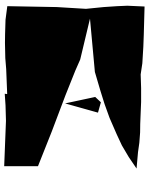

Universitat Autònoma de Barcelona

Seguimiento visual robusto en entornos complejos

Memoria de Tesis presentada por **Javier Varona Gómez** en la Universitat Autònoma de Barcelona para obtener el título de **Doctor en Informática**.

Bellaterra, 21 de noviembre de 2001

Director: **Dr. Juan José Villanueva Pipaón**
Universitat Autònoma de Barcelona
Dept. Informàtica, Centre de Visió per Computador (CVC)

This document was typeset by the author using LaTeX2ε.

The research described in this book was carried out at the Computer Vision Center, Universitat Autònoma de Barcelona.

ISBN 84-932156-1-9

Printed by Servei de Publicacions, UAB

Para la mujer más maravillosa, mi madre...

Agradecimientos

Una Tesis Doctoral es uno de los trabajos intelectuales individuales más largo al que se puede enfrentar una persona. Sin embargo, aunque puede llevar a un estado de absoluta soledad, siempre se está rodeado de personas que de una forma u otra influyen en su realización. En mi caso personal este hecho aún es más importante porque todo lo que ocurre a mi alrededor me afecta. Incluso en los peores momentos de este trabajo el que ocurra algo a mi alrededor influye en el desarrollo de mi trabajo. Cualquiera que lea estas líneas pensará que no llegaré muy lejos siendo de esta manera. Pero aquí presento una Tesis, que considero el reto intelectual más grande que alguien se puede plantear en su vida. La cuestión no es sólo aprender, el problema es "crear". Por eso, en estos momentos siento una satisfacción doble, el poder haber realizado este trabajo y el haber ayudado a todo aquel que durante este tiempo me ha necesitado. De todas maneras, este trabajo ha contado con la ayuda fundamental de muchas personas, y a todas ellas y a las que seguramente me olvidaré quiero dedicarlo con la oportunidad que me brinda este espacio dentro de la Tesis.

En primer lugar a la persona que confió que yo podía llegar a realizar este trabajo, el Dr. Juan José Villanueva. Que siguió confiando en los peores momentos que he pasado durante la Tesis y cuya aportación final a este trabajo ha sido decisiva para su conclusión. Gracias por todos los consejos e ideas que ha realizado en esta Tesis y que espero que haya sabido reflejar bien. Seguramente, si hubiera sabido aprovechar muchos de sus consejos hace tiempo que habría finalizado el trabajo. Gracias también por la creación del Centro de Visión por Computador, el que durante mucho tiempo ha sido mi segunda casa y donde tengo muchos de mis mejores amigos y un gran número de excelentes compañeros.

Precisamente, el hecho de que mi segunda casa haya sido el CVC, implica que muchas de las personas que están en él, o que durante estos años han pasado por aquí, hayan intervenido de una manera u otra a este trabajo. Uno de los momentos culminantes de este trabajo fue una discusión durante una noche de verano. Allí estaban Albert, Paco y Poal escuchando el planteamiento final que quería dar a este trabajo. Escuché todos los consejos que me dieron y parte del contenido de esta memoria es el resultado de aquella noche. Pero su aportación fundamental ha sido en el plano personal. A Albert le tengo que agradecer su compañía en el despacho y las innumerables veces que hemos tenido que darnos ánimos mútuamente para seguir adelante. Espero que consigas todo lo que te propongas a partir de ahora, estoy seguro

que lo conseguirás. A Paco le tengo que agradecer muchas cosas, pero sobre todo su lealtad conmigo. Siempre ha estado a mi lado en todo momento. Gracias por ser como eres y no dejes que nada cambie tu manera de ser. Espero que algún día consigas demostrar a todo el mundo todo lo que vales. Y de Poal, que puedo decir que no le haya dicho ya, sus pensamientos son geniales. Es difícil encontar a alguien tan sensible y con un corazón tan grande. Espero que encuentres lo que buscas porque te lo mereces.

A Xavi y Maria les tengo que agradecer su apoyo durante todos estos años. De Maria me asombra su ilusión por su trabajo, muchas veces sin tiempo para poder escuchar sus valiosos consejos y sus ánimos, hace que compartas su ilusión con ella. De Xavi su pasión por todo lo que hace, da lo mismo lo que tenga que hacer, lo importante es que pone todo de su parte para que salga bien. Nunca le podré agradecer suficiente la invitación que me hizo para ir a ver un partido del Barcelona, me hizo el hombre más feliz del mundo. A Ramon Baldrich le tengo que agradecer su compañía este último verano y espero que no olvide la comida pendiente que tenemos. A Felipe y a Josep, el haber podido trabajar con ellos en ICAR, y por todo lo que he aprendido de ellos desde que entré por la puerta de su despacho. Siento no haber podido ayudaros a tirar adelante el proyecto, pero estoy seguro que saldrá bien, y sabeis que podeis contar conmigo a partir de ahora. A Judit por su compañia en el despacho y por los ánimos y consejos que me ha venido dando desde que la conozco.

A Andresito y Dani, compañeros de penas al principio de este trabajo su ayuda en todo momento que los he necesitado. A todo el equipo de fútbol del CVC: Antonio, Ernest, Coen, Marc y Juanra por recordarme cada martes que había partido. Al Miguel por convencerme de participar en la Liga Fantástica para que él no quedara el último. Que sepas que este año te vigilaré muy de cerca. A Eva Costa por que ha sido un placer dirigir su proyecto y su trabajo final de máster. A Jordi Saludes sus valiosas revisiones. A Joan Serrat el haber confiado dos veces en mi, ojalá pudiera tener su forma de trabajar. A Ernest Estruga su simpatía y el cartel que hizo para mi fiesta de cumpleaños, cuando se pone es un artista. A Fernando por compartir conmigo su amistad y sus charlas filosóficas, cuanta razón tiene cuando dice que no somos el centro del mundo.

Al resto de gente del CVC, que siempre han tenido un momento cuando los he necesitado: Enric, Jordi Vitrià, Petia, Xavi Binefa, Xavi Sánchez, Ricardo, Gemma, Oriol, David, Marco, Anna, Jordi Arnabat, Debora, Xavi Otazu, Cristina, Robert, Juanma, Guillamet, Xevi, Llus Barceló, Ramon Felip, David Rotger, Antonio Tovar, Àlex, Agnes, Francesc, Joan Queralt y Vicente. Y a los que han pasado por él: Carme, Dina, Thais, Joan Ramon y Nacho. A los Vyra: Sergi, Miquel, Ochi y Franpa, ya vereis como el sistema se vende como churros. A Silvia porque siempre es muy agradable charlar con ella. A Selma por las charlas por icq desde los USA y sus consejos siempre que le cuento mis problemas.

A todo el equipo de administración y soporte porque han sido y son fundamentales para que el CVC funcione bien. Joan Masoliver, el auténtico mago que hace que todo funcione. Pili porque hace que siempre encuentre el libro o el artículo que necesito.

A Maria José por estar siempre dispuesta a ayudarme cuando he tenido problemas económicos y por permitir que moleste a sus pupilos cuando hacía un descanso. A Monste porque es maravillosa, como persona y en su trabajo, ojalá hubiera más gente como ella en el mundo. A Pedro porque siempre me ha tratado bien desde que llegó y porque no olvidaré nunca que me limpió el azucarero. A Raquel por escucharme siempre que la he necesitado y porque siempre me deja pasar primero por la puerta. A Loli por darme los buenos días cuando hacía el turno de noche. A Carme Ramírez y a Toni Guerra por ayudarme en todos los papeleos universitarios.

A Ana Celia, como compañera de trabajo por su sonrisa cuando llego al CVC, sólo eso es suficiente para alegrarme el día, y por ayudarme siempre en las tareas administrativas que he tenido que hacer. Espero que todo el mundo llegue a apreciar lo buena que eres en tu trabajo. A Ana Celia como amiga especial por devolverme a la vida cuando pasaba por mis peores momentos personales. Por darme fuerzas para poder levantarme por las mañanas para ver su sonrisa. Por todos los momentos que he estado con ella y que han sido fundamentales para soportar la presión de la parte final de este trabajo. Gracias por todo y no cambies nunca, eres genial.

A todos mis amigos que han hecho que durante mi tiempo libre me pudiera olvidar de la Tesis. Pepín por estar a mi lado en los buenos y malos momentos. Por su nobleza, que hace que pueda confiar en él a ciegas. Por haber confiado en mi todos estos años. A Andreu, mi maestro, por enseñarme a vivir en el "exterior". A Mari Pepi y Ernesto porque son geniales. A Ricard, Maria, Loli, Cristian, Nuri, Azu y Vanessa por todas las noches de fiesta y las cenas de los domingos. A Jesús por mantener mi amistad aunque le he tenido muy olvidado.

A mi "familia"de Vilanova: Carmen y Pepín, porque han sido como unos padres para mi. Dani y Esther, por ayudarme y animarme cuando los he necesitado. Mónica y Arcadi por su amistad y los momentos que hemos compartido. Os hecho de menos a todos. A Andreu, que aunque no le conozco tiene algo de especial para mi. A Luci, por compartir tantos años conmigo. Pasara lo que pasara, no olvidaré nunca los momentos que pasé con ella. Me siento orgulloso de haber participado en que cumplieras tu sueño. Por ayudarme en lo que has podido en este final de Tesis.

A toda mi familia. Especialmente a mi tía Rosi por el placer de escucharla y por su ayuda todos estos años. A mi tío Jose y mis primos Nacho y Ana, por hacer que mi estancia en Madrid y en Astorga fuera tan agradable como estar en casa. A mi tía Gela por los consejos que me ha dado sobre la vida y sus ánimos para acabar la Tesis. A mi tío Popi por ser un excelente anfitrión. A mi primo Antonio, que sabe lo que es una Tesis. A mis primos Angel y Adolfo.

A mis hermanos Dani, Jose y Ruben porque son maravillosos. Su apoyo incondicional en los peores momentos es lo que me permitía seguir adelante. A Maybel porque es una hermana para mi. A Virginia por su vitalidad. A Isa por su interés. A Magí por sus ánimos y por ser tan buena persona. Finalmente, a la persona más importante del mundo para mi: mi madre Marisa. Porque ha sufrido tanto como yo

para que pudiera acabar la Tesis. Y porque has sido la única persona que siempre he tenido a mi lado durante todo el tiempo. Desde que me tuvo la primera vez en sus brazos no ha dejado de cuidarme. Sólo espero tener la oportunidad de devolverle algún día todo lo que ha hecho por mi.

Resumen

El seguimiento visual de objetos se puede expresar como un problema de estimación, donde se desea encontrar los valores que describen la trayectoria de los objetos, pero sólo se dispone de observaciones. En esta Tesis se presentan dos aproximaciones para resolver este problema en aplicaciones complejas de visión por computador. La primera aproximación se basa en la utilización de la información del contexto donde tiene lugar el seguimiento. Para mostrar como el conocimiento del entorno es una aproximación válida para resolver el problema del seguimiento visual, se presenta una aplicación de anotación de vídeo: la reconstrucción 3D de jugadas de un partido de fútbol.

A continuación, previo a la presentación de la segunda aproximación, se repasan las bases teóricas utilizadas en el resto de este trabajo. Se presenta la solución del problema de estimación desde un enfoque probabilístico que pretende encontrar los valores más probables condicionados a las observaciones obtenidas hasta ese momento. Por tanto, la solución se expresará en términos Bayesianos. Para poder completar la definición es necesaria la formalización en términos probabilísticos del modelo dinámico y la función de *likelihood*. Una vez definido de forma completa el problema, se repasa la forma de poder realizar los cálculos que permiten la obtención de la estimación de las variables por medio de la representación muestral de funciones de densidad de probabilidad.

Escogiendo el esquema Bayesiano de seguimiento visual, la segunda aproximación que se presenta en esta Tesis es un algoritmo que utiliza como observaciones directamente los valores de apariencia de los píxels de la imagen. Este algoritmo, que denominaremos *iTrack*, se basará en la construcción y ajuste de un modelo estadístico de la apariencia del objeto que se desea seguir. Este modelo permite la definición de una función de *likelihood* robusta a oclusiones parciales o totales del objeto. Los resultados del algoritmo se mostrarán en comparación con los filtros de estimación más utilizados y se muestra como la introducción de los valores de la imagen en el proceso de corrección del algoritmo mejora los resultados. Por último, se amplía la definición original del algoritmo, para poder realizar el seguimiento de múltiples objetos.

En la última parte de esta Tesis se presenta una aplicación de vídeo vigilancia automática. Este problema es difícil de resolver debido a la diversidad de escenarios y de condiciones de adquisición. Con los algoritmos basados en la información del contexto

no podría resolverse la aplicación para todos los escenarios. El objetivo es mostrar la utilidad del algoritmo *iTrack*. La aplicación se divide en tres partes principales: localización, seguimiento visual y reconocimiento. Para resolver la localización, se muestra como el modelado de escenas es la metodología más extendida y se presenta un algoritmo que resuelve el problema de trabajar con sistemas con cámara activa. A continuación, se utiliza el algoritmo *iTrack* para realizar la fase de seguimiento visual. Su adaptación al problema consistirá en la definición de una densidad *prior* a partir de los resultados del algoritmo de localización. Se propondrán un conjunto de medidas con las que evaluaremos el rendimiento del algoritmo en secuencias de vídeo vigilancia. Estas secuencias forman parte de un estándar que se está desarrollando para la evaluación de los algoritmos de seguimiento visual en aplicaciones de vídeo vigilancia. Finalmente, se propone un método de representación de actividades humanas. A partir de esta representación es posible realizar una descripción de alto nivel de lo que ocurre en la escena.

Abstract

Visual tracking can be stated as an estimation problem. The main goal is to estimate the values that describe the object trajectories, but we only have observations of their true values. In this Thesis, we present two approaches to solve the problem in complex computer vision applications. The first approach is based on using the application context information. To show how the environment knowledge is a tracking valid approach, we present a video annotation application: the 3D reconstruction of a soccer match.

Next, prior to present the second approach, we state the theoretical basis that we have used in the rest of this work. We choose a probabilistic scheme to obtain the most probable values conditioned to their observed values until this time. Therefore, the problem solution is expressed in Bayesian terms. To complete this scheme, it is necessary to express probablistically the dynamic model and the likelihood function. Once the probabilistic scheme is defined, a Particle Filter is used to make computationally feasible the method.

Based on the Bayesian probabilistic scheme, the second approach that we present is a visual tracking algorithm that uses directly the image values like observations. This algorithm, that we named *iTrack*, is based on a statistical model of the object appearance. The appearance model allows the definition of a robust likelihood function to partial or total object oclussions. The results of this algorithm are evaluated in comparison with other traditional tracking algorithms. We show like the use of image values onto the correction function outperfoms the algorithm results. Also, we extend the method to track multiple objects.

In the last part of our work, we present an automatic video surveillance system. This problem is difficult due to the possible different scenes and the environment conditions. It is not possible to solve the application for all the scenarios using the context based algorithms. Our goal is to show how the *iTrack* algorithm can be adapted easily to the system. The final application is divided onto three basic issues: localization, visual tracking and recognition. The scene models are the most used methods to obtain the object localization. We present a new method to make feasible the creation of a scene model when using an active camera. Next, we use *iTrack* to track the localized objects. With the definition of a prior density based on the results of the localization method, we can apply easily the tracking algorithm. We propose a

set of measures in order to evaluate the tracking results. The sequences used constitute a standard to tracking algorithm evaluation in video surveillance applications. Finally, we propose a human activity respresentation that can be used to make a high level scene description .

Índice General

Capítulo 1

Introducción.

1.1 Motivación.

Existen muchos tipos de sensores para recoger de forma automática la información del mundo que nos rodea. En la sociedad actual, es habitual la utilización de los lectores de códigos de barras para identificar los productos de consumo o los emisores de radio frecuencia para evitar los robos de los artículos en los comercios. Todos estos sistemas de etiquetado son baratos y funcionan correctamente. Sin embargo, requieren una cierta proximidad y proporcionan una información limitada. Existen otros sistemas de seguridad, como los circuitos cerrados de televisión (CCTV), que proporcionan más información y que cubren un mayor área de vigilancia. Estos sistemas, compuestos por varias cámaras, reciben el nombre de sistemas de vídeo vigilancia, un ejemplo de la información que pueden proporcionar es la identificación de personas.

Podemos encontrar sistemas de vídeo vigilancia en edificios oficiales, zonas portuarias, transportes y vías públicas, parques, bancos y cajas de ahorro, grandes almacenes, estadios deportivos y muchos más lugares. La gente actúa con normalidad ante la presencia de una cámara. Sin embargo, a diferencia de los dispositivos descritos anteriormente, es necesaria la presencia de profesionales que monitoricen cada cámara. Los sistemas de visión por computador aparecen como una alternativa muy útil para gestionar de forma automática esta gran cantidad de imágenes. Existen sistemas de visión capaces de localizar, seguir, reconocer e interpretar lo que está ocurriendo en la escena adquirida por la cámara. Por tanto, parece natural la necesidad de estos sistemas de visión en la sociedad actual.

1.2 Trabajos previos.

Dentro de la visión por computador, la expresión "*Looking at People*"se utiliza para agrupar todos los métodos que analizan imágenes en las que intervienen personas[70]. Este dominio cubre desde el reconocimiento de caras hasta la estimación del movimiento humano y el reconocimiento de actividades y acciones humanas. El reconocimiento

de personas y sus acciones no es una tarea nueva dentro de la visión por computador. Existen dos aproximaciones básicas: basadas en restricciones físicas y en apariencia.

La aproximación basada en restricciones físicas suele implicar la localización y el seguimiento de las articulaciones y las extremidades del cuerpo. Todo, bajo el control de un mecanismo que optimiza el seguimiento respecto a ciertas restricciones físicas impuestas por el modelo de cuerpo escogido. Esta aproximación requiere un entorno controlado debido a las dificultades para identificar las partes del cuerpo en una secuencia del mundo real. También es difícil el desarrollo de métodos computacionales eficientes por la gran cantidad de parámetros que implica un modelo físico.

La alternativa es la utilización de modelos de apariencia del movimiento humano. Las principales dificultades de estos métodos son la dependencia del punto de vista, la variación de la apariencia y el reconocimiento en presencia de oclusiones. La variabilidad de la apariencia es debida a cambios en la ropa, tamaño del cuerpo, proporciones entre individuos y sombras. A diferencia de la mayoría de métodos basados en restricciones físicas, la mayor parte de modelos de apariencia son 2D [48, 30]. Esto es lo que provoca la dependencia del punto de vista. Los sistemas basados en la apariencia suelen utilizar métodos probabilísticos para resolver estos problemas. La ventaja de estos métodos es que son capaces de aprender los modelos de apariencia a partir de secuencias de ejemplo.

Podemos dividir un sistema de reconocimiento de acciones en tres modulos diferentes. Primero, es necesario realizar la localización de las personas en la escena. Una vez eliminada la información redundante de la escena, el modulo de seguimiento visual mantiene la trayectoria a lo largo del tiempo de la persona. Finalmente, a partir de la información extraída es posible abordar el proceso de reconocimiento utilizando información contextual.

En la fase de localización, el primer paso es realizar un modelo de escena. Para este propósito es necesario conocer las condiciones de adquisición, es decir, el sistema de adquisición de imágenes utilizado para captar la escena. Algunas alternativas son cámara monocromática o color, estática o dinámica, una sóla vista o múltiples vistas, y 2D o 3D. En nuestro trabajo, consideraremos dominios complejos donde no es posible garantizar unas determinadas condiciones de adquisición.

Pfinder (*person finder*) es un trabajo clásico de localización y seguimiento de personas basado en un modelo de escena[90]. En este trabajo se utiliza un esquema probabilístico para realizar el modelo de la persona y de la escena. Cada píxel de la imagen se modela por medio de sus estadísticos de segundo orden durante un determinado número de imágenes de aprendizaje. El modelo de escena se utiliza para identificar la región de la imagen que corresponde a la persona porque los valores de los píxels no tienen el valor esperado. Estos píxels se agrupan por su proximidad espacial para formar *blobs*. El resultado final es un sistema de localización y seguimiento de una persona en entornos controlados y para cámara estática. Los problemas básicos del algoritmo son la dependencia a la postura para reconocer las partes del cuerpo y

su sensibilidad a las condiciones del entorno.

Otro trabajo basado en la apariencia es el sistema W^4 de Haritaoglu et al.[31]. W^4 es un sistema de tiempo real para localizar y seguir personas basado en cámara estática y monocromática. En W^4, las regiones de la imagen pertenecientes a las personas se detectan utilizando una combinación de modelado de escena y procesado de bajo nivel. El modelo de escena se construye a partir del máximo, mínimo y máxima derivada temporal para cada píxel de la imagen durante el tiempo de aprendizaje. Este modelo se va actualizando durante el proceso de funcionamiento del método actualizando los píxels detectados como pertenecientes a la escena. La asunción que se realiza en este trabajo es que todos los objetos localizados se clasifican como personas. El seguimiento visual se realiza por medio de modelos dinámicos simples. Para resolver los problemas que implica el seguimiento de múltiples personas se utilizan heurísticas dependientes de la aplicación. Finalmente, se utiliza un modelo 2D [48] para estimar la localización de las partes del cuerpo.

Posteriormente, se extiende W^4 a un sistema estéreo para mejorar el proceso de localización eliminando las sombras y controlando los cambios de iluminación en la escena y de apariencia de la persona[32]. La extensión de W^4 para mejorar la localización de las partes del cuerpo es *Ghost* [30]. W^4 y otros sistemas previos como *Pfinder* asumen una determinada postura para localizar las partes del cuerpo. *Ghost* utiliza un representación jerárquica de posibles posturas para estimar la posición de la persona. El resultado es que cualquier postura humana puede dividirse en postura principal y vista. El método basado en la forma de *blob* se basa en dos observaciones: es bastante probable que las partes del cuerpo estén en la silueta de la persona, y el cuerpo humano, en cualquier postura, tiene una estructura topológica que condiciona la localización de las partes del cuerpo.

Oliver y otros en [68] realizan un modelo de escena basado en el análisis de componentes principales, *EigenBackground*. La ventaja del método es que no es necesario un gran conjunto de imágenes de aprendizaje para modelar la escena. También es importante el hecho de que la presencia de objetos en movimiento no provocan que el modelo sea incorrecto.

Todos estos métodos previos de modelado de escenas están basados en cámara estática. Una posible extensión para cámara dinámica es la utilización de una representación de la escena basada en mosaicos[41]. Otra cuestión abierta, es cual es la mejor representación del cuerpo humano para poder reconocer sus movimientos[50]. La más utilizada es la representación basada en *blobs*[90, 31, 16, 21, 46, 52, 66, 72]. Otras representaciones se basan en siluetas[5, 63, 93], contornos[51] y esqueletos[29, 25]. Todas estas descripciones son 2D y basadas en la apariencia.

Una vez localizada la persona en la escena es necesario mantenerla localizada en las imágenes siguientes. Para realizar este proceso se utilizan algoritmos de seguimiento visual. El algoritmo más utilizado es el Filtro de Kalman [49]. Las características utilizadas en el seguimiento son las articulaciones o puntos característicos como el

centroide del *blob*, la mediana o la caja envolvente (*bounding box*). Sin embargo, este filtro sólo permite el seguimiento de una persona. Una escena típica contiene muchas personas y entonces es necesaria la utilización de multiples filtros y deben aplicarse técnicas de asociación de datos[10] para determinar las trayectorias de cada persona. Una propuesta alternativa es utilizar un Filtro Bayesiano[82]. Su utilización en visión por computador es reciente y fue presentada por Isard y Blake como el algoritmo CONDENSATION [42]. Este algoritmo ha sido aplicado con éxito en problemas de seguimiento visual de formas en escenas complejas[44, 43].

Los algoritmos de seguimiento visual se utilizan para corregir los errores de localización debidos a oclusiones o escenas con mucho ruido. La dificultad de estos métodos es la necesidad de definir un modelo dinámico de la persona. En la mayoría de casos se opta por utilizar modelos de velocidad o aceleración constante, esto provoca un paso previo de calibración de la escena. Sin embargo, existen métodos de seguimiento basados sólo en características visuales del objeto. Es decir, se basan en tener localizado siempre al objeto aunque esté parcialmente ocluído, y no necesitan la utilización de un filtro de estimación. Las características visuales utilizadas por los métodos basados sólo en localización son el color[74, 58], los contornos[38], y la apariencia[8].

Finalmente, destacar que una dirección de futuro es la combinación de las técnicas de seguimiento visual con las de reconocimiento de acciones. Esta combinación implica un modelo de movimiento con más restricciones para el algoritmo de seguimiento visual[9].

1.3 Aproximación al problema.

Como hemos visto anteriormente un sistema de reconocimiento de acciones se compone de tres partes diferenciadas: localización, seguimiento visual y reconocimiento. Esta división, ha provocado que estas tres partes se hayan abordado básicamente de dos formas. La primera es estudiar cada módulo por separado. Este hecho se muestra claramente en los trabajos de reconocimiento, ya que en ellos se supone que la persona está localizada correctamente en todo instante de tiempo. La segunda aproximación es desarrollar todos los módulos para resolver una determinada aplicación.

Ambas aproximaciones tienen ventajas e inconvenientes. El estudio de cada módulo por separado asume una resolución correcta de los otros módulos. En muchas aplicaciones prácticas esta suposición no es correcta. Esto puede llevar al desarrollo de métodos que no sea posible implementar en ninguna aplicación práctica debido a la presunción de asunciones incorrectas. La ventaja es que es posible centrar la investigación en el módulo que se desea resolver. Por otro lado, las aproximaciones dirigidas por la aplicación facilitan el desarrollo de los métodos al utilizar información contextual, eliminando el estudio de los casos que no ocurren en la aplicación. Sin embargo, es difícil reaprovechar los métodos desarrollados en diferentes aplicaciones. Esto implica la necesidad de estudiar cada caso particular.

El seguimiento visual es un problema clásico de la visión por computador, debido al propio interés que tiene y a la larga lista de aplicaciones que es posible realizar. En esta Tesis se aborda la resolución de este problema bajo los dos puntos de vista explicados anteriormente. Mostraremos cómo la utilización de la información contextual permite la resolución de una aplicación compleja de seguimiento visual de forma relativamente sencilla, es decir, sin la utilización de modelos dinámicos complejos.

Sin embargo, si se desea cambiar de aplicación, o ésta no dispone de información contextual se han de utilizar la técnicas de seguimiento de objetos. Estas técnicas están fundamentadas en una aproximación probabilística. Esto es debido a que el movimiento del objeto no puede modelarse de forma determinista sin errores, y a que la obtención de las medidas también suele contener errores, provocados por los dispositivos de adquisición, o por el propio entorno.

La mayoría de estas aproximaciones se basan en filtros de estimación de parámetros en presencia de ruido. Para poder realizar esta estimación se utiliza un modelo de movimiento y una función de medida. En la mayor parte de los métodos que hemos visto en la sección anterior la función de medida está basada en un proceso de extracción de características. Esto tiene el problema de que un error en el proceso de extracción de características no es tenido en cuenta en el filtro de estimación.

En esta Tesis se presenta un algoritmo de seguimiento visual basado directamente en los valores de la imagen. Para poder manejar esta función de medida, se utiliza como esqueleto básico del algoritmo un enfoque bayesiano. Se mostrará cómo poder implementar este algoritmo de forma computacionalmente eficiente por medio de una representación muestral de las densidades de probabilidad implicadas en el cálculo de la estimación.

El método presentado se ha desarrollado de forma general, sin tener en cuenta el resto de módulos de reconocimiento de acciones. Sin embargo, se abordará la resolución de aplicaciones de vídeo vigilancia automática utilizando directamente nuestro método. De esta forma, se intenta demostrar que, aunque se ha desarrollado de forma general, es adecuado para la resolución de una aplicación de reconocimiento de acciones.

1.4 Estructura de la Tesis.

Esta Tesis se divide básicamente en cuatro partes. En el capítulo 2 se muestra cómo la utilización del contexto permite la resolución de una aplicación compleja de seguimiento visual. Esta aplicación consiste en el seguimiento de jugadores de fútbol. Veremos que la definición de un método de localización y de seguimiento visual basados en las restricciones que impone el problema, permite una resolución correcta de la aplicación.

A continuación, en el tercer capítulo se presentan los métodos de seguimiento visual genéricos desde un punto de vista probabilístico. Se repasan la definición de los filtros más utilizados, y se establece un marco de trabajo para el desarrollo de nuestro método, que se presenta en el capítulo 4. Hemos denominado a nuestro algoritmo *iTrack*, de **i**mage-based **Track**ing, porque se basa fundamentalmente en los valores de apariencia de cada imagen de la secuencia. También se presenta una extensión para poder aplicar el algoritmo al seguimiento de múltiples objetos.

En el último bloque de la Tesis se presenta una aplicación de vídeo vigilancia automática. Se repasan los principales trabajos del módulo de localización de objetos y se propone un nuevo método de localización para cámara activa basado en la creación de un panorama. También se aprovecha la aplicación para mostrar cómo el algoritmo *iTrack* es posible utilizarlo de forma muy sencilla. Finalmente se muestra una descripción de actividades humanas que es posible utilizar a partir de los resultados del seguimiento visual para realizar el reconocimiento.