43**, 433–441 (2004).
 21. B. K. Hom, *Robot Vision* (MIT Press, 1986).
 22. Y. Tian, H. T. Tsui, S. Y. Yeung, and S. Ma, "Shape from shading for multiple light sources," *J. Opt. Soc. Am. A.* **16**, 36–52 (1999).
 23. S. K. Zhou and R. Chellappa, "Image-based face recognition under illumination and pose variations," *J. Opt. Soc. Am. A.* **22**, 217–229 (2005).
 24. D. Lefebvre, H. H. Arsenault, P. Garcia-Martinez, and C. Ferreira, "Recognition of unsegmented targets invariant under transformations of intensity," *Appl. Opt.* **41**, 6135–6142 (2002).
 25. D. Lefebvre, H. H. Arsenault, and S. Roy, "Nonlinear filter for pattern recognition invariant to illumination and to out-of-plane rotations," *Appl. Opt.* **42**, 4658–4662 (2003).
 26. S. Roy, D. Lefebvre, and H. H. Arsenault, "Recognition invariant under unknown affine transformations of intensity," *Opt. Commun.* **238**, 69–77 (2004).
 27. J. D. Foley, A. van Dam, S. K. Feiner, and J. F. Hughes, *Computer Graphics, Principle and Practics*, 2nd ed. (Addison-Wesley, 1996).
 28. F. M. Dickey and L. A. Romero, "Normalized correlation for pattern recognition," *Opt. Lett.* **16**, 1186–1188 (1991).
 29. H. H. Arsenault and D. Lefebvre, "Homomorphic cameo filter for pattern recognition that is invariant with change of illumination," *Opt. Lett.* **25**, 1567–1569 (2000).
 30. <http://mathworld.wolfram.com/Gram-SchmidtOrthonormalization.html>.
 31. <http://mathworld.wolfram.com/VectorSpaceProjection.html>.

Phase Fourier vector model for scale invariant three-dimensional image detection

José J. Vallés, Pascuala García-Martínez, Javier García, and Carlos Ferreira

Departamento de Óptica. Universitat de València. C/. Dr. Moliner 50. 46100Burjassot Spain
pascuala.garcia@uv.es

Abstract: A scale invariant 3D object detection method based on phase Fourier transform (PhFT) is addressed. Three-dimensionality is expressed in terms of range images. The PhFT of a range image gives information about the orientations of the surfaces in the 3D object. When the object is scaled, the PhFT becomes a distribution multiplied by a constant factor which is related to the scale factor. Then 3D scale invariant detection can be solved as illumination invariant detection process. Several correlation operations based on vector space representation are applied. Results show the tolerance of detection method to scale besides discrimination against false objects.

©2007 Optical Society of America

OCIS codes: (070.2590) Fourier transforms; (070.5010) Pattern recognition and feature extraction; (100.6890) Three-dimensional image processing

References and links

1. B. Javidi, ed., *Image Recognition and Classification: Algorithms, Systems, and Applications*, (Marcel Dekker, New York, 2002).
2. J. Rosen, "Three-dimensional electro-optical correlation," *J. Opt. Soc. Am. A* **15**, 430–436 (1998).
3. J. Rosen, "Three-dimensional joint transform correlator," *Appl. Opt.* **37**, 7538–7544 (1998).
4. T. Poon and T. Kim, "Optical image recognition of three-dimensional objects," *Appl. Opt.* **38**, 370–381 (1999).
5. B. Javidi and E. Tajahuerce, "Three-dimensional object recognition by use of digital holography," *Opt. Lett.* **25**, 610–612 (2000).
6. B. Javidi, I. Moon, S. Yeom, and E. Carapezza, "Three-dimensional imaging and recognition of microorganism using single-exposure on-line (SEOL) digital holography," *Opt. Express* **13**, 4492–4506 (2005).
7. M. Takeda and K. Mutoh, "Fourier transform profilometry for the automatic measurement of 3-D object shapes," *Appl. Opt.* **22**, 3977–3882 (1983).
8. J. J. Esteve-Taboada, D. Mas, and J. García, "Three-dimensional object recognition by Fourier transform profilometry," *Appl. Opt.* **38**, 4760–4765 (1999).
9. J. J. Esteve-Taboada, J. García and C. Ferreira, "Rotation invariant optical recognition of three-dimensional objects," *Appl. Opt.* **39**, 5998–5352 (2000).
10. J. García, J. J. Vallés and C. Ferreira, "Detection of three-dimensional objects under arbitrary rotations based on range images," *Opt. Express*, **11**, 3352–3358 (2003).
11. J. J. Esteve-Taboada, N. Palmer, J.-Ch. Giannesini, J. García and C. Ferreira, "Recognition of polychromatic three-dimensional objects," *Appl. Opt.* **43**, 433–441 (2004).
12. J. J. Esteve-Taboada, J. García and C. Ferreira, "Optical recognition of three-dimensional objects with scale invariance using classical convergent correlator," *Opt. Eng.* **41**, 1324–1330 (2002).
13. M. Rioux, "Laser range finder based on synchronized scanners," *Appl. Opt.* **23**, 3837–3844 (1984).
14. E. Paquet, M. Rioux and H. H. Arsenault, "Invariant pattern recognition for range images using the phase Fourier transform and a neural network," *Opt. Eng.* **34**, 1178–1183 (1995).
15. E. Paquet, P. García-Martínez, and J. García, "Tridimensional invariant correlation based on phase-coded and sine-coded range images," *J. Opt.* **29**, 35–39 (1998).
16. S. Chang, M. Rioux, and C. P. Grover, "Range face recognition based on the phase Fourier transform," *Opt. Commun.* **222**, 143–153 (2003).
17. Y. Li and J. Rosen, "Scale invariant recognition of three-dimensional objects by use of a quasi-correlator," *Appl. Opt.* **42**, 811–819 (2003).
18. D. Lefebvre, H. H. Arsenault, P. García-Martínez, and C. Ferreira, "Recognition of unsegmented targets invariant under transformations of intensity," *Appl. Opt.* **41**, 6135–6142 (2002).
19. H. H. Arsenault and P. García-Martínez, "Intensity-invariant nonlinear filtering for detection in camouflage" *Appl. Opt.* **44**, 5483–5490 (2005).

20. H. H. Arsenault and D. Lefebvre, "Homomorphic cameo filter for pattern recognition that is invariant with change of illumination," *Opt. Lett.* **25**, 1567-1569 (2000).
 21. D. Lefebvre, H. H. Arsenault, and S. Roy, "Nonlinear filter for pattern recognition invariant to illumination and to out-of-plane rotations," *Appl. Opt.* **42**, 4658-4662 (2003).
 22. S. Roy, D. Lefebvre, and H. H. Arsenault, "Recognition invariant under unknown affine transformations of intensity," *Opt. Commun.* **238**, 69-77 (2004).
 23. J. J. Vallés, J. García, P. García-Martínez, and H. H. Arsenault, "Three-dimensional object detection under arbitrary lighting conditions," *Appl. Opt.* **45**, 5237-5247 (2006).
 24. F. M. Dickey and L. A. Romero, "Normalized correlation for pattern recognition," *Opt. Lett.* **16**, 1186-1188 (1991).
-

1. Introduction

Recent interest in three-dimensional (3D) optical information systems has increased because of its vast potential in applications such as object recognition, image encryption as well as 3D display [1]. Regarding 3D object optical recognition, joint transform correlator (JTC) architecture in combination to electronic processing can be used [2, 3]. However it is a complex setup and it is computationally intensive, although it provides localization of 3D targets in 3D space. Other techniques use digital holography as a recording method to perform correlation between planar holograms of the 3D functions [4-6]. In addition, techniques based on Fourier transform profilometry (FTP) are used for 3D object detection [7,8]. The FTP technique deals with the projection of gratings onto the 3D object besides a registration of a 2D image carrying the 3D information. It has been applied to many pattern recognition tasks, as rotation invariance [9, 10], color pattern recognition [11] and scale invariance [12].

Other recognition procedures are based on the application of optical or digital processing to range imagery. A range image contains the depth information of an object from a given view line, that defines the z axis [13]. One of the main advantages of the range image is that all 3D information is stored in a 2D image containing only geometrical information. The encoding of the depth information has been used in the literature to extend the possibilities of range images recognition using the phase Fourier transform (*PhFT*) [14,15]. They showed that it is possible to characterize a 3D object by using a limited number of normals which they calculated using the *PhFT* operation. The idea is to detect the planar surface of a 3D range image when the Fourier transform of a phase-coded image (the phase being proportional to the elevation) is calculated using feed-forward neural networks [14]. Another application of this technique is face recognition applications, for which the range image plays an important role since it contains richer and more accurate shape information than 2D intensity images [16].

However, a main limitation of most 3D matching methods is the limited tolerance to scale changes in the object. Several methods have been shown in the literature to show scale tolerance detection applied to 3D images [12, 14, 17]. In Ref. [12] the authors combine profilometry techniques with optical matched filtering based on Mellin radial harmonics to obtain 3D scale invariant pattern recognition. Moreover, other optical methods based on performing a correlation between a set of images of a tested 3D scene with a logarithmic radial harmonic filter is applied for scale invariance [17]. However the method suffers from complexity and a certain limitation of the space bandwidth product. To avoid such inconveniences, the use of *PhFT* of range images is studied for shift, rotation and scale invariance using feed-forward neural networks [14]. However, the process of training the neural network may be complex if one wants to provide a good discrimination result.

In this paper we have addressed scale invariant 3D object detection using the *PhFT* applied to range imagery in combination with nonlinear statistical operations that we have defined recently for intensity invariant 2D pattern recognition [18,19]. We show that scaling a 3D range object implies the multiplication of the *PhFT* distribution by a constant factor. So, a change of scale for a range image is transformed in a change in illumination (multiplication by a constant) of the *PhFT* amplitude. In common 2D correlation, if the illumination model consists in multiplying a target by an unknown constant factor, the correlation peak will change by the same amount. In such cases dark targets can be missed. Arsenault and

Lefebvre [20] used a homomorphic transformation to change a multiplicative-intensity problem into an additive-intensity problem that can be addressed with the synthetic discriminant filter mentioned above. Moreover, Lefebvre et al [18] defined a nonlinear filtering method known as the LACIF (Locally Adaptive Contrast Invariant Filter), which is invariant under any linear intensity transformation. This LACIF operation uses three correlations involving local statistics and nonlinearities. It was applied directly to scenes containing unsegmented targets. One of the advantages of the LACIF method is that no a priori information about the constant values involved in the linear illumination model is assumed. The LACIF method can also be combined with synthetic discrimination filters to achieve both illumination invariance and out-of-plane rotation invariance [21]. In Ref. [22] the authors generalized the LACIF filtering for situations where a linear intensity gradient across an object is presented. It is interesting to consider the LACIF technique in the context of a vector space interpretation. In addition, in Ref. [19], the authors applied LACIF algorithm to unsegmented natural camouflage scenes while maintaining intensity invariance. Recently, we have applied another filtering connected with LACIF to 3D shading to achieve illumination invariance [23]. In section 2 we review the basis of *PhFT* and how affects a change of scale in range imagery. LACIF filtering is described in Section 3. The application of the method to 3D scaled targets is in Section 4.

2. Scale changes for range image phase Fourier transforms

A range image, $z(x, y)$, contains the depth information of an object from a given view line, that defines the z axis. One of the main advantages of a range image is that three-dimensional information is stored in a 2D image containing only geometrical information. In fact, the range image is considered as a set of facets which may be described by their normals to the surface. The encoding of the depth information has been used in the literature to extend the possibilities of range images for pattern recognition [10, 14, 15]. Those techniques are based on phase coding. We encode the range as a phase distribution as follows:

$$P(x, y) = \exp[imz(x, y)] \quad (1)$$

where m is a constant that permits the adjustment of the phase slope of the object. From now, without losing generality, we will assume $m=1$.

A way to deal with range images keeping translation invariance is to use their Fourier transform. The Fourier transform of the phase encoded range image (*PhFT*) is

$$PhFT_z(u, v) = F_{2D} \{ \exp[iz(x, y)] \} \quad (2)$$

where F_{2D} stands for two dimensional Fourier transform.

One of the most important properties of the *PhFT* is that it contains information of all the orientations of the surfaces that defines a given 3D object [14, 15]. For instance, if the object is defined by a planar surface, after a phase encoding, it will become a planar phase distribution. Thus its Fourier transform will be peaked around a well defined location. So *PhFT* maps a facet into a peak. The position and distribution of the peak represent the orientation and the boundary of the facet, respectively. The intensity of the *PhFT* exhibits a crucial property: it is invariant to arbitrary translations of the object. This property is obvious for translation in the (x, y) plane, as they will produce just a linear phase factor in the *PhFT*. On the other hand, from the definition of the phase encoded range image [see Eq. (1)], a shift along the view line (z axis) will influence just as a constant phase factor. In both cases the *PhFT* is just altered by a phase that is irrelevant in intensity [15]. Furthermore, changing the orientation of the object (i.e. rotating all the facets that compose the range image) implies a translation of angular *PhFTs*. This connection between rotation in spatial domain and translation in angular *PhFT* domain makes it advantageous for 3D object correlation based recognition methods [9, 10].

As an example we consider Fig. 1(a), that contains a simple range image (pyramid) with four facets and the *PhFT* amplitude of Fig. 1(a) is shown in Fig. 1(b). The four facets (four normals) correspond to four location peaks in the Fourier plane. On the other hand, if the curvature of the object's surface is a continuous function (smooth object surface, i. e. with large number of facets), the *PhFT* will be continuous.

Now we show the effect of a scale change in *PhFT* domain. Let assume that the scaled object

is given by $z'(x, y) = z\left(\frac{x}{k}, \frac{y}{k}\right)$ with k being the scale factor. Although the scale is changed,

the shape of the object will remain the same as well as the orientation of the normals, except for a constant value which is related to the scale factor. In Appendix A, we show the connection between *PhFTs* of scaled objects. In fact, from Appendix A, the *PhFT* of the scaled object can be approximated as

$$PhFT_z'(u, v) \cong k^2 PhFT_z(u, v) \quad (3)$$

So a scale change in the range image will only lead to a multiplication of the amplitude of the non-scaled *PhFT* domain by a constant factor, with no change in the pattern distribution of the *PhFTs*. Only a global change in the intensity is observed.

Fig. 1(c) shows a scaled version of Fig. 1(a), and Fig. 1(d) shows the *PhFT* of Fig. 1(c). Both *PhFTs* [Fig. 1(b) and Fig. 1(c)] are the same except for a global multiplicative constant factor.

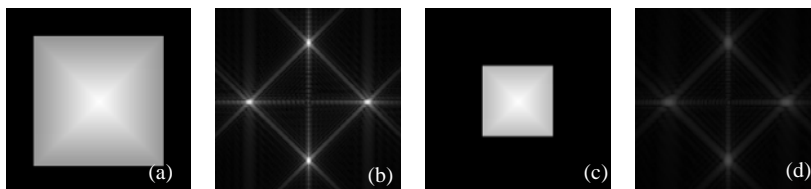


Fig. 1. (a). Range image of a diamond shaped object. (b). *PhFT* amplitude of Fig. 1(a). (c) Scaled version of Fig. 1(a). (d) *PhFT* amplitude of Fig. 1(c).

Moreover, if the surface of the 3D object is a bounded curve, the change of scale causes some new problems. Because of digitalization processes, a larger curved surface contains more small facets than those corresponding to a smaller curved surface (See Fig. 2).

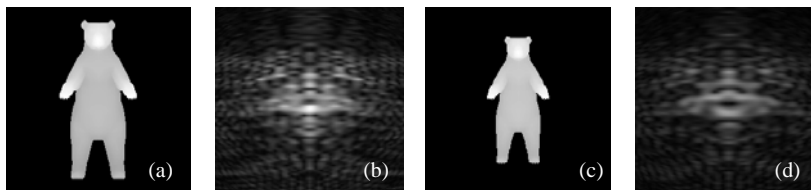


Fig. 2. (a). Range image. (b). *PhFT* amplitude of Fig. 2(a). (c) Scaled range image. (d) *PhFT* amplitude of Fig. 2(c)

Note that the relation between the *PhFTs* shown in Figs. 2(b) and 2(d), respectively, is not a global constant factor. In section 4 we will deal with this factor in more details. But, except for those digitalization errors, changes in scale involve changes in intensity or amplitude. There are several methods for intensity invariant pattern recognition based on correlations [18-22, 24]. In this paper we will apply the LACIF method. Following section briefly review the essentials.

3. Intensity invariant correlation method

Images may be considered as vectors in a Hilbert space. Then any vector can be expressed in terms of a given basis. Correlation is a measure of the similarity between images, because one

way to understand the correlation operation is to consider it as an inner product between two functions, the object and the reference. From the point of view of vector spaces, intensity invariant pattern recognition consists of recognizing vectors independently of their length, which can be viewed as an angle measurement between vectors in vector spaces. The vector space defined here consists of *PhFT* amplitude distributions.

Because a change of intensity implies a change in the length of the vector, two equal targets with different illuminations will be parallel, so the normalized correlation and the cosine will be equal to one. For 2-D images, Lefebvre et al [18] defined LACIF operation, which is invariant under any linear intensity transformation. This LACIF uses three correlations involving local statistics and nonlinearities. It was applied directly to scenes containing unsegmented targets. One of the advantages of the LACIF method is that no a priori information about the constant values involved in the linear illumination model is assumed. We will consider that the target is the *PhFT* amplitude in the Fourier domain. So, a linear transformation of intensity over a target can be expressed as

$$PhFT_z(u, v) = \alpha PhFT_z^o(u, v) + \beta \diamond(u, v) \quad (4)$$

where $\diamond(u, v)$ is the binary support which is equal to unity over the support of the target $PhFT_z(u, v)$, and equals to zero everywhere else, and α, β are unknown constants. An orthogonal basis for the subspace is selected. We define $PhFT_z^o(u, v)$ as

$$PhFT_z^o(u, v) = PhFT_z(u, v) - \mu_f \diamond(u, v) \quad (5)$$

where μ_f is the mean of $PhFT_z(u, v)$. Note that $PhFT_z^o(u, v)$ is a zero-mean target in the region of support. Then, the target can be defined as a linear combination of two orthogonal images (a silhouette and a zero-mean target) as

$$PhFT_z(u, v) = \alpha' PhFT_z^o(u, v) + \beta' \diamond(u, v) \quad (6)$$

where α' and β' are constants. The basis defined by $\{PhFT_z^o(u, v), \diamond(u, v)\}$ is not orthonormal, but we can normalize it to unit length as $\{\phi_1(u, v) = \diamond(u, v) / \|\diamond(u, v)\|; \phi_2(u, v) = PhFT_z^o(u, v) / \|PhFT_z^o(u, v)\|\}$, where

$\|PhFT_z^o(u, v)\| = \sqrt{PhFT_z^o(u, v) * PhFT_z^o(u, v)}$, and $*$ denotes the correlation, which is the inner product. Taking into account this orthonormal basis, $\{\phi_1(u, v), \phi_2(u, v)\}$, the target can now be defined as

$$PhFT_z(u, v) = \alpha'' \phi_1(u, v) + \beta'' \phi_2(u, v) \quad (7)$$

LACIF filtering operation at the output is defined as Refs. [18,19]

$$C_{LACIF}(u, v) = \frac{(PhFT_s(u, v) * \phi_2(u, v))^2}{\sqrt{N (PhFT_s(u, v)^2 * \phi_1(u, v) - (PhFT_s(u, v) * \phi_1(u, v))^2)}} \quad (8)$$

where N is the number of pixels inside the region of support. Then for a given range target $s(x, y)$, if $PhFT_s(u, v)$ is a linear combination of the orthonormal basis, then the correlation peak will be equal to one, and it will be smaller than one if it is not. We have applied LACIF filtering for scale invariant 3D object detection codified in terms of *PhFTs*. Note that for our 3D detection process only constant α is considered (see Eq. (3), where $\alpha = k^2$), whereas the constant β equals zero. It means that there is no global constant added to the *PhFT*, but multiplied. Moreover, the definition of Eq. (8) in terms of correlations makes the approach feasible to be implemented optoelectronically. The correlations can be performed by

conventional optical correlators like Vander Lugt correlator or joint transform correlator architectures and the local calculations can be obtained using computer interface.

4. Results of detection

We have carried out experiments to show the performance of the detection method. All of those experiments are based on the calculation of Eq. (8) when the reference target is scaled between 0.6 and 1.2 factors. We have used a square window around the *PhFT* amplitude distribution as the region of support for all the experiments. We have chosen as reference target the average between different scaled reference targets in order to minimize the possible sampling errors due to scaling digitalization process. Results are shown in Fig. 3. As we see from Fig. 3(a), the LACIF is almost constant for all scale factors (60%-120%) and its value oscillates between 0.7 and almost 0.9 correlation peak value. A multimedia file shows the appearance of the range images and the result of the detection.

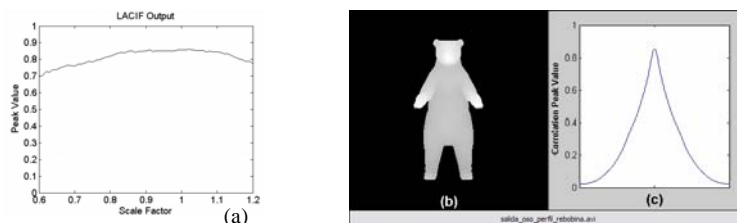


Fig. 3. (a). LACIF value for different scale factors (b). Multimedia file (1.09MB). Scaled Range Targets (c). Profile 1-D with the LACIF correlation peak value

Note from Fig. 3 that the LACIF correlation is not equal to one. In terms of vector space, a correlation value which is not a maximum means that the average target is not an element of the basis. However, the average target is a good candidate to represent of all the scaled targets, and it contains enough information to give a higher correlation peak value.

Another experiment deals with face recognition. Since the surface of human face can be expressed as a set of facets, the *PhFT* of the human face provides a new signature of the face. Ref. [16] deals with invariant to scale, translation and rotation for face human recognition. However the tolerance of the method presented in Ref. [16] to all the changes is quite limited. Our results with LACIF show an improvement of the robustness of the method. At the same time we will consider other false object to verify the discrimination capability of the LACIF method.

Figure 4(a) shows a human face target and Fig. 4(b) shows a false target (a different person) to discriminate. Results are shown in Fig. 4(c) and 4(d), respectively. Note that the different scaled faces are correctly detected above around 70%, whereas the false target is rejected with 35% of discrimination. This means that choosing a detection threshold of around value 0.3 the correct object will be detected and the false rejected. So the method is robust for detection and discrimination for all scale factors considered.

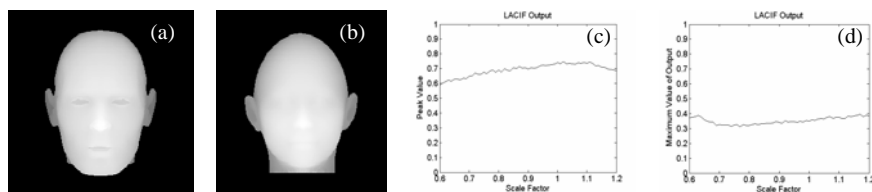


Fig. 4. (a). Target range image. (b). False range image. (c). LACIF peak value for different scaled targets of Fig. 4(a). (d). Idem for the false target.

In summary, scale invariant 3D object recognition is converted to intensity invariant pattern recognition technique. This can be possible if the recognition method implies a phase codification and a Fourier transformation of the 3D range images. A scale change of the *PhFT* of a range images is a multiplication of a constant factor in the *PhFT* of the reference range target. LACIF filtering detects targets which have been multiplied by constant values. Then, applying LACIF to Fourier phase encoded range images will solve the scale changes in 3D range images detection. Various experiments were carried out to validate the scale invariance. For real applications, the method has applied to face human recognition. We successfully tested the method when other false targets are tested.

Acknowledgments

The authors gratefully acknowledge the editor's comments. This work was supported by the Spanish "Ministerio de Educación y Ciencia" (Ref: FIS2004-06947-C02-01) and Generalitat Valenciana (Ref: GV06/077). José J. Vallés acknowledges a FPU grant from the Ministry of Education and Science.

Appendix A

This appendix justifies Eq. (3), which shows the effect of a scale change in the *PhFT* distribution. The *PhFT* contains information of all the orientations of the surfaces that defines a 3D object. It can be assumed that a 3D object is composed by multiples planar facets in different position and orientation. All of them define the object. Then, the *PhFT* of the complete 3D object is the contribution of all the *PhFTs* of the different facets.

Let consider one individual facet defined as

$$f(x, y) = (a \cdot x + b \cdot y + c) \cdot w(x, y) \quad (\text{A.1})$$

where $w(x, y)$ is a boundary function equals to one if the point (x, y) belongs to the plane and zero otherwise. Its phase Fourier transform is

$$\begin{aligned} \text{PhFT}_f(u, v) &= FT \left\{ e^{i \cdot f(x, y)} \right\} = \int_{\mathbb{R}^2} e^{i(ax+by+c)w(x,y)} e^{-i2\pi(ux+vy)} dx dy \\ &= \int_{(x,y) \in w} e^{i(ax+by+c)} e^{-i2\pi(ux+vy)} dx dy + \int_{(x,y) \notin w} e^{-i2\pi(ux+vy)} dx dy \\ &= e^{ic} \delta \left(u - \frac{a}{2\pi}, v - \frac{b}{2\pi} \right) \otimes W(u, v) \end{aligned} \quad (\text{A.2})$$

where $\delta(u, v)$ is the delta function, \otimes is the convolution and $W(u, v) = FT \{ w(x, y) \}$. Note from Eq. (A.2) that the integral outside the boundary area is infinite and the exponential is a periodical function, so the integral of the second term of Eq. (A.2) is zero. Then, the *PhFT* of an individual facet is centered in a peak given by the orientation of the normal to the plane and the shape of the peak is given by a certain factor of form which is the Fourier transform of the boundary. This result is similar to the interpretation of diffraction limited coherent optical imaging due to the pupil of the lenses.

A scale change can be described as

$$f'(x, y) = (a \cdot x + b \cdot y + c) \cdot w(x/k, y/k) \quad (\text{A.3})$$

where k is the scale factor. Note that a scale change has no influence in the orientation of the individual facets. The *PhFT* of the scaled object, $f'(x, y)$, is

$$PhFT_f(u, v) = k^2 e^{ic} \delta\left(u - \frac{a}{2\pi}, v - \frac{b}{2\pi}\right) \otimes W(ku, kv) \quad (\text{A.4})$$

The *PhFT* of the scaled object is given by a peak in the same position than the non-scaled object, but it is convolved by a scaled factor of form. Considering that a continuous 3D object is formed by numerous facets and the range of scale factors is not too large (i.e. k varies from 0.6-1.2), it can be shown from the experiments (see Fig. 1) that a scale change in the factor of form will be negligible in comparison with the influence of the peaks position, so in a first approximation we can assume that

$$PhFT_f(u, v) \cong k^2 e^{ic} \delta\left(u - \frac{a}{2\pi}, v - \frac{b}{2\pi}\right) \otimes W(u, v) = k^2 PhFT_f(u, v) \quad (\text{A.5})$$

Spherical nonlinear correlations for global invariant three-dimensional object recognition

José J. Vallés, Pascuala García-Martínez,* and Carlos Ferreira

Departament d'Òptica, Universitat de València, c/o Dr. Moliner, 50 46100-Burjassot, València, Spain

*Corresponding author: pascuala.garcia@uv.es

Received 30 April 2007; revised 13 September 2007; accepted 19 September 2007;
posted 20 September 2007 (Doc. ID 82605); published 18 October 2007

We define a nonlinear filtering based on correlations on unit spheres to obtain both rotation- and scale-invariant three-dimensional (3D) object detection. Tridimensionality is expressed in terms of range images. The phase Fourier transform (PhFT) of a range image provides information about the orientations of the 3D object surfaces. When the object is sequentially rotated, the amplitudes of the different PhFTs form a unit radius sphere. On the other hand, a scale change is equivalent to a multiplication of the amplitude of the PhFT by a constant factor. The effect of both rotation and scale changes for 3D objects means a change in the intensity of the unit radius sphere. We define a 3D filtering based on nonlinear operations between spherical correlations to achieve both scale- and rotation-invariant 3D object recognition. © 2008 Optical Society of America
OCIS codes: 070.5010, 070.2590, 100.6890.

1. Introduction

Three-dimensional (3D) optical information systems are receiving considerable attention nowadays because of their vast potential for applications, such as object recognition, image encryption, and 3D display [1–6]. Generally, the information of a 3D object is contained in a set of triplets with the (x, y, z) coordinates of the object surface. Object detection means scanning over the six degrees of freedom (three rotations and three translations) for the best matching with the reference set of points. However, a significant reduction of the complexity can be achieved by collecting a few views of the object as planar images. For example, a range image can contain full 3D geometrical information, although only from a given point of view. Range images are obtained from most of the 3D digitizing techniques [2,3]. The encoding of the depth (range) information has been used in the literature to extend the possibilities of range image recognition using the phase Fourier transform (PhFT) [6,7]. The idea is to detect the planar surface of a 3D range image when the Fourier transform of a phase-coded image (the amount of phase being pro-

portional to the elevation) is calculated. Another application of this technique is face recognition, for which the range image plays an important role since it contains richer and more accurate shape information than those in a two-dimensional (2D) intensity image [8].

However, a main limitation of most 3D matching methods is the limited tolerance to rotations and scale changes in the object. In [9] and [10] the authors showed a method for detection and orientation evaluation of 3D objects that suffer an arbitrary rotation in space. The method is based on computing a map or a signature of the 3D object based on the local orientation of its surfaces by means of range images. Moreover, the detection is done by correlation techniques [9,10]. The map (named 3D object orientation map or 3DOOM) contains the information of all viewpoints for any possible orientation of the 3D object. In this paper we use the 3DOOM information to perform 3D filtering based on vector space representation to achieve tolerance to scale changes in addition to rotation invariance. The idea is to combine some particular correlations on the 3DOOM unit sphere with nonlinear local statistical operations that we have defined for intensity-invariant 2D pattern recognition [11,12]. We will show that scaling a 3D range object implies the multiplication of the amplitude of

the PhFT distribution by a constant factor. Thus, a change of scale is transformed into a change in the illumination (multiplication by a constant). In common correlation, if the illumination model consists of multiplying a target by an unknown constant factor, the correlation peak will change by the same amount. Lefebvre *et al.* [11] defined a nonlinear filtering method known as the locally adaptive contrast-invariant filter (LACIF), which is invariant under any linear intensity transformation or, to be more precise, affine intensity transformation. This LACIF operation uses three 2D correlations involving local statistics and nonlinearities. It was applied directly to scenes containing unsegmented targets. One of the advantages of the LACIF method is that no *a priori* information about the constant values involved in the linear illumination model is assumed. The LACIF has noteworthy properties in the context of vector space interpretation [11,12]. We have also defined a local angular distance correlation connected with LACIF for illumination-invariant 3D range images [13]. Changing lighting conditions for a 3D object implies multiplying Lambertian reflectance functions obtained with distant light sources by a constant factor. Those reflectance functions are expressed in terms of an orthonormal basis defining a vector subspace where local angular distance correlation measures the angle between a target and a projection to that subspace [13]. Recently, we used a phase Fourier vector model for scale-invariant 3D image detection [14].

In this paper we propose to extend 2D LACIF correlation to 3D spherical surfaces to detect a 3D object that has been arbitrarily rotated and scaled simultaneously. The method is based on the definition of a certain spherical nonlinear correlation. In Section 2 we review the basis of the 3DOOM calculation. The effect of a change in the scale of a 3D rotated object is presented in Section 3. Invariant spherical correlation is described in Section 4. The application of the method and the results are in Section 5.

2. Three-Dimensional Object Orientation Map (3DOOM)

We define a range image as $x = f(y, z)$. It contains the depth information of an object from a given view line

that defines the x axis. The range image is considered a set of facets that may be described by their normals to the surface. The codification of the depth information has been used in the literature to extend the possibilities of range images for pattern recognition [6,7,9,10]. Those techniques are based on phase coding. We encode the range as phase as follows:

$$P(y, z) = \exp[ikx(y, z)], \quad (1)$$

where k is a constant to adjust the phase slope of the object. A way to deal with range images keeping translation invariance is to use their Fourier transforms. The Fourier transform of the phase-encoded range image (PhFT) is

$$\text{PhFT}(\theta, \varphi) = F_{2D}\{\exp[ikx(y, z)]\}, \quad (2)$$

where F_{2D} stands for 2D Fourier transform and (θ, φ) are conventional spherical coordinates in the Fourier domain. A detailed description of the coordinate changes in the PhFT domain can be found in [9] and [10]. The k factor gives the scaling of the Fourier transform frequencies, and we will assume $k = 1$. The value of the k factor is chosen to ensure that the important frequency information fits the size of our images. After a phase codification, a planar surface will become a linear phase factor whose Fourier transform will give a peak located in a certain position. Let us consider Fig. 1(a), which represents our coordinates and axis definitions. The x axis defines the view line, the polar axis is the z axis, and the azimuth is measured with respect to the y axis. Figure 1(b) shows a range image of a certain nonrotated 3D object, and Fig. 1(c) shows the amplitude of the PhFT for the (θ, φ) coordinates. Note that from now on we will assume that the amplitude of the PhFT is expressed in angular coordinates, where a certain location of a peak gives directly the orientation of a corresponding facet [10]. The effect of a rotation of the 3D object around the z axis implies a shift in the amplitude of the PhFT domain defined in Eq. (2) [9]. If the full 3D information of the object is known, it is possible to calculate in advance the amplitude of the

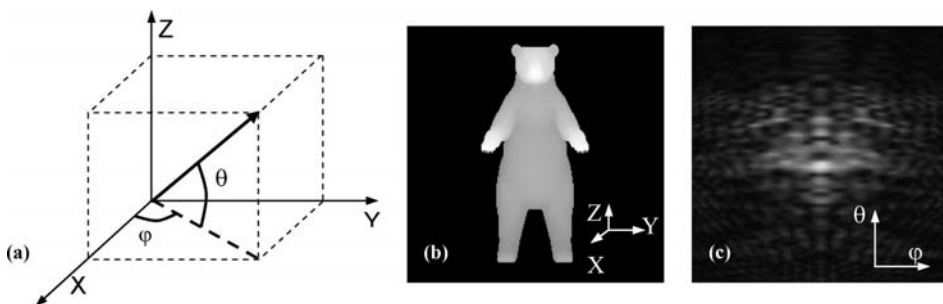


Fig. 1. (a) Definition of angular coordinates. (b) Example of a nonrotated and nonscaled range image. (c) Amplitude of the PhFT of the range image in (θ, φ) coordinates.

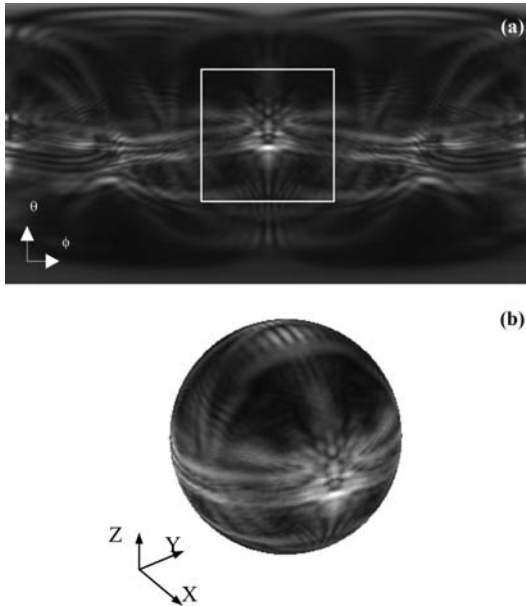


Fig. 2. (a) Planar 3DOOM showing the position of Fig. 1(c) in a white square. (b) Full 3DOOM on the unit sphere for all possible orientations of Fig. 1(b).

PhFT for any possible orientation. This makes it possible to build the 3DOOM that contains the information for all viewpoints by displacing and pasting on a large image the different amplitudes of the PhFTs expressed in spherical coordinates. The angular variations of θ and φ are $[0, 2\pi]$ and $[-\pi, \pi]$, respectively. This 3DOOM contains the information about the normals for all possible angles in a single image [9,10]. Moreover, we have performed the 3DOOM by changing θ and φ at intervals of 1° . Figure 2(a) shows the planar 3DOOM in angular coordinates showing the location of the amplitude of the PhFT of the reference range image [see Fig. 1(b)]. Note that the rotation of Fig. 1(b) is $\varphi = 0^\circ$ and $\theta = 0^\circ$, so its amplitude of the PhFT is shown in the center of Fig. 2(a). It is marked with a white square. Figure 2(b) is the spherical view of the 3DOOM on the unit sphere. The possibility to represent the map on a plane prevents the possibility of making template matching by correlation just by conventional correlation on the plane (θ, φ) . A rotation of the object will be converted into a rotation on the 3DOOM in the unit sphere, but the planar representation will not undergo a simple transformation because the mapping will be affected by changing the nonlinear coordinates from spherical to rectangular coordinates [9]. Moreover, to achieve fully rotation-invariant recognition, a correlation between the 3DOOM of different objects can be defined [10]. In the further sections we review the coordinate transformations from rectangular to angular variables to define the correlation between spheres. This is what we name a spherical correlation.

3. Scale Changes for 3DOOM

The effect of a scale in a range image only leads to a change by a constant global factor in the amplitude of the PhFT domain, with no change in the pattern distribution [14]. Let us assume that the scaled range image is given by

$$x'(y, z) = x\left(\frac{y}{m}, \frac{z}{m}\right),$$

where m is the scale factor. Although the scale is changed, the shape of the object will remain the same as well as the orientation of the normals, except for a constant value, which is related to the scale factor [6]. It can be shown mathematically that the amplitude of the PhFT of a scaled object can be expressed in terms of the amplitude of the PhFT of the nonscaled object as [14]

$$|\text{PhFT}_{x'}(\theta, \varphi)| \cong m^2 |\text{PhFT}_x(\theta, \varphi)|. \quad (3)$$

Thus, a scale change in the range image will only lead to a multiplication of the amplitude of the non-scaled PhFT domain by a constant factor, with no change in the pattern distribution of the PhFTs. Only a global change in the intensity is visualized. To verify Eq. (3) we have simulated a change of scale for the range image shown in Fig. 1(b) and the effect of the amplitude of the PhFT for that range image. The result is shown in Fig. 3. Figure 3(a) shows a range image and Fig. 3(b) its amplitude of the PhFT distribution drawn on the unit sphere for that particular

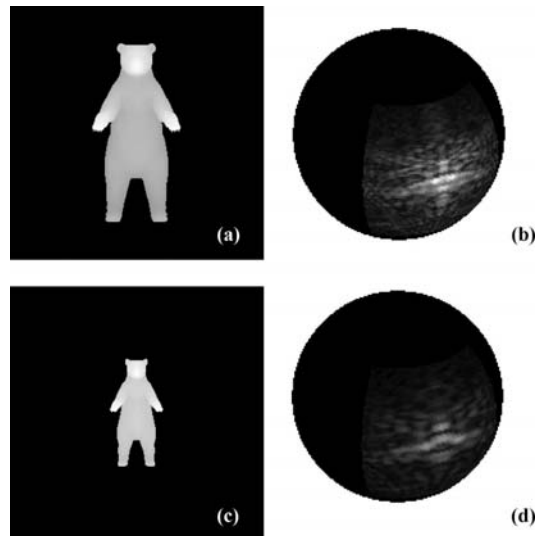


Fig. 3. (a) Range reference image at $\theta = 0^\circ$ and $\varphi = 0^\circ$. (b) Amplitude of the PhFT distribution of Fig. 3(a) drawn on the unit sphere. (c) Scaled range image (scale factor 0.6) at $\theta = 0^\circ$ and $\varphi = 0^\circ$. (d) Amplitude of the PhFT distribution of (c) drawn on the unit sphere.

orientation of the 3D object. Similarly, Fig. 3(c) shows a scaled version of Fig. 3(a); the scale factor is 0.6. The effect of scaling, according to Eq. (3), is a change in the amplitude pattern of the PhFT distribution drawn on the unit sphere [see Fig. 3(d)]. The particular view drawn in the 3DOOM for Fig. 3, corresponds to Fig. 1(b) in which the object has no rotation ($\theta = 0^\circ, \varphi = 0^\circ$). Note that we only represent that particular view as a patch in the 3DOOM. On the other hand, as the surface of the 3D object is a bounded curve, the change of scale may cause some problems due to digitalization processes. Then, some information can be lost in the mapping of the 3DOOM; nevertheless the description is sufficient to provide detection, as we will show in Section 5.

We will define a method to detect a 3D object with independence of the rotation and the scale. The method is based on performing spherical correlations between the 3DOOM distribution and different patches on the unit sphere as well as the introduction of a nonlinear function to incorporate the intensity-invariant detection to the 3DOOM distributions.

4. Invariant Spherical Correlation

For a conventional 2D correlation, the images to be correlated as well as the correlation plane are 2D distributions. The 2D matching process is performed by scanning the spatial coordinates of one of the two distributions on the other. Similar concepts can be applied for correlations between unit spheres, where the scanning process using shifting operations are replaced by rotations by means of Euler angles (α, β, γ) [10]. The definition of Euler angles is shown in Fig. 4. A mathematical definition of the correlation on the unit sphere between two real distributions $f(\theta, \varphi)$ and $g(\theta, \varphi)$ defined on the surface of the sphere is [15]

$$C_{fg}(\alpha, \beta, \gamma)_{\in SO(3)} = f(\theta, \varphi)_{SO(3)}^* g(\theta, \varphi) \equiv \sum_{j,m,m'} d_{m,m'}^j(\beta) F_{j,m'} G_{j,m} \times \exp(im\gamma) \exp(im\alpha), \quad (4)$$

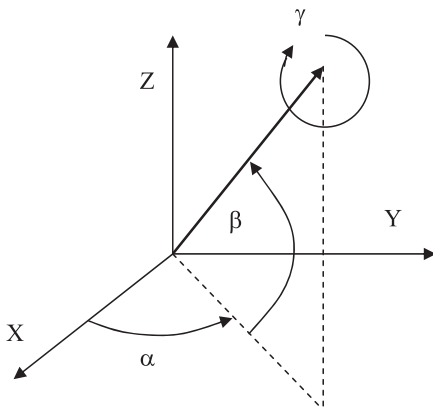


Fig. 4. Definition of Euler angles.

where $SO(3)$ is the rotation group, ${}_{SO(3)}$ is the spherical correlation, $d_{m,m'}^j(\beta)$ are the reduced rotation matrices [16], and $F_{j,m'}$ and $G_{j,m}$ are the spherical Fourier transforms defined as [10]

$$F_{j,m'} = \int f(\theta, \varphi) Y_{lm}^*(\theta, \varphi) \sin(\theta) d\theta d\varphi, \quad (5)$$

where $Y_{lm}(\theta, \varphi)$ are the spherical harmonics functions. Note that on the unit sphere the Fourier expansion involves the use of spherical harmonics instead of linear exponential functions. Following [10], f and g are the amplitude of the PhFT from a given (unknown) point of view and the 3DOOM, respectively. The output correlation will be 3D, and the location of the correlation admits a rapid implementation using the fast Fourier transform algorithm [17].

In this paper, we apply the correlation on the unit sphere [Eq. (4)] to intensity-invariant 3DOOM detection. The basic idea is to extend the 2D LACIF process, which is invariant under any affine intensity transformation [11,12], to the 3D spherical process. An affine transformation of intensity over a unit sphere target can be expressed as

$$g(\theta, \varphi) = af(\theta, \varphi) + bw(\theta, \varphi), \quad (6)$$

where a and b are unknown constants, $f(\theta, \varphi)$ is the amplitude of the PhFT from a given point of view [see Fig. 3(b), for instance] and $w(\theta, \varphi)$ is a binary 3D object support that is equal to unity over the support of the target $f(\theta, \varphi)$ and equal to zero everywhere else.

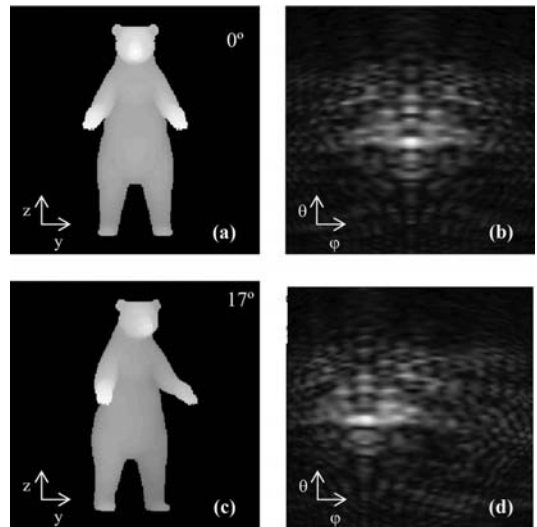


Fig. 5. (a) Range image of a 3D object rotated 0° around the z axis. (b) Amplitude of the PhFT obtained for the range image shown in (a). (c) Range image of the 3D object rotated 17° around the z axis. (d) Amplitude of the PhFT obtained for the range image shown in (c).

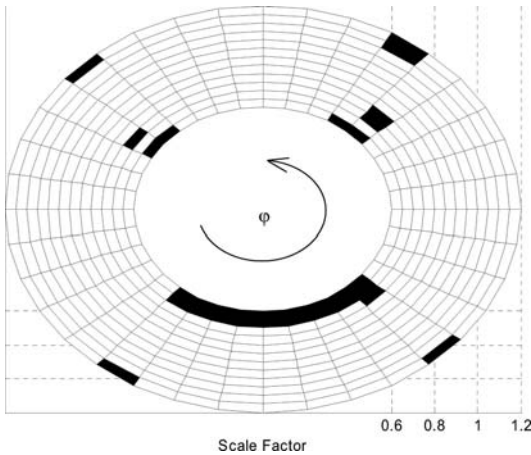


Fig. 6. Polar detection diagram showing LACIF output thresholded to a value of 0.7 for the reference range object. The radial coordinate is the scale factor [0.6–1.2] and the angular coordinate is φ varying between $[0-2\pi]$. The scale factor is changed every 0.05.

The final expression for the 3DOOM LACIF is

$$L_{gf}(\alpha, \beta, \gamma) = \frac{[g(\theta, \varphi) *_{so(3)} f_o(\theta, \varphi)]^2}{\sqrt{N} [g(\theta, \varphi) *_{so(3)} w(\theta, \varphi)] - [g(\theta, \varphi) *_{so(3)} w(\theta, \varphi)]^2} \quad (7)$$

where $L_{gf}(\alpha, \beta, \gamma)$ is the 3DOOM LACIF correlation, N is the area involved in the particular section on top of the unit sphere, and $f_o(\theta, \varphi) = f(\theta, \varphi) - \mu_f w(\theta, \varphi)$, where μ_f is the mean of $f(\theta, \varphi)$. Then, $f_o(\theta, \varphi)$ is a zero-mean image. If $g(\theta, \varphi)$ is expressed in terms of affine transformations ($g = af + b$) as Eq. (6), then the correlation of Eq. (7) is equal to unity [11], so the method is invariant to the intensity transformation

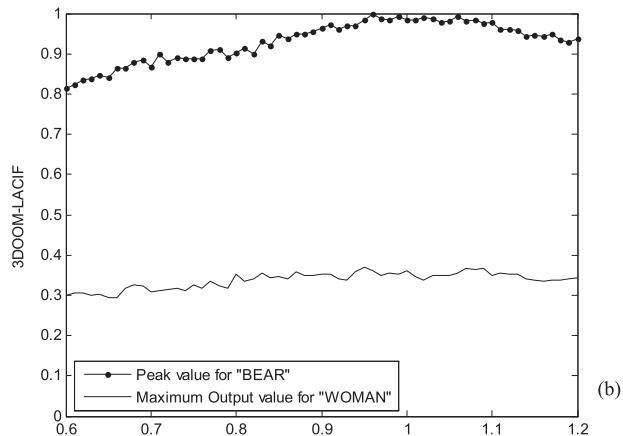
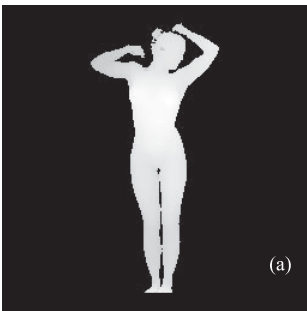


Fig. 7. (a) Range image of another object (woman) rotated $\alpha = 0^\circ$ and $\beta = 0^\circ$. (b) Correlation output for different scale factors and a rotation of $\alpha = 0^\circ$ and $\beta = 0^\circ$ for the two images.

considered. According to Eq. (3), the relation between scale factor, m , and intensity factor, a , is $m^2 = a$. In addition, constant b is equal to zero. The values of the constant factors have no importance in recognition because the LACIF method will obtain intensity invariance whatever the constant values are.

Although Eq. (7) implies a kind of locally adaptive 3D processing, the operations carried out involve only the calculation of three spherical correlations according to Eq. (4). Our method will detect a range target that has been rotated and scaled arbitrarily. To be more precise, we will correlate the 3DOOM with the target that will be a patch on the surface of the unit sphere. If the rotated target coincides with the reference object, the target patch will be contained in the 3DOOM. Moreover if the target is also scaled, it will give a change in the intensity of the patch and it will also be detected by the intensity-invariant correlation of Eq. (7). In addition, the LACIF techniques also yield good discrimination against false targets, as we will show in the following section.

5. Results

Because the number of degrees of freedom to be tested is significant, it could be complex to represent all together. Because of that, we have carried out two sets of experiments. The first experiment deals with the detection of a 3D target that has been scaled and rotated around the z axis and the second experiment is applied to fully rotational and scaled 3D targets using the invariant spherical correlation described in Section 4.

A. Rotation around the z Axis and Scale Changes

Coming to the PhFT definition of a range image in terms of angular coordinates, a rotation around the z axis implies just a shift of the 3DOOM along the φ coordinate. This effect is shown in Fig. 5. As we can see, any shift in the amplitude of the PhFT domain, which corresponds to a specific rotation, will be con-

tained in the planar 3DOOM [Fig. 2(a)]. Because of that, to detect a particular view of the object, common 2D linear correlation can be applied between the planar 3DOOM of Fig. 2(a) and the PhFT amplitude of the corresponding rotated object. The basic idea of this recognition method is to take advantage of the shift invariance properties of common linear 2D correlation [9]. Moreover, adding the scale invariance to the previous issue implies a change in intensity of the

PhFT amplitude distributions. We have applied the 2D LACIF correlation described by us in [14] between the complete planar 3DOOM and a particular rotated and scaled target to obtain both scale and rotation around the z axis invariances. Two distributions are involved in LACIF correlation: For the reference input we will use $f(\theta, \varphi) = \text{Planar 3DOOM}(\theta, \varphi)$, that is, the planar 3DOOM. For the input target, $g(\theta, \varphi) = |\text{PhFT}_{x_R}(\theta, \varphi)|$ we used the amplitude of the PhFT

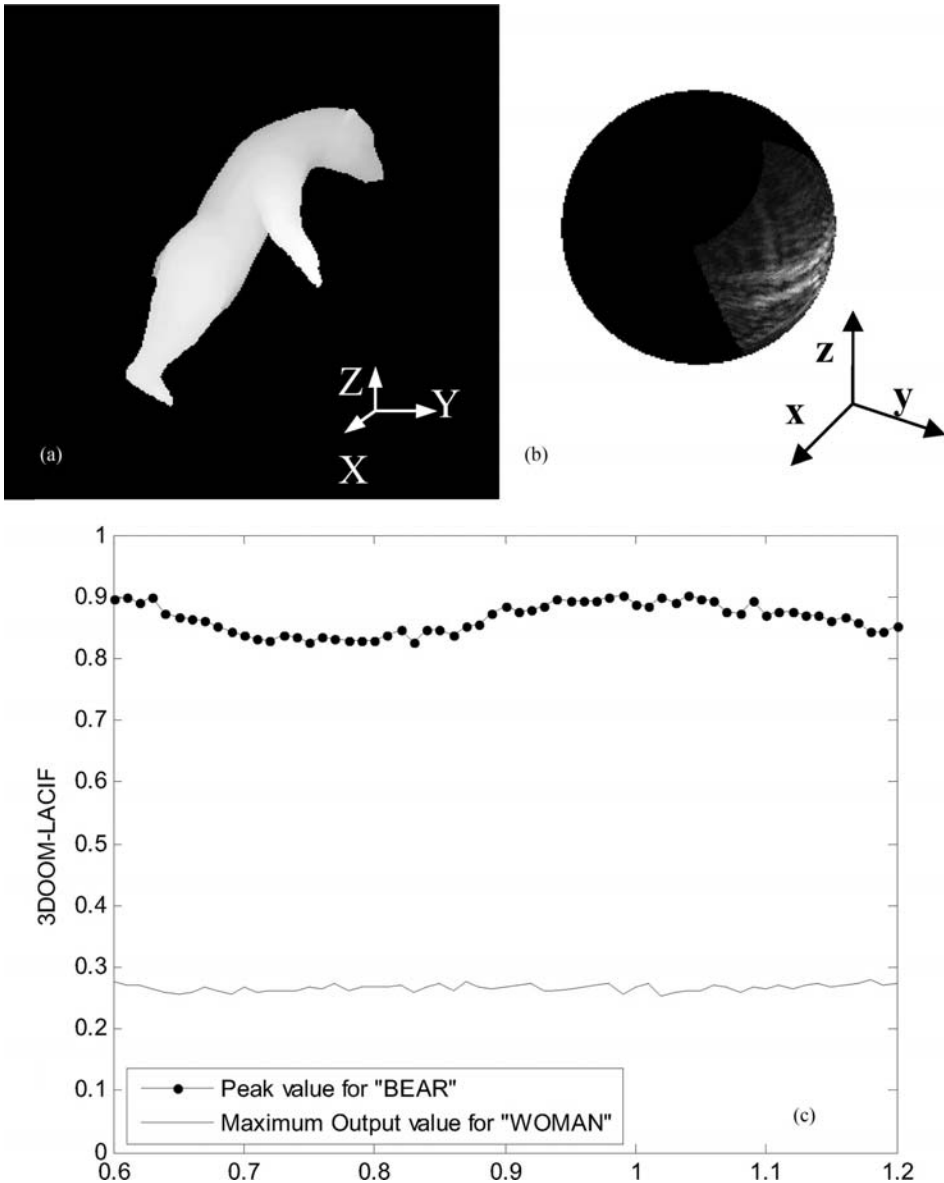


Fig. 8. (a) Range image of the 3D object rotated with $\alpha = 90^\circ$ and $\beta = -30^\circ$. (b) Amplitude of the PhFT of (a) drawn on the unit sphere. (c) Correlation output for that fixed orientation and different scale values.

of the specific scaled and rotated object. Note that $x_R'(y, z)$ is the range image of the object that has been rotated and scaled.

Figure 6 shows the correlation results with LACIF. We have used a polar diagram to show all possible rotations around the z axis as well as different scale factors. The radial coordinate is the scale and the angular coordinate is the rotation around the z axis. For each scale and rotation value, we calculate the

2D LACIF correlation peak value. We remind that LACIF correlation values are normalized to unity by definition. Finally, a threshold is applied to binarize the final result. We have applied a 0.7 threshold value, so, if the correlation peak value is above 0.7, then we associate value 1, and 0 otherwise. We have changed the scale range between [0.6–1.2] and the rotation angle $\varphi \in [0, 2\pi]$ for the bear 3D range image, respectively. In spite of a few black sectors, the

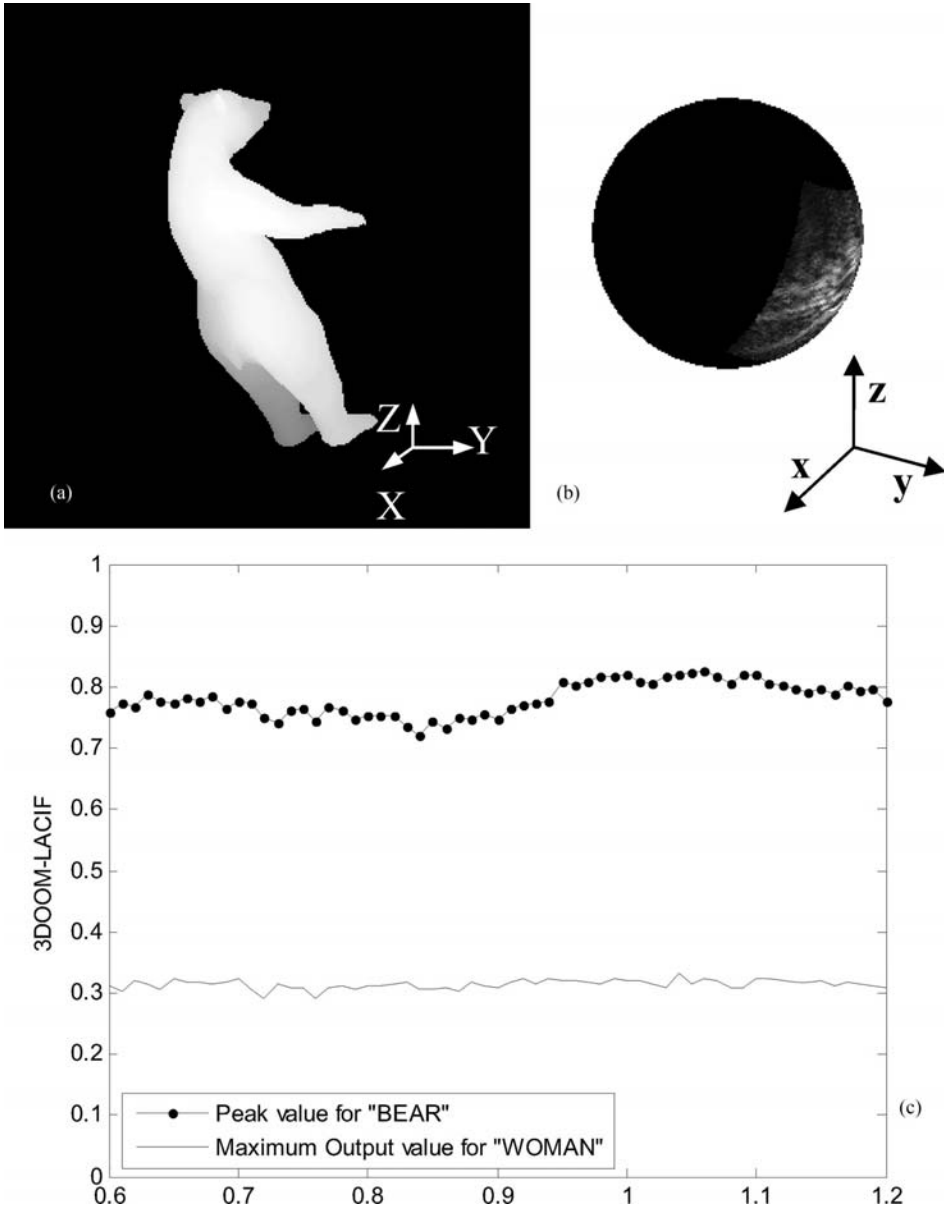


Fig. 9. (a) Range image of the 3D object rotated with $\alpha = 120^\circ$ and $\beta = 30^\circ$. (b) Amplitude of the PhFT of (a) on the unit sphere. (c) Correlation output for that fixed orientation and different scale values.

method has detection tolerance to changes of rotations around the z axis and to scale.

Because the shift of the amplitude of the PhFT in the planar 3DOOM distribution is not exact, and due also to the digitalization process, the correlation peak output is not exactly one. Thus, the method exhibits a certain tolerance. Because of that, we have applied a threshold. Another interpretation of the polar recognition diagram is that almost all combinations between changes of scale and φ angle are above that value. Only for small objects (inner radius or scale factor of 0.6) some combinations are below 0.7, and then it appears in the polar diagram representation as black sectors. In spite of those few sectors, the method has detection tolerance to changes of rotations around the z axis and to scale. Moreover as the LACIF process applied in this subsection is also defined in terms of 2D correlations, conventional optical correlators, such as the Vander Lugt correlator or joint transform correlator architectures, can be used. The next subsection will be related to applying the technique when the 3D object is rotated arbitrarily.

B. Global Rotation and Scale Changes

In this section we show the results of the invariant spherical correlation expressed in Eq. (7). The reference range image used to test the rotation and scale changes is in Fig. 1(b). To simplify the interpretation of the results, without loss of generality, the rotations are limited to the rotations around the z axis and the y axis. Moreover, for giving a general interpretation of the method it would be convenient to show the correlation output value for the three variables, i.e., the (θ, φ) angles and the scale factor. This is impossible to illustrate all together, since it would be a four-dimensional representation. To provide a clear explanation of the results, we have chosen some orientations of the reference range image and we have calculated the correlation output when the scale is the variable that changes. In addition, we have tested the discrimination abilities of the method using another object with the same orientation as the reference. Although the coordinates of the PhFT amplitude distributions are defined using (θ, φ) angles, the fully rotational criterion will be expressed using the Euler system [10]. Moreover, we have chosen as the reference target the average between different scaled reference targets to minimize the possible sampling errors due to the scaling digitalization process. This is the reason why in all the results the correlation value will not reach unity.

Figure 7 is the first group of results when the rotation of the object is $\alpha = 0^\circ$ and $\beta = 0^\circ$ [see Fig. 1(b)]. Figure 7(a) shows the woman range image to test discrimination, and Fig. 7(b) is the correlation results when the scale changes between ranges [0.6–1.2]. The output values are almost constant for all scale factors, whereas the false range object is clearly isolated.

Different rotations are chosen and the scale changes analyzed in Figs. 8 and 9. Note that Figs.

8(b) and 9(b) show the location of the rotations in 3DOOM, we also mark in the figures the axis coordinate system (x, y, z) to clarify the exact position of the area on top of the sphere. The LACIF 3DOOM correlation implies spherical correlations between those particular views and the complete 3DOOM shown in Fig. 2(b). Although we finally present three particular rotations, the method was successfully tested for all possible (α, β) rotations.

6. Conclusion

Scale- and rotation-invariant 3D object recognition has been converted to intensity-invariant pattern recognition. A 3DOOM distribution that contains all possible orientations of the object in terms of the amplitude of the phase Fourier transform of range images was calculated. From that description it is possible to achieve detection and rotation estimation even when only partial object information is presented by performing a correlation between unit spheres. However correlation is proportional to target intensity, so when multiplying a target by an unknown constant factor, the correlation peak height will change by the same amount. A scale change of rotated range images will imply a change of intensity in the 3DOOM unit sphere. We define a nonlinear operation based on combining three spherical correlations in a nonlinear way to achieve both scale- and rotation-invariant 3D object recognition. Various experiments were carried out to validate the approach. We successfully tested the method when other false targets were used.

The authors acknowledge Henri H. Arsenault, José J. Esteve-Taboada, and Javier Garcia for useful discussion. This work was supported by the Spanish Ministerio de Educación y Ciencia under the project FIS2007-60626. José J. Vallés acknowledges a Formación de Profesorado Universitario grant from the Ministry of Education and Science.

References

1. B. Javidi, ed., *Image Recognition and Classification: Algorithms, Systems, and Applications* (Marcel Dekker, 2002).
2. M. Rioux, "Laser range finder based on synchronized scanners," *Appl. Opt.* **23**, 3837–3844 (1984).
3. M. Takeda and K. Mutoh, "Fourier transform profilometry for the automatic measurement of 3-D object shapes," *Appl. Opt.* **22**, 3977–3882 (1983).
4. J. Rosen, "Three-dimensional electro-optical correlation," *J. Opt. Soc. Am. A* **15**, 430–436 (1998).
5. T. Poon and T. Kim, "Optical image recognition of three-dimensional objects," *Appl. Opt.* **38**, 370–381 (1999).
6. E. Paquet, M. Rioux, and H. H. Arsenault, "Invariant pattern recognition for range images using the phase Fourier transform and a neural network," *Opt. Eng.* **34**, 1178–1183 (1995).
7. E. Paquet, P. Garcia-Martinez, and J. Garcia, "Tridimensional invariant correlation based on phase-coded and sine-coded range images," *J. Opt.* **29**, 35–39 (1998).
8. S. Chang, M. Rioux, and J. Dorney, "Face recognition with range images and intensity images," *Opt. Eng.* **36**, 1106–1112 (1997).

9. J. J. Esteve-Taboada and J. García, "Detection and orientation evaluation for three-dimensional objects," *Opt. Commun.* **217**, 123–131 (2003).
10. J. Garcia, J. J. Vallés, and C. Ferreira, "Detection of three-dimensional objects under arbitrary rotations based on range images," *Opt. Express* **11**, 3352–3358 (2003).
11. D. Lefebvre, H. H. Arsenault, P. Garcia-Martinez, and C. Ferreira, "Recognition of unsegmented targets invariant under transformations of intensity," *Appl. Opt.* **41**, 6135–6142 (2002).
12. H. H. Arsenault and P. García-Martínez, "Intensity-invariant nonlinear filtering for detection in camouflage," *Appl. Opt.* **44**, 5483–5490 (2005).
13. J. J. Vallés, J. García, P. García-Martínez, and H. H. Arsenault, "Three-dimensional object detection under arbitrary lighting conditions," *Appl. Opt.* **45**, 5237–5247 (2006).
14. J. J. Vallés, P. García-Martínez, J. García, and C. Ferreira, "Phase Fourier model for scale invariant three-dimensional image detection," *Opt. Express* **15**, 7818–7825 (2007).
15. B. D. Wandelt and K. M. Górski, "Fast convolution on the sphere," *Phys. Rev. D* **63**, 123002 (2001).
16. J. J. Sakurai, *Modern Quantum Mechanics* (Addison Wesley, 1985), pp. 221–223.
17. J. R. Driscoll and D. M. Healy, "Computing Fourier transforms and convolutions on the 2-sphere," *Adv. Appl. Math* **15**, 202–250 (1994).