

MODELADO DE ENTORNOS 3D PARA PLANIFICACIÓN DE TRAYECTORIAS DE ROBOTS MEDIANTE VISIÓN ESTEREOSCÓPICA Y CONTROL ACTIVO DE PARÁMETROS ÓPTICOS

Luis M. Jiménez,⁽¹⁾ Rafael Aracil,⁽¹⁾ L. F. Peñín,⁽¹⁾ M. Ferre,⁽¹⁾ F. Torres⁽²⁾

⁽¹⁾ Div. Ingeniería de sistemas y Automática (DISAM), E.T.S.I. Industriales, Universidad Politécnica de Madrid.

⁽²⁾ Dpto. Ingeniería de Sistemas y Comunicaciones, Universidad de Alicante.

Resumen. *Se presenta un sistema de modelado de entornos tridimensionales para tareas de teleoperación, con una aplicación a la reparación en carga de torres de distribución de media tensión. Se incide en los aspectos de reconstrucción tridimensional del entorno mediante la integración de las técnicas de visión estereo y el control activo de los parámetros del cabezal estereoscópico con control de convergencia y enfoque, a fin de conseguir un sistema robusto que permita la generación de mapas visuales integrables dentro del conjunto robot-planificador de trayectorias desempeñando tareas en entornos complejos y no estructurados. El resultado de la reconstrucción actualiza una base de datos que permite implementar un control de colisiones de los brazos robotizados, realizar una planificación de trayectorias para tareas estándar, y generar información visual al operador con el objetivo alcanzar una telepresencia suficiente en las tareas de reparación.*

1 INTRODUCCIÓN

El objetivo de todo sistema de percepción es crear un modelo del mundo real y usar éste modelo para interactuar con el entorno. En la construcción de un modelo del entorno, un sistema utiliza su conocimiento sobre los objetos, conocimiento sobre los sensores y un conocimiento general del dominio en el que se desenvuelve. Las tareas de automatización avanzada suponen la actuación de mecanismos robotizados operando sobre entornos no estructurados y/o dinámicos, en estos casos el sistema precisa adquirir previamente un modelo del entorno real.

Un mecanismo primario para controlar el movimiento de un robot es la información visual. No obstante, gran parte de los sistemas desarrollados, embarcados en vehículos autónomos, perciben el entorno de forma estática. La posibilidad de fijar la mirada en un determinado motivo de la escena, móvil o no, o la de variar rápidamente la atención de una zona a otra, aptitudes inspiradas en los órganos visuales del ser humano, constituyen áreas de creciente interés en la comunidad científica [Bac-88,92][Krot-89][Coombs-92]. Un sistema sensorial visual de tales características disminuye notablemente la complejidad del tratamiento e interpretación de los datos visuales.

El desarrollo de Sistemas Robóticos Inteligentes ha sido objeto de un extenso desarrollo científico en su aplicación a la navegación autónoma de robots, pero existen otras tareas que implican la actuación de sistemas robotizados y que comparten una serie de técnicas algorítmicas similares. Este campo incluye la operación de sistemas robotizados para *Teleoperación Inteligente* que permitan la actuación humana en entornos peligrosos con una supervisión inteligente de su actividad sobre el entorno, aumentando el nivel de seguridad. Su utilización potencial es extensa y se espera que desempeñen tareas en entornos complejos, por ello, estos sistemas requieren un elevado número de aptitudes para lograr un determinado objetivo: deben percibir y modelar su entorno, planificar y razonar usando la información proporcionada por los sensores y ejecutar las correspondientes acciones. La realización de estos sistemas incluye el diseño de diferentes módulos, tanto mecánicos como computacionales, y su posterior integración en el marco de complejas arquitecturas de control.

En este artículo se describe el desarrollo de los algoritmos de reconstrucción 3D fruto del proyecto de investigación CICYT TAP92-0239 así como una implementación de estas técnicas dentro del marco del proyecto ROBTET en colaboración con las empresas IBERDROLA S.A. y COBRA S.A. para el diseño y desarrollo de un sistema

robotizado para teleoperación de trabajos en tensión en redes de distribución eléctrica. Las tareas de reparación en carga son en la actualidad muy costosas y peligrosas debido al peligro de shock eléctrico. Estas labores están siendo sustituidas por sistemas teleoperados, especialmente en aquellos países donde las normativas de seguridad son más estrictas, pero su utilización eficiente requiere la adopción de nuevas estrategias de control en lo que se puede denominar *Teleoperación Inteligente*. Estas tareas suponen el cálculo automático de trayectorias para tareas estándar, la detección de colisiones antes de que se produzcan, la realimentación de esfuerzos y datos visuales precisos, y la generación de trayectorias reales libres de colisión. Para poder realizar estas tareas el sistema precisa de un modelo exacto del entorno. Las técnicas de visión tridimensional permiten abordar estos problemas.

2 DESCRIPCIÓN DEL SISTEMA ROBTET

La figura 2 presenta un esquema de la arquitectura ROBTET. Todos los elementos, excepto los dos manipuladores esclavos y el sistema de adquisición que se encuentran en el extremo de la grúa, están situados en la cabina del camión de mando. Se pueden distinguir tres componentes principales: el puesto del operador, la estación de control de los manipuladores y el sistema de procesamiento de imágenes. El conjunto está construido sobre una arquitectura distribuida, con enlaces VME - Ethernet, según los requerimientos de comunicación de cada tarea. La comunicación entre la cabina y la cesta superior se realiza mediante un enlace de fibra óptica para mantener el aislamiento eléctrico entre zonas.

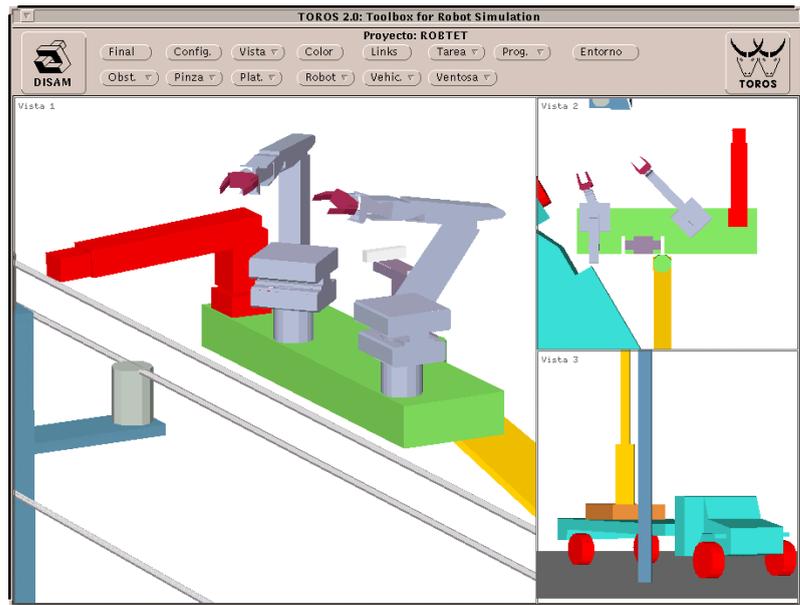


Figura 1 Simulador y planificador de tareas. El sistema para teleoperación ROBTET está constituido por una plataforma integrada por dos brazos manipuladores, una grúa de carga, un sistema sensor estereoscópico, y una cabina de control situado en la base.

El objetivo del sistema es aproximar al máximo las reacciones y condiciones de trabajo sobre la línea eléctrica a las

que percibe el operador (*telepresencia*). El puesto del operador está constituido por dos brazos maestros con realimentación de esfuerzos que controlan los dos manipuladores, un sistema de reconocimiento de voz como entrada de órdenes, y un interfase multimedia que presenta información visual del área de trabajo e información sintética procedente del modelo del entorno adquirido. Esta representación permite al operador realizar un chequeo de coherencia entre el modelo y el entorno real. El sistema permite la ejecución de tareas de reparación bajo el control del sistema de detección de colisiones. Este *módulo de control* chequea las ordenes de movimiento generados por el operador con el modelo presente en la base de datos del entorno, analizando la posibilidad de colisiones o contactos con zonas en carga, corrigiendo dichas acciones. Este módulo permite también generar trayectorias automáticas para determinadas tareas estándar. Dentro del esquema general del sistema ROBTET existe un núcleo central constituido por la base de datos que almacena el conocimiento del entorno. Este conocimiento está representado por medio de modelos tridimensionales de un conjunto de primitivas básicas y por relaciones entre ellas. Esta información es utilizada en el chequeo de colisiones y en la generación de trayectorias de acuerdo a tareas estándar, así como para la realimentación de información visual al operador. Uno de los puntos clave en la capacidad para alcanzar el objetivo de un sistema para *Teleoperación inteligente*, estriba en la fiabilidad en la correspondencia entre los datos almacenados del entorno y su estado real. Esta tarea consistente en la percepción del entorno para generar un modelo tridimensional preciso, será realizada por el *módulo de Visión Artificial*. El objetivo de dicho sistema es la reconstrucción de un modelo suficientemente descriptivo del entorno a partir de las imágenes captadas por un sensor estereoscópico. Detrás de esta tarea subyace toda la problemática del procesamiento de las imágenes captadas a fin de segmentar y extraer los detalles característicos de la escena, la reconstrucción de la información tridimensional a partir de las proyecciones bidimensionales de los sensores de las cámaras, y el reconocimiento y localización de los objetos de interés presentes en la escena.

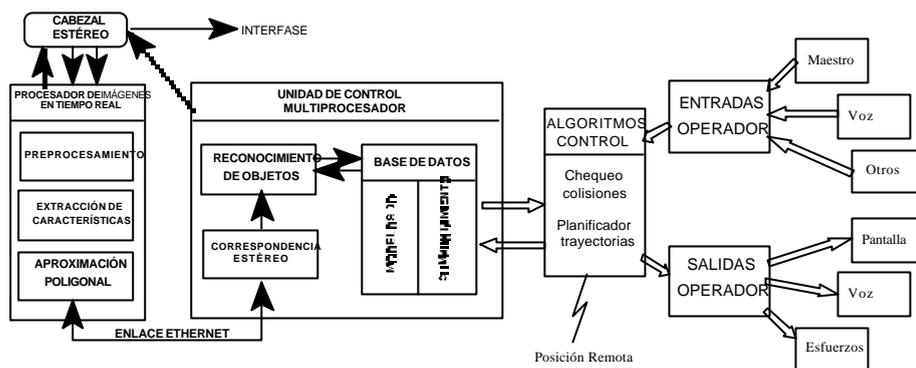


Figura 3 Arquitectura funcional del sistema ROBTET, constituido por: Base de datos con el modelo , módulo de reconstrucción 3D, chequeo de colisiones, planificador de trayectorias, interfase con el operador con realimentación de esfuerzos e información visual y de voz.

Los módulos de control e interfase multimedia se describen en detalle en [Aracil-95]. El presente artículo se centra en la descripción del módulo de visión 3D encargado de actualizar la base de datos del modelo. El tipo de entornos a percibir están limitados por el alcance del proyecto, estando este entorno constituido por una torre de alta tensión, los elementos aisladores y los cables. Se pretende por tanto segmentar estos objetos de la escena, reconocer su estructura y calcular su localización para la actualización de la base de datos. El primer elemento a localizar en la escena es la estructura envolvente de la torre que permita generar áreas de exclusión en el entorno de trabajo de los robots. No se trata pues de un elemento dinámico, o sujeto a cambios durante la reparación, por lo cual constituye una referencia fija que permite calibrar la estabilidad de la plataforma frente al área de trabajo. El objetivo en este caso es, en un paso previo a comenzar la reparación, la modelización de la torre y su envolvente, ayudada, para ciertos tipos de torres estándar, por una base de datos de modelos de torres predefinidas, y en segundo lugar un seguimiento de la posición de puntos característicos de la torre para actualizar su posición durante la operación de reparación. Los modelos a utilizar suponen una modelización poliédrica de su estructura.

El segundo elemento/os a localizar son el aislador y las fijaciones móviles del cable. Se trata en este caso de un elemento dinámico y activo, por tanto su posición puede variar debido a la labor de reparación e incluso desaparecer. El objetivo es generar un área envolvente inicial del aislador, con su localización. Debido a la complejidad motivada por la existencia de multitud de tipos de aisladores y la no estandarización de sujeciones, este área envolvente es una aproximación grosera del área ocupada, ya que en esta primera fase es el operador el que realizará la teleoperación de la reparación, no siendo por tanto un área de exclusión por colisión debido a la necesidad de que las herramientas de los robots toquen y sujeten estos elementos. En una segunda fase se planteará la localización exacta para ciertos modelos estandarizados que permita la realización de labores de teleoperación automatizadas. El tercer elemento del entorno son los cables anclados en el aislador. Se trata de un elemento dinámico y que constituye un elemento de colisión, por lo tanto se exige una actualización de su localización de forma continua.

3 SISTEMA DE VISIÓN 3D

3.1 INTRODUCCIÓN

La modelización de entornos implica, por tanto, la utilización de nuevas técnicas sensoriales que permitan extraer la información tridimensional, estas técnicas agrupadas dentro del concepto de Visión Artificial 3D tratan de resolver la pérdida de información de profundidad producida por la transformación proyectiva sobre el plano de imagen de la cámara. El desarrollo de un sistema de Visión 3D requiere la resolución de una serie de aspectos ó etapas: recuperación de la estructura tridimensional de la escena, modelado y representación de objetos, reconocimiento y localización, y la interpretación de la escena.

Las diferentes técnicas existentes para la recuperación de la estructura tridimensional de la escena presentan características específicas en todos los niveles del proceso de interpretación visual, desde la etapa de formación de la imagen, análisis e interpretación de la misma, cada método requiere tanto de equipos como de algoritmos específicos. La bibliografía desarrollada en los últimos veinte años ha permitido disponer de un conjunto de técnicas

que se podrían clasificar en cinco grupos:

% *Visión estéreo pasiva*: permite extraer la información tridimensional mediante la puesta en correspondencia de las informaciones bidimensionales procedentes de dos o más captadores de imagen.

% *Cámara móvil*: denominada en alguna bibliografía como técnicas de visión activa, permite extraer la información 3D a partir del flujo de imagen obtenido por un sensor, conocido el flujo de velocidades de la cámara.

% *Técnicas de luz estructurada*: son técnicas de visión activa en cuanto que modifican las condiciones del entorno. La distorsión producida por la proyección de patrones simples (rayos o planos), generados mediante luz coherente o luz láser, permite la extracción de la información tridimensional.

% *Telemetría láser*: permiten determinar el mapa de profundidad de la escena en base al tiempo transcurrido entre la emisión y detección de un pulso láser.

% *Control de parámetros ópticos (Análisis enfoque/desenfoque)*: permiten determinar el mapa de profundidad de la escena en a partir del nivel de enfoque en cada pixel de la escena.

Cada una de estas técnicas se ha desarrollado tratando de resolver problemas concretos de la visión 3D y presentan características específicas en todos los niveles del proceso de interpretación visual, por ello gran parte de las investigaciones actuales se dirigen hacia la integración de la información obtenida mediante diferentes sensores o técnicas de extracción, de forma que el sistema pueda trabajar con entornos complejos y no estructurados, en los cuales cada técnica por separado presenta graves dificultades tanto algorítmicas como en tiempo de cálculo y precisión cuando se trabaja con entornos no controlados.

Uno de los principios de un sistema de visión activa es permitir el procesamiento de regiones significativas para el comportamiento global del sistema. Un sistema de visión activa no puede ser considerado ya como un elemento aislado, sino como un sistema que coopera con el sistema mecánico que controla varios grados de libertad: posición en el espacio, convergencia, zoom, apertura, etc, de forma que la percepción de la escena se acomode a su propósito. Esta integración permite por ejemplo mantener el foco de atención sobre regiones de interés de la escena, la adaptación del sistema a cambios de iluminación, mantenimiento del enfoque de los objetos, etc.

3.2 VISIÓN ESTÉREO

El desarrollo de las técnicas de Visión Estéreo ha sido objeto de un gran esfuerzo de investigación en los últimos años. Nishihara y Poggio [NP-84], Aggarwal [DA-89] y Faugeras [Fau-91] presentan una excelente revisión del trabajo realizado en visión estéreo binocular hasta la fecha. El objetivo de la Visión Estéreo es resolver dos problemas [MP-79] [Faugeras-93]: el *problema de correspondencia* consistente en decidir para un punto del plano de imagen izquierdo, que punto en el plano de imagen derecho es correspondiente (son imágenes del mismo punto físico). El segundo problema es el *problema de reconstrucción* que tratar de obtener, dados dos puntos correspondientes en ambos planos de imagen, las coordenadas 3D del punto en el espacio respecto a un sistema de coordenadas del mundo. Dentro de este segundo problema subyace el problema de calibración del sistema visual que determina con qué aproximación conocemos las posiciones de los centros focales de cada sistema óptico, y la posición de los planos de imagen en el sistema de coordenadas del mundo. La solución del problema de la visión estéreo debe asumir desde el primer momento que las posiciones de los puntos en la imagen y los parámetros de calibración del sistema visual son conocidos de forma imperfecta. Por lo tanto no se puede confiar en modelos geométricos rígidos del proceso de formación y reconstrucción de la imagen.

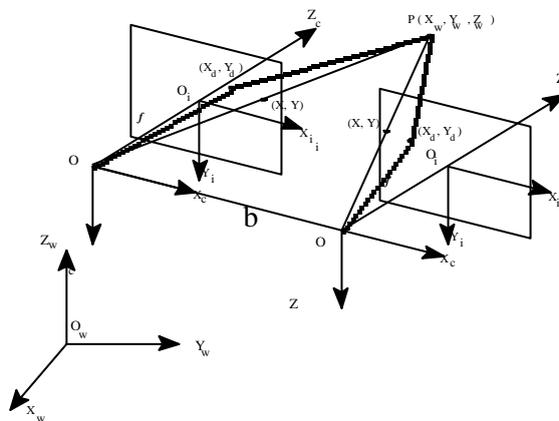


Figura 4. Modelo par estereoscópico alineado con distorsión.

El Problema de Correspondencia.

El problema de correspondencia tal como se ha definido es en principio un problema indeterminado: dado un punto

m_1 en el plano de imagen R_1 puede ser el correspondiente de cualquier punto m_2 en el plano de imagen R_2 . Para resolver esta dificultad se deben identificar un conjunto de restricciones que permitan reducir el número de correspondencias potenciales. Estas restricciones se pueden clasificar en cuatro tipos básicos [Fau-93][MT-94]:

& Restricciones geométricas impuestas por el sistema de formación de imagen.

La más importante es la condición de *epipolaridad* por la cual los puntos correspondientes deben estar contenidos en la línea epipolar correspondiente [Marr-82]. Esta restricción permite reducir el problema de una búsqueda bidimensional a uno unidimensional. Otras restricciones son las impuestas por la convergencia de los ejes focales de las cámaras que impone la condición de disparidad nula en el centro de la imagen [Coomb-92].

& Restricciones geométricas precedentes de los objetos visualizados.

Se han identificado un conjunto amplio de condiciones de este tipo que permiten de forma general resolver la ambigüedad del problema estéreo. La aplicación de estas condiciones depende en gran medida del tipo de escenas analizadas:

- *Condición de unicidad*: si no existen superficies transparentes la correspondencia para un punto m_1 en el plano de imagen R_1 debe ser única en el plano de imagen R_2 [MP-79].
- *Condición de continuidad de superficie*: La distancia al sistema de imagen varía suavemente dentro de las superficies, por lo tanto las disparidades deben variar también suavemente [MP-79].
- *Condición del gradiente de disparidad*: restringe el máximo gradiente de disparidad permitido entre correspondencias de primitivas [Jul-80].
- *Condición de continuidad figural*: la disparidad varía continuamente a lo largo de los contornos de las superficies, pero puede cambiar abruptamente entre contornos [MF-81][MN-89][Otha-85].
- *Condición de orden*: la correspondencia de primitivas debe ocurrir en el mismo orden izquierda-derecha a lo largo de las líneas epipolares [Baker-81] [Otha-85]. La inversión de orden puede ocurrir en ciertos entornos (p.e. objetos transparentes, objetos estrechos, etc.) pero son escasas en escenas naturales.
- *Condición de posición general*: ciertos eventos ocurren bastante infrecuentemente, en un sentido estadístico, de forma que permite desestimar correspondencias improbables [MT-94].

& Restricciones físicas.

Procedentes de modelos de como los objetos interactúan con la iluminación. La luz reflejada por una superficie es función del tipo de superficie, su orientación, la posición de los focos de luz y el punto de vista [Horn-86]. Parten de un modelo de reflectancia de las superficies (generalmente un modelo Lambertiano) y un modelo de iluminación (difusa). El problema surge por el hecho de que las superficies reales no son Lambertianas y la iluminación nunca es perfectamente difusa.

& Restricciones locales de la primitiva.

Estas condiciones determinan el grado de probabilidad de correspondencias de primitivas seleccionadas de la escena, como pueden ser signo y orientación del gradiente de intensidad, o condiciones asociadas a primitivas de mayor nivel como líneas o regiones que atesoran un conjunto más rico de características.

Técnicas de Correspondencia.

Los algoritmos de reconstrucción estéreo dependen en gran medida del tipo de primitivas candidatos de la imagen. Medioni y Nevatia [MN-85] distinguen entre dos tipos de técnicas estéreo según la primitiva utilizada: las basadas en regiones, y las basadas en características dispersas de la escena.

Técnicas basadas en regiones.

Idealmente, el método más simple sería tratar de buscar correspondencias para cada pixel de la imagen, pero por un lado la información semántica almacenada en un pixel es muy escasa (nivel de intensidad) y por otro es el resultado de una discretización de un fenómeno continuo como es la proyección de la luminosidad reflejada por las superficies de la escena. En esta discretización subyace un error de precisión y posible enmascaramiento de proyecciones próximas. Por lo tanto para poder atacar el problema de correspondencia las técnicas basadas en regiones utilizan la información en un área o entono de vecindad de cada pixel. La obtención de una correspondencia se realizaría mediante la búsqueda en la otra imagen mediante una medida de correlación cruzada suma de las diferencias absolutas de intensidad u otras medidas similares [Nev-82]. Este tipo de técnicas hacen uso de simplificaciones del espacio de búsqueda mediante la condición de epipolaridad y la utilización de estrategias multinivel «*tosco-a-fino*» que se analizarán más adelante. Deriche y Cocquerez [DC-87] utilizan regiones, contornos cerrados sobre la imagen de intensidad luminosa. Estas regiones se segmentan mediante alguna característica de textura, tamaño, forma, etc. Estas técnicas plantean problemas de correspondencia de los puntos dentro de las regiones, pero facilitan la

correspondencia entre regiones al ser entidades de mayor nivel, con más características diferenciadoras.

Los sistemas basados en correlación de regiones sufren de las siguientes limitaciones:

- & Requieren la presencia de texturas dentro de cada ventana de correlación, por lo tanto tienden a comportarse mal ante entornos con falta de características o por la presencia de texturas repetitivas.
- & Tienden a confundirse por la presencia de discontinuidades en superficies dentro de la ventana de correlación.
- & Son sensibles a intensidades absolutas, contraste e iluminación.
- & Tienen problemas con campos de profundidad que cambian rápidamente.

Técnicas basadas en características.

Estas técnicas se basan en la extracción de un conjunto de primitivas características de cada imagen para realizar la correspondencia entre estos datos a partir de las propiedades diferenciadoras. Las primitivas pueden ser pixels que representen características especiales de la escena como son los pixels de bordes, correspondientes a cambios bruscos de intensidad [Marr-82] o los pasos por cero del operador Laplaciana. Este tipo de datos presentan unas características estables frente a variaciones fotométricas. Como características se utilizan el contraste del gradiente, orientación, radiancia [Horn-86].

Un paso adicional consiste en la agrupación de pixels en primitivas de mayor nivel de abstracción. El caso más simple es la línea recta que es invariante a la proyección de perspectiva, esta representación es compacta, intrínseca, discriminable, precisa y densa. Esta representación permite modelar fácilmente modelos poliédricos, pero tiene problemas con contornos curvos, para los cuales se debe utilizar aproximaciones polinomiales de orden superior [ShePel-90] para preservar la compacidad de la descripción y discriminar mejor la característica. Aún así pueden presentar problemas adicionales debido a oclusiones parciales entre primitivas. Los segmentos de línea y curvas disponen de características geométricas y de intensidad incorporadas como longitud, orientación, curvatura, contraste medio, etc. Entre las ventajas de los sistemas basados en primitivas se encuentran:

- % Son más rápidos que los métodos basados en área, ya que se reduce el número de primitivas a considerar.
- % La correspondencia es más precisa ya que la primitiva puede ser localizada con precisión subpixel.
- % Son menos sensibles a variaciones fotométricas, ya que representan propiedades de la escena.

Por otra parte presentan el inconveniente de utilizar datos dispersos de la escena, por lo cual para obtener un mapa denso de profundidad de la escena será necesario interpolar los datos obtenidos en primitivas de superficie.

Selección y Representación de Primitivas.

Respecto a la selección de las primitivas de la escena se pueden establecer un conjunto de propiedades que permitan evaluar su bondad e idoneidad para resolver el problema de correspondencia. Para ello se evalúan dos características básicas: la potencia descriptiva para resolver ambigüedades, y la invarianza ante pequeñas variaciones en el punto de vista. Para evaluar estas propiedades se puede identificar un conjunto de estimadores [Barnard-82][MT-94]:

- **Dimensionalidad:** Los puntos se ven menos afectados por distorsiones de perspectiva, pero las características tipo borde permiten encontrar correspondencias más fácilmente.
- **Tamaño:** (frecuencia espacial) Para primitivas prominentes es más fácil encontrar la correspondencia.
- **Contraste:** primitivas de bajo contraste son difíciles de distinguir.
- **Contenido semántico:** que permita resolver falsas correspondencias.
- **Densidad de ocurrencia:** (disperso frente a denso). Primitivas dispersas son más fáciles de corresponder pero generan sólo disparidades dispersas.
- **Localización y estabilidad** del detector de primitivas. El detector debe localizar la primitiva en la posición real correspondiente a la proyección de la primitiva tridimensional, y esta debe ser estable entre las diferentes imágenes.
- **Distinguibilidad/Unicidad.** La primitiva debe ser distinguible en ambas imágenes y cada característica de la escena debe generar una única primitiva en cada imagen.

Cabe destacar los experimentos biológicos de Bel Julesz [Jul-60,71] sobre estereogramas aleatorios que indican que la estereovisión humana actúa desde el inicio del proceso visual en lugar de seguir un análisis detallado de cada imagen monocular. Esto indica que son las primitivas de bajo nivel las que determinan el inicio del proceso estereoscópico. Ejemplos de utilización de primitivas incluyen la utilización de intensidades de pixel [Gennery-77][Movarec-83], signo del filtro Laplaciana-Gausiana [Nishihara-84], zero-crossings de la Laplaciana-Gaussian [Marr-82].

y Poggio [MP-79] [Grimson-85][KA-87], elementos de borde Baker y Bindford [Baker-81], segmentos enlazados de borde [Ayache-87,91][Faugeras-91], Horaud y Skordas [HoSk-84] Medioni y Neviata [MN-85] Sherman y Peleg [ShePel-90]. Además de obtener las características (contornos o regiones), es necesario utilizar una representación estructurada de esta información.

Estrategias de Correspondencia.

Independientemente de la técnica de correspondencia utilizada se han desarrollado un conjunto de estrategias cuyo objetivo es reducir la ambigüedad y mejorar la correspondencia estéreo. Entre estos métodos se pueden citar:

% **Estrategias tosco-a-fino.** Son estrategias multinivel que utiliza la información obtenida en una escala tosca que se utiliza para guiar y limitar la búsqueda de correspondencia de las primitivas en la escala fina. En esta aproximación la correspondencia se inicia en la escala grosera donde el número de primitivas es muy inferior, lo cual disminuye el espacio de búsqueda, pero su precisión es muy pequeña [MP-79][Nish-84][Grim-85].

% **Estrategias estructurales.** utilizan relaciones estructurales entre las primitivas, dependen en general del tipo de escena analizar [Boyer-88][PonHar-89].

% **Estrategias jerárquicas.** Estos métodos utilizan diferentes niveles de descripción de la imagen formados por primitivas de diferente nivel de abstracción. En este caso la reducción de la complejidad de la correspondencia se logra mediante la utilización de estructuras de más alto nivel que sirven para decidir correspondencias en un nivel inferior [Binford-87][MT-94].

% **Integración,** de las diferentes fases de extracción de primitivas, correspondencia e interpolación de superficies [Ahuja-93].

% **Detección y corrección de errores en las disparidades.** Medioni y Neviata [MN-89].

% **Estrategias híbridas regiones-características.** Cochran y Medioni [CocMed-89].

% **Estrategias iterativas de predicción-propagación.** Estos métodos realizan una primera predicción de correspondencias sobre algunos buenos candidatos y tratan de propagarlas a sus vecinos mediante grafos de disparidad. Ayache y Faugeras [Aya-87][Fau-91].

Una parte importante de los algoritmos de correspondencia suele ser la técnica de "*relajación*", que permite reducir el número de correspondencias hipotéticas, disminuyendo el cálculo computacional necesario. Marr y Poggio son los pioneros en el uso de algoritmos de relajación para visión estéreo [MP-76] [Marr-82]. Este algoritmo se inicializa con todas las correspondencias que satisfagan la condición epipolar. Se asocia un peso a cada correspondencia (modelo neuronal) y se utiliza un mecanismo iterativo: en cada iteración y para cada correspondencia, un proceso "INHIBITORIO" reduce los pesos de correspondencias conflictivas (aquellas que violan la condición de unicidad con la correspondencia actual) y un proceso "EXCITATORIO" incrementa los pesos de las correspondencias con disparidades similares (condición de continuidad). Las correspondencias con pesos insuficientes son eliminadas. El algoritmo converge a soluciones en pocas iteraciones. Los mejores resultados se obtienen con estereogramas de puntos aleatorios. Julesz [Jul-71] demuestra que en el sistema visual humano el proceso de correspondencia funciona en primitivas elementales (puntos) en ausencia de cualquier reconocimiento de formas. Estos algoritmos suelen tener dificultades en presencia de discontinuidades, por ello se han presentado algoritmos mejorados que añaden una condición adicional (preservación de orden) Drumheller y Poggio [DP-86]. Otras implementaciones como la de Grimson [Grim-81] utilizan puntos de contorno obtenidos por zero-crossings y una técnica multiresolución: las correspondencias de baja resolución se utilizan para resolver correspondencias conflictivas en una resolución superior. Mayhew y Frisby [MF-81] sustituyen la condición de continuidad por un test estadístico que llaman "*la condición figural*", y la coincidencia entre canales de diferente frecuencia espacial.

3.3 CONTROL DE CONVERGENCIA

En la solución del problema de reconstrucción 3D mediante visión estéreo binocular, muchos autores asumen que los ejes ópticos de las cámaras son paralelos y están en un plano ortogonal a los planos de imagen. Esta simplificación reduce la complejidad en la búsqueda de la correspondencia al convertir la restricción epipolar a una línea horizontal, y la reconstrucción 3D al establecer una sencilla ecuación entre la disparidad horizontal y la profundidad. A pesar de ello, esta simplificación es difícil de lograr en la práctica en un sistema mecánico real. Generalmente se introduce la técnica de rectificación para re proyectar las imágenes originales sobre un plano. Esta re proyección plantea el problema de cuantificación de las imágenes y no supone una disminución del coste computacional ya que este se traslada al proceso de rectificación.

Por otra parte, para la solución de la correspondencia estéreo es deseable hacer que los ejes ópticos converjan de

forma que la superposición de los campos visuales sea mayor y la disparidad disminuya. Esto coincide con los estudios psicofísicos sobre la visión binocular humana en el cual la fusión estereó solo se produce dentro de un rango de disparidad máximo (límite de disparidad de Panum [Marr-82]). Adicionalmente, cuando los planos de imagen son verticales la fusión estereó presenta problemas para los bordes horizontales de la escena (bastante comunes en entornos artificiales), por lo cual es deseable que las líneas epipolares no sean paralelas a las líneas de rastreo del sensor. Por tanto es preferible incorporar la geometría convergente dentro del problema estereó pudiendo lograr la reconstrucción 3D mediante una adecuada calibración del sistema binocular.

El objetivo de un sistema de convergencia es controlar la distancia desde las cámaras al punto de fijación en una orientación determinada, manteniéndolo próximo a la superficie del objeto. El punto de fijación viene dado por la intersección de los dos ejes focales que formarán un ángulo α de convergencia. Debido a que este ángulo está relacionado con la distancia, el sistema de convergencia puede dar una estimación de distancia del objeto principal de la escena (situado en el centro de los sensores) de manera inversa, una estimación previa de la distancia puede ayudar a mantener el punto de fijación en la superficie del objeto. Esta relación permite el planteamiento de un algoritmo cooperativo junto a otra técnica de extracción de profundidad como la estereoscopia o el enfoque. Abbott y Ahuja [Ahuja-89,93] utilizan la convergencia cooperativamente con el enfoque y la estereoscopia extendiendo un mapa parcial en cada fijación, Krotkov [Krot-89] mejora un mapa completo aproximado mediante el control de convergencia y enfoque. De este modo la convergencia permite al sistema concentrarse ("foveate") en áreas de interés para obtener una mayor precisión y confianza en la medida de profundidad, presentando una serie de ventajas:

- **Simplificación matemática:** los puntos del objeto de interés están próximos al eje óptico de ambas cámaras, permitiendo simplificaciones en el modelo geométrico del par estereoscópico (perspectiva débil).
- **Facilita la correspondencia estereó:** el punto de fijación tiene disparidad nula, y las zonas vecinas tendrán disparidades pequeñas. Esto disminuye el número de candidatos en la correspondencia y por tanto disminuye la posibilidad de falsas correspondencias, especialmente cuando existen patrones repetitivos en la escena [Nishijara-84][Olson-91]. Los estudios de Julesz [Julesz-71] sobre la visión binocular humana corroboran esta apreciación, ya que la fusión estereó solo se produce dentro de un rango máximo de disparidades (límite de fusión de Panum).
- **Posibilitan la utilización de un sistema de coordenadas ligado al objeto** [Bal-89].

3.3.1 Estrategias para el control de convergencia.

El proceso de convergencia se puede dividir en dos tareas: el mantenimiento de la convergencia sobre un objeto de la escena y la fijación de un nuevo objeto. En el primer caso presenta un comportamiento determinado por las dinámicas del observador y el objeto. Estas están restringidas en escenas reales al ser las velocidades y aceleraciones finitas y las superficies de los objetos varían suavemente en profundidad. En cambio en la fijación de un nuevo objeto se producen cambios bruscos en el ángulo de convergencia y discontinuidades en su derivada temporal. Por tanto deben usarse dos algoritmos de control, uno optimizado para cambios suaves y continuos para el mantenimiento y saltos bruscos para la fijación de nuevos objetos.

Básicamente se distinguen dos fases: la primera consiste en seleccionar el objetivo que puede realizarse mediante un filtrado para detectar los contornos de las superficies, o bien tomando una cámara como dominante. La segunda fase consiste en variar el ángulo de convergencia en función de un estimador de error hasta que éste se anule. Para poder decidir el ángulo de convergencia para poder alcanzar el punto de fijación es necesario una estimación del error de convergencia. El objetivo es calcular este error a partir de la información visual y ejecutar una acción de control que lo haga nulo, es decir, lograr que la intersección de los dos ejes focales esté situada sobre la superficie del objeto. En la literatura se pueden encontrar diferentes técnicas para dar una estimación de este error [Coombs-92][Ahuja-93]:

Í Una primera aproximación es utilizar una medida de profundidad obtenida a partir de la **disparidad binocular** entre los centros focales de las imágenes pero requiere resolver el problema estereó previamente, lo cual es difícil y costoso si la disparidad es elevada. La ventaja es que la relación entre el ángulo de convergencia y la disparidad es simple y directa, e incluso no es necesario utilizar valores absolutos de ángulo de convergencia. El inconveniente es que los estimadores de disparidad estereó suelen estar optimizados para obtener una buena localización y densidad de características pero no para ser robusto y capaz de ejecutarse en tiempo real al necesitar una etapa de optimización global.

Í Otro tipo de técnicas son las denominadas "registration" que operan sobre el espacio de la imagen trasladando la imagen hasta que las regiones solapadas son similares. **La correlación cruzada** entre ventanas de ambas imágenes

es utilizada como medida de disparidad. La Ecuación 3.1 nos da una medida de correlación entre el centro de la imagen izquierda y un punto de la imagen derecha desplazado (s,t) . La medida de correlación se realiza en una ventana entorno al centro de la imagen con una máscara de ponderación w_v que da más peso a los valores centrales.

$$d^2(s,t) = \frac{\left[\int w_v I_L(x,y) I_R(x+s,y+t) dx dy \right]^2}{\left[\int w_v I_L^2(x,y) dx dy \right] \left[\int w_v I_R^2(x+s,y+t) dx dy \right]} \quad (\text{Ecu. 3.1})$$

Este tipo de medidas se utilizan también en correspondencia estéreo cuando las variaciones de intensidad son relativamente continuas y suaves (generalmente esto es cierto en imágenes aéreas)[Ahuja-89]. estas medidas tienen problemas cuando el gradiente de las superficies es elevado o cuando hay oclusión del punto de fijación en cada una de las imágenes. El fenómeno de oclusión está ligado con el límite del gradiente de disparidad [Mayhew-83][Fau-91], problema que subyace también en la visión estereoscópica. Sólo el enfoque, al ser monocular, es robusto frente a estas situaciones, razón por la cual es conveniente incorporarlo en la solución del problema estéreo.

El **La transformada de Fourier** resulta útil para realizar esta estimación de error debido a que permite detectar el eco espacial entre las dos imágenes. Las dos imágenes convergentes serán similares y con un cierto desplazamiento que se traducirá en un eco en el espectro de frecuencias. el problema radica el coste computacional de la transformada de Fourier. Una de las implementaciones que utilizan esta transformada es la del filtro **Cepestal** (Yeshurum y Schwartz) [YS-89]. El "cepestrum" de una señal es la transformada de Fourier del logaritmo del espectro de potencia de la señal. El espectro de potencia es así mismo la transformada de Fourier de la función de autocorrelación de la señal $*F^*$). Fue inicialmente desarrollado por Bogert [Bogert-63] para analizar señales con eco. Sea $S(t)$ una señal unidimensional y $R(t)$ la señal con ecos en intervalos t_0, t_1, \dots $R(t)$ será la convolución de la señal $S(t)$ con un tren de impulsos.

$$R(t) = S(t) * (\delta(t) + a_0 \delta(t-t_0) + a_1 \delta(t-t_1) + \dots) \quad (\text{Ecu. 3.2})$$

Cuando se toma el logaritmo del espectro de potencia se transforma en dos términos, uno que depende solo de $S(t)$ y el otro que es una combinación de sinusoides distorsionadas con frecuencias relacionadas con t_0, t_1, \dots Yeshurum y Schwartz [YS-89] han aplicado el filtro cepestral bidimensional como estimador de disparidad. Toma ventanas $(h \times w)$ en cada imagen y las une lado a lado formando un imagen $(h \times 2w)$. Considerando que solo existe desplazamiento se puede considerar como la imagen original en $(0,0)$ más un eco en $(w+d_h, d_v)$, donde d_h, d_v son las disparidades horizontal y vertical. El término periódico del logaritmo del espectro de potencia tendrá frecuencias fundamentales de $(w+d_h)$ en horizontal y d_v (señales de alta frecuencia). La componente dependiente de la imagen estará compuesta de bajas frecuencias. por tanto el filtro cepestral de la señal presentará picos aislados en $(\pm(w+d_h), \pm d_v)$. Coombs [Coombs-92] realiza una implementación con imágenes muestreadas. El estimador funcionaba bien dentro de un rango de $\pm 7^\circ$ de disparidad.

El **Correlación de fase**. Coombs [Coombs-92] presenta un estimador derivado del filtro cepestral caracterizado por un comportamiento similar pero con un coste computacional menor. Este estimador está relacionado con la correlación de fase y se basa en un estudio del efecto producido por el filtro cepestral. El espectro de potencia, al ser la transformada de Fourier de la función de autocorrelación, es real y par, y por tanto la transformada de Fourier y su inversa son equivalentes. Aplicando esto tenemos que el filtro cepestral equivale a la transformada inversa de Fourier del logaritmo de la transformada de Fourier de la función de autocorrelación. El término logarítmico es lo que lo diferencia de la autocorrelación.

$$\hat{C}^{-1} [\log^* \hat{C}^{*2}] = \hat{C}^{-1} \left[\frac{\sqrt{\log^* \hat{C}^{*2}}}{* \hat{C} *} \right]^2 \quad (\text{Ecu. 3.3})$$

El segundo término de la ecuación es la transformada de Fourier inversa del espectro de potencia de la función original filtrada con un filtro no lineal. Este filtro reduce la contribución de señales de banda estrecha producida por patrones periódicos y objetos suaves y grandes que tienen pobres características para correlar, y deja las señales de banda ancha casi inalteradas. A partir de esta aproximación se derivan los filtros de correlación de fase, que consistiría en llevar el filtro al extremo de que el resultado sea una constante. la transformada de Fourier es un impulso en $(0,0)$ (la imagen se correla así misma con disparidad cero). En este caso se debe utilizar la información de

fase que se pierde en el espectro de potencia. (Kuglin y Hines [KH-75]) . Person [Person-77] describe una implementación de este algoritmo. Este método requiere menor coste computacional al procesar ventanas de dimensión $(h \times w)$ y no $(h \times 2w)$ del filtro cepestral.

3.4 PROFUNDIDAD A PARTIR DEL ENFOQUE

Una tarea fundamental dentro de un sistema de visión activa es mantener enfocados los objetos de interés dentro de la superficie sensora. En fisiología se denomina "Acomodación" a la acción de las lentes oculares para enfocar el ojo. Este fenómeno de acomodación, mantiene enfocados los objetos de primer plano ó interés con el movimiento de la cabeza y constituye en sí mismo un fuente de información de la distancia de los objetos. Las preguntas que trata de resolver un sistema de enfoque en visión activa son dos. Dada la proyección P' de un punto $P=(X,Y,Z)^T$ (Z desconocido) perteneciente a la superficie de un objeto de la escena:

- 1.- Sin información acerca de Z , ¿cómo se puede determinar la posición del motor de enfoque para lograr la mejor definición de P' ?
- 2.- Dada la posición del motor de enfoque que da la proyección P' más contrastada, ¿cómo se puede calcular la componente Z de P ?

Este proceso se denomina en la literatura "profundidad a partir del enfoque" y ha sido tratado por varios investigadores. Krotkov [Krotkov-89] discute varios criterios para determinar el mejor enfoque de la imagen de la escena y el cálculo de la profundidad a partir del motor de enfoque. Pentland [Pent-87] y Grossmann [Gros-87] ha tratado el problema desde un punto de vista diferente , considerando la extracción de profundidad a partir del desenfoque, dando una estimación de la profundidad de la escena a partir de la borrosidad de cada zona de la imagen

3.4.1 Análisis del Desenfoque.

La figura 3.4 presenta un modelo general de la geometría de una lente física. Al ser el mismo medio (aire) el existente entre el sensor y las lentes y las lentes y la escena, las distancias focales frontal y posterior son iguales (f). La ley de las lentes Gaussianas relaciona d_{ext} (distancia desde el plano principal primario al objeto), y d_{int} (distancia desde el plano principal secundario a la imagen de dicho objeto) con la distancia focal f mediante la ecuación 3.4.

$$\frac{1}{d_{ext}} + \frac{1}{d_{int}} = \frac{1}{f} \quad (\text{Ec. 3.4})$$

Cuando un objeto se encuentra a una distancia d_{ext} (distancia del objeto) del punto principal primario de las lentes, su imagen aparece a una distancia d_{int} (distancia de la imagen) detrás del punto principal secundario (lentes convergentes). d_{int} y d_{ext} son consideradas distancias conjugadas. La superficie sensora (situada en el plano normal al eje óptico) se encontrará a una distancia d_{sen} desde el plano principal secundario. A la distancia conjugada del plano sensor enfrente de la lente (distancia de enfoque) los puntos estarán perfectamente enfocados en la imagen del sensor. Para un punto fuente en el infinito $d_{ext}=\infty$, de la ecuación 3.4 tenemos $d_{int}=f$.

Existen dos formas de abordar el proceso de formación de imagen: *Óptica Geométrica*, se basa en el trazado de rayos y sus resultados son una aproximación de primer orden. *Óptica Física*, se basa en la teoría de difracción y sus resultados son exactos. En el presente estudio se utiliza la primera técnica, alternativa utilizada en la mayoría de los trabajos que abordan este problema, debido al límite de precisión del resto de los elementos de proceso de formación de imagen. Si las lentes tienen un diámetro A (apertura de la lente) y el plano sensor está desplazado $\Delta = d_{int} - d_{sen}$ entonces la imagen formada sobre él será un círculo de diámetro d :

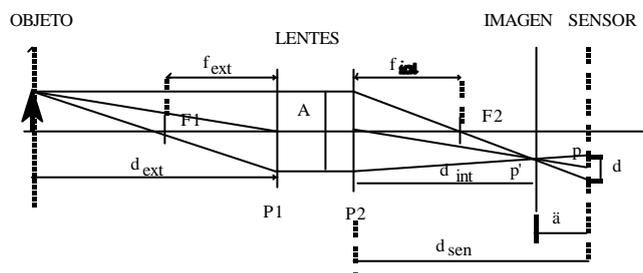


Figura 3.5. Modelo de lente física. P_1 , P_2 son los planos principales primario y secundario. F_1 , F_2 son los planos focales, primario y secundario. A: apertura de las lentes, d: diámetro del círculo de confusión.

$$d \approx A \left(\frac{1}{f} + \frac{1}{d_{ext}} \right) \quad (\text{Ec. 3.5})$$

Las lentes reales cortan ciertas altas frecuencias espaciales a causa de la difracción y también atenúan otras a causa de aberraciones. Esto supone que la intensidad en el plano sensor no es uniforme. Suponiendo que hasta 1^{er} orden la intensidad es uniforme, el círculo de confusión puede ser modelado matemáticamente mediante la convolución con la función de distribución de intensidad descrita en la ecuación 3.6. La transformada de Fourier de la función círculo se puede expresar con la función de Bessel de primer orden [HZ-74].

La transformada es simétrica circularmente con un punto central y anillos concéntricos de amplitud decreciente. Actúa como un filtro paso-bajo. De la función de desenfoque se pueden extraer las siguientes conclusiones:

$$I_b(x,y) = I_g(x,y) \text{ circ}_r(s), \quad I_g(x,y) \text{ Imagen ideal} \\ \text{circ}_r(s) = \frac{4}{\delta d^2}, \quad s = \sqrt{x^2 + y^2} \quad 0 < s < d/2 \quad (\text{Ec. 3.6})$$

! Cuando r aumenta, la integral del espectro de potencia decrece y el primer cero de $J_1(rf_a)$ se aproxima al origen (se concentra más la energía en la parte central).

! Cuando δ crece, r crece, por consiguiente más se atenúan las altas frecuencia espaciales y más borrosa será la imagen.

$$f_a > 2\delta r^2 \left[\frac{J_1(rf_a)}{rf_a} \right], \quad f_a = \sqrt{f_x^2 + f_y^2} \quad (\text{Ecu. 7})$$

J_1 es Función de Bessel de 1^a clase de primer orden

Portanto desenfocar un sistema óptico atenúa las altas frecuencias espaciales. Esta característica se puede utilizar para medir el desenfoque de la imagen.

3.4.2 Funciones de Medida de la calidad del enfoque.

Se debe definir una métrica de agudeza visual o enfoque. La idea general es utilizar el contenido de altas frecuencias. Existen diferentes métodos que permiten medir el desenfoque, a continuación se presenta un resumen de estos métodos con sus características principales [Krotkov-89].

1.- **Sensores de rango**, sonares, sensores de infrarrojos. Son sensores activos que actúan sobre el entorno midiendo la distancia a los objetos de la escena. Este tipo de dispositivos se utilizan en los sistemas de enfoque de cámaras comerciales y realizan el proceso inverso calculando la distancia para a partir de ella calcular la posición del motor de enfoque.

2.- Transformada de Fourier.

En el apartado anterior vimos como el enfoque está relacionado directamente con la información de altas frecuencias. El cálculo del enfoque se realizaría maximizando la energía de altas frecuencias. Este método supone un costo computacional muy elevado debido a la necesidad de calcular la transformada de Fourier de la imagen.

$$S(x,y) = *LI(x,y)* = \sqrt{I_x^2 + I_y^2} \\ i_x = \frac{1}{4} \begin{bmatrix} 1 & 0 & 1 \\ 2 & 0 & 2 \\ 1 & 0 & 1 \end{bmatrix} \quad i_y = \frac{1}{4} \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \quad (\text{Sobel})$$

3.- Maximización de la Magnitud de Gradiente.

Ya que la calidad del enfoque afecta a las características de borde, es natural utilizar un detector de borde para su medida. Tenenbaum [Ten-70] y Schlag [Schl-83] describen su utilización. (es descrito por estos autores como operador Tenengrad).

Realiza la estimación del gradiente mediante el módulo del operador de Sobel, sumando todas las magnitudes superiores a un cierto umbral dentro de un entorno del punto a enfocar.

$$I_x = i_x(I(x,y)) \quad I_y = i_y(I(x,y)) \\ \max_j \sum_x \sum_y S(x,y) \quad \text{para } S(x,y) > T \quad (\text{Ec. 3.8})$$

$$L(x,y) = \nabla^2 I(x,y) = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2}$$

$$L = \frac{1}{6} \begin{bmatrix} 1 & 4 & 1 \\ 4 & -20 & 4 \\ 1 & 4 & 1 \end{bmatrix} \quad (\text{laplaciana})$$

$$\max_{j_x, j_y} \sum_{j_x, j_y} *L(x,y)* \quad \text{para } L(x,y) > T \quad \max_{j_x, j_y} \sum_{j_x, j_y} I(x,y) & T \quad \text{para } I(x,y) > T \quad (\text{Ec. 3.9})$$

los cuadrados de las intensidades está relacionada mediante el Teorema de Parseval con la energía del espectro de potencia.

6.- Entropía del Histograma.

En una imagen desenfocada el histograma tiende hacia una distribución uniforme. La entropía es una medida de la uniformidad de un histograma. Cuanto mayor sea E , más uniforme es el histograma y más desenfocada estará la imagen. Este método falla con imágenes con texturas complicadas.

$$E = - \sum_I P(I) \ln(P(I)) \quad P(I) \dots 0 \quad (\text{Ec. 3.11})$$

$P(I)$ es la frecuencia de la ocurrencia del nivel de gris I .

7.- Histograma de Variaciones Locales.

A mayor variación local del nivel de gris, tendremos un mejor enfoque. El método consistiría en:

- 1) Calcular un histograma de las diferencias locales de nivel de gris en una dirección dada.
- 2) Comprimir el histograma con una transformación logarítmica.
- 3) Ajustar el histograma comprimido a una línea recta L ,
- 4) Recuperar la inclinación de L . Cuanto menor es esta inclinación tendremos mayores variaciones, por tanto buscaremos el mínimo de esta inclinación.

$$\sigma^2 = \frac{1}{N^2} \sum_{x=1}^N \sum_{y=1}^N (I(x,y) - \mu)^2$$

μ media de la distribución de nivel de gris

(Ec. 3.12)

9.- Diferencia Modulo-Suma.

Es una medida propuesta por Jarvis [Jar-76]. Se calcula sumando las primeras diferencias de intensidad entre pixels vecinos a lo largo de un línea de rastreo. El criterio es maximizar SMD_x . Tiene pobres resultados si las variaciones de nivel de gris no se producen en las líneas de rastreo. En este caso se puede modificar:

8.- Varianza del Nivel de Gris.

Alta varianza supone una imagen contrastada, mientras que una baja varianza supone una imagen borrosa.

El objetivo es maximizar la varianza (σ^2) lo que corresponde a maximizar la integral del espectro de potencia de la distribución de intensidad (Teorema de Parseval). Un problema que presenta es que la varianza no depende de la distribución espacial de los niveles de gris.

$$SMD_x = \sum_x \sum_y *I(x,y) - I(x,y+1)*$$

$$SMD_y = \sum_x \sum_y *I(x,y) - I(x,y+1)*$$

$$SMD = SMD_x + SMD_y$$

(Ec. 3.13)

3.4.3 Obtención del mejor enfoque de la imagen.

La selección del estimador de enfoque más adecuado se realiza en función de la curva de respuesta en el rango de posiciones del motor de enfoque. El objetivo en un sistema de auto-enfoque es calcular la posición del motor de enfoque que da un valor máximo en la función de respuesta, pero estas curvas de respuesta están sometidas a perturbaciones provenientes por un lado de variaciones temporales en la iluminación, proceso de digitalización, ruido o interferencias. Estos problemas se deben tener en cuenta ya que varias de las funciones anteriores calculan derivadas y amplifican los efectos del ruido. Otro elemento que influye en el tipo de curva de respuesta obtenida es el tamaño de la ventana sobre la que opera el estimador. Una imagen puede contener la proyección de puntos situados a diferente distancia, esto produce que se presenten en la función criterio varios picos. Por tanto para obtener un enfoque correcto debemos seleccionar el área que se debe enfocar. Otro problema es que la distancia de enfoque cambia con la magnificación de la lente (zoom) y los puntos característicos pueden cambiar de coordenadas. Otro efecto es la reducción de brillantez al variar el estado de las lentes. Esto se puede compensar normalizando los valores de intensidad (media en la ventana).

Todos estos efectos dan lugar a que la curva de respuesta presente máximos locales, varios de los cuales pueden ser pronunciados. Krotkov [Krotkov-89] presenta un análisis comparativo de los métodos anteriores obteniendo como resultado que el método de Tanenbaum (magnitud del gradiente) es el que presenta una curva de respuesta con un máximo muy pronunciado dando un comportamiento más robusto. Este resultado es corroborado por otros autores y por la experimentación realizada en la cual se han comparado las medidas de la magnitud del gradiente, entropía del histograma, laplaciana y varianza de los niveles de gris. Este comportamiento monotónico permite utilizar algoritmos de búsqueda basados en el gradiente ascendente que permiten encontrar el máximo en pocas iteraciones. La búsqueda de **Fibonacci** permite seleccionar los incrementos de forma óptima, fijando el máximo número de iteraciones para la resolución del accionador del motor de enfoque.

Para minimizar el ruido se plantean varias soluciones. Una opción es establecer la curva de ruido temporal, filtrar las intensidades muestreadas y/o los valores criterio (media de la imagen), o utilizar valores umbral para la función basada en el gradiente para eliminar los picos. Tenenbaum [Ten-70] utiliza un parámetro adaptativo, Schalg [Schlag-83] utiliza un múltiplo de la desviación estándar de la distribución de Rayleigh del ruido en una imagen de bordes.

3.4.4 Cálculo de la Profundidad a Partir del Enfoque.

El problema que se plantea es este caso consiste en dados el punto P' proyección de un punto $P(X,Y,Z)^T$, la distancia focal y la posición del motor de enfoque que da el mejor enfoque de P' , calcular la distancia Z a la que se encuentra el punto P . Pentlan [Pent-85] presenta una solución basada en la ley de las lentes para extraer la distancia absoluta. Requiere medir la cantidad de borrosidad de un punto particular de la imagen (profundidad a partir del desenfoque). Esto es difícil ya que es función tanto de las características de la escena como de las lentes. Jarvis [Jar-76, Jar-83] presenta también una solución a este problema. Para lentes finas sin aberraciones la ley de las lentes (Ecu. 3.7) puede resolverse:

$$\frac{1}{f} = \frac{1}{\tilde{a}} + \frac{1}{\tilde{a}p} \quad \text{Ley de las lentes} \quad (\text{Ecu. 14})$$

f : distancia focal, determina el grado de magnificación (zoom). Se considera fija.
 p : posición del motor de enfoque. La distancia se considera función lineal de p
 $p=0$: distancia de enfoque infinita

- \tilde{a} : es un parámetro desconocido por lo que se requiere un proceso de calibración.
- \tilde{a} : es la distancia del punto a un punto fijo de la estructura de la cámara.
- β : es la distancia del plano principal primario al punto donde se mide \tilde{a} .

Despejando $Z = \tilde{a} + \beta$, distancia del objeto a lo largo del eje óptico desde el punto principal primario, el error en la medida del punto i vendrá dado por:

$$e_i = \tilde{a}_i \beta & Z_i = \frac{(\tilde{a}_i p_i f)}{\tilde{a}_i p_i} \quad (\text{Ec. 3.15})$$

Donde: $\beta - Z_i$ es el valor estimado.
 \tilde{a}_i valor real medido.
 Si disponemos de N medidas (\tilde{a}_i, p_i) podemos determinar los parámetros β y \tilde{a} , que satisfacen:

$$\min_{j=1}^N \left[e_i^{*2} \right] \quad \min E = \min_{j=1}^N \left[\left(\hat{a}_i \frac{(\tilde{a} p_i / f) f}{\tilde{a} p_i} \right)^2 \right] \quad (\text{Ec. 3.16})$$

3.4.5 Precisión y Profundidad de Campo.

Debido a las limitaciones del sensor (resolución espacial finita) y las lentes, puntos a diferente distancia se proyectarán sobre el mismo punto sensor. La distancia entre los planos más cercanos y lejanos en los que la definición es satisfactoria se denomina profundidad de campo.

Tal como se muestra en la fig. 3.5 para aquellos puntos fuera de la distancia de enfoque tendremos un círculo de confusión, cuyo diámetro dependerá de la distancia de los planos del objeto a la distancia de enfoque.

Sea c el diámetro máximo admisible del círculo de confusión (determinado por el tamaño del punto sensor).

Sean x_1, x_2 distancias máximas y mínimas.

$$\frac{c}{a} = \frac{y_1 y_2}{y_1 x_1} = \frac{y_2 y_1}{y_2 x_2} \quad (\text{Ec. 3.17})$$

c es directamente proporcional al diámetro de apertura a . Mediante la Ley de las lentes podemos obtener:

$$x_1 = \frac{y_1^2}{y_1 + f}, \quad x_2 = \frac{y_2^2}{y_2 + f}, \quad x = \frac{y^2}{y + f} \quad D = D_1 \cdot D_2 = \frac{2 x a f c (x + f)}{a^2 f^2 + c^2 (x + f)^2} \quad (\text{Ec. 3.18})$$

$$y_1 = \frac{a y}{c + a}, \quad y_2 = \frac{a y}{a - c}, \quad y = \frac{x f}{x + f} \quad (\text{Ec. 3.19})$$

Que representa la precisión alcanzable y da una medida del error producido, necesario en los algoritmos de integración.

$$x_1 = \frac{x a f}{c + a}, \quad x_2 = \frac{x a f}{a - c}$$

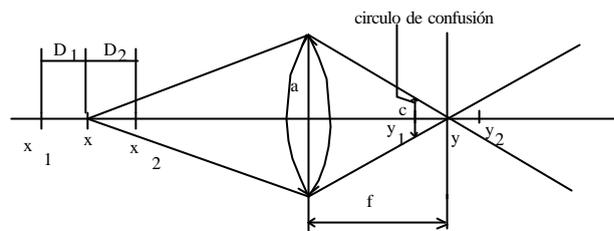


Figura 3.6 Profundidad de campo.

3.5 FUSIÓN SENSORIAL

Las características y limitaciones de cada proceso de extracción de profundidad de la escena llevan a la necesidad de integración, no solo de medidas, si no también de técnicas que faciliten la fusión estereoscópica. Los estudios sobre la visión humana permiten observar ese proceso de integración. Por ejemplo la fusión binocular humana no se produce si la disparidad es superior al límite de Panum [Jul-71]. Por lo tanto el movimiento ocular es importante en la fusión estereoscópica. La integración de técnicas ha sido objeto de un gran tratamiento en los últimos años, así Krotkov [Krotkov-89] presenta un sistema estéreo utilizando la información de estéreo, enfoque y convergencia de forma cooperativa. Abbot y Ahuja [Ahuja-93] integran enfoque, estéreo, convergencia e interpolación de superficies para ir extendiendo el mapa del entorno.

Cada uno de los métodos descritos presenta unas características específicas en tres aspectos diferentes [Ahuja-93]: características de las imágenes requeridas, capacidad y complejidad, y precisión.

! Características de las imágenes requeridas:

- Estéreo:
- Requiere la localización de características (puntos, bordes) que son detalles de alta frecuencia. La

- imagen debe estar bien enfocada por lo que precisa una gran profundidad de campo.
- Presenta problemas con periodicidades en las imágenes al requerir la correspondencia.
- Es preferible una longitud focal pequeña para tener un campo de vista mayor.
- Enfoque:
 - Requiere pequeña profundidad de campo para alcanzar cierta precisión.
 - Requiere una longitud focal elevada para que la función criterio presente picos pronunciados.

! Capacidad y complejidad:

- Convergencia: permite dar una estimación grosera y reducir la disparidad estereo entre ambas imágenes. Disminuye el espacio de búsqueda en la correspondencia.
- Estéreo y convergencia: no permiten calcular la profundidad para aquellos elementos no visibles en ambas cámaras.
- Estéreo : precisa el conocimiento de la posición y orientación de las cámaras.
- Enfoque y Convergencia: producen estimaciones para puntos simples, mientras que Estéreo puede estimar muchos puntos simultáneamente.

! Precisión:

- El método de enfoque está limitado por la profundidad de campo de las lentes (área del pixel). Es menos sensible al ruido ya que la función criterio opera sobre regiones. La precisión disminuye con la distancia al objeto de interés.
- Estéreo también depende del área del pixel pero se puede aumentar la resolución utilizando primitivas de mayor nivel (bordes) mediante técnicas subpixel. Depende también de los parámetros de posicionamiento relativo de la cámara.

La estereoscopia basada en características (bordes o segmentos de borde) requiere una localización de detalles de alta frecuencia lo que supone una escena bien enfocada en los objetos principales de la escena. Esto implica que la profundidad de campo del sistema de formación de imagen debe incluir todos los objetos, lo cual supone un valor elevado de profundidad de campo. Esta característica se contrapone a la de un sistema de enfoque que precisa una pequeña profundidad de campo para obtener una precisión aceptable.

La segunda interrelación importante entre estereoscopia y enfoque es el carácter monocular de la segunda que le permite obtener una estimación de profundidad más robusta en escenas donde la falta de visibilidad de alguna característica se produce en alguna de las imágenes, o cuando en las imágenes existen zonas con gradiente de disparidad elevado en las cuales los algoritmos estereo tienden a producir falsas correspondencias debido a la imposición de las restricciones de continuidad de la escena. Este efecto es más acusado en los objetos huecos formados por estructuras finas, donde se suele incumplir la condición de orden y pueden existir bordes muy próximos pero con disparidades significativamente diferentes.

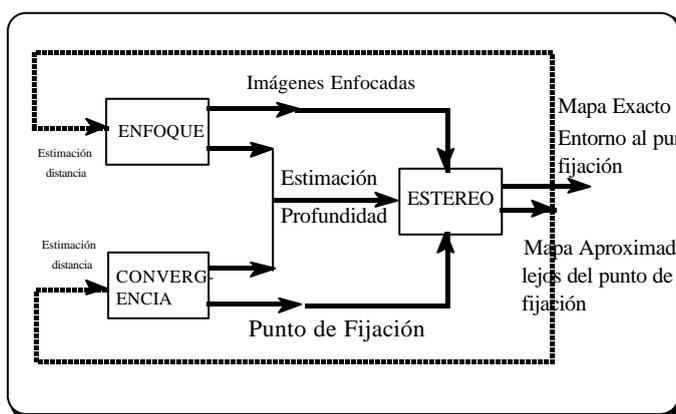


Figura 5. Sistema cooperativo de extracción de profundidad

La tercera interacción importante se produce entre la convergencia y la visión estereo. La convergencia permite dar una estimación de profundidad en el punto de fijación pero esto carece de mucha importancia al tratarse de un único punto. En cambio, la convergencia permite reducir el rango de disparidades en la zona central de la imagen, paso clave para lograr un comportamiento robusto en la visión estereoscópica. Estas interrelaciones también se producen en sentido contrario, es decir la convergencia necesita un mapa de disparidad aproximada para decidir el punto de fijación para iniciar el proceso de convergencia, si no pueden presentarse problemas, tendiendo a soluciones no deseadas. igualmente uno de los

problemas de los algoritmos de enfoque es decidir qué hay que enfocar y a qué distancia aproximada se encuentra de forma que no se enclave en máximos locales o en objetos de fondo.

El objetivo de la fusión sensorial es la combinación coherente de medidas tomadas en tiempos diferentes o con

sensores diferentes, incorporando la incertidumbre en las mediadas a las propias estrategias de navegación, en la construcción y adaptación de representaciones globales, en el reconocimiento de los objetos y sus localizaciones. La solución consiste en una representación del entorno mediante primitivas y medidas de incertidumbre en sus parámetros.

4 RESULTADOS.

Se describe a continuación la implementación realizada de las técnicas descritas anteriormente para el sistema de teleoperación ROBTET en tres aspectos diferenciados: Sistema de adquisición , Arquitectura Hardware y Software de procesamiento, y Algoritmos de Visión desarrollados. Parte de los desarrollos realizados son fruto del proyecto de investigación CICYT TAP92-0239, constituyendo la base de la presente implementación.

4.1 Sistema de adquisición.

Un primer resultado del presente proyecto ha sido la consecución de un potente sistema de adquisición estereoscópico que permite dar el salto cualitativo hacia algoritmos de visión activa. Para tal fin, y tras una etapa de búsqueda de equipos de tales características se optó por partir del desarrollo del grupo de trabajo del HIRST RESEARCH LABORATORY de GEC-Marconi encargado del desarrollo de la base estéreo binocular con control activo dentro del proyecto ESPRIT 5390 "REAL TIME GAZE CONTROL". Las investigaciones realizadas dentro de este proyecto han permitido realizar un sistema mecánico de adquisición con un sistema de control orientado a visión, en el cual las estrategias de adquisición están íntimamente ligadas a los algoritmos de procesamiento e interpretación de imágenes, principalmente orientados a seguimiento de blancos.

El sistema de adquisición integrado, constituido por una base binocular con control de acimut, inclinación, convergencia, enfoque, apertura y zoom para cada una de las cámaras permite la implantación práctica de los paradigmas de visión activa, pero también supone la necesidad de una cuidadosa y precisa calibración de los parámetros intrínsecos y extrínsecos del sistema visual. La calibración se ha abordado tanto estática (para un estado dado de las articulaciones del cabezal), como dinámicamente (transformación entre el sistema de coordenadas del extremo de cada actuador y la base del cabezal en función de los parámetros de las articulaciones). Esta última presenta una complejidad adicional para la generación de datos que se mantengan dentro del campo visual de las cámaras para un rango de valores de las articulaciones suficiente.

Parámetro	Ejes de Convergencia	Eje de Elevación	Eje de Acimut
Libertad del eje	+/- 25°	-27° - +64°	+/- 190°
Máxima velocidad	180 grados /s	150 grados /so	110 grados por segundo
Backlash debido al final flotante	0.04 grados	0.03 grados	N/A
Aceleración Máxima	1100 grados por seg ²	450 grados por seg ²	500 grados por seg ²
Precisión/resolución	0.1 grados / 0,0055 grados (16-bit)	0.02 grados / 0,0055 grados (16-bit)	0.04 grados / 0,0055 grados (16-bit)

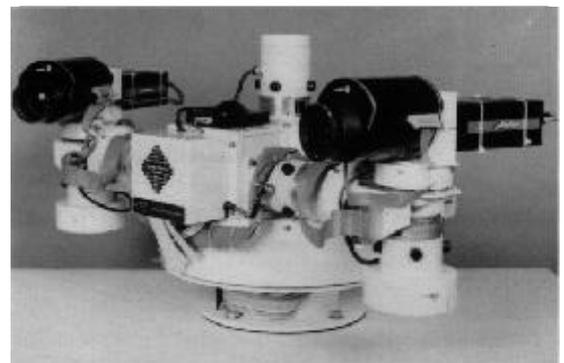


Figura 1 Cabezal estéreo binocular

Tabla 1 Especificaciones técnicas de la base estéreo.

4.2 Arquitectura.

Una etapa en el diseño del sistema ha sido la definición de la arquitectura hardware y software del sistema. La figura 2 muestra un esquema del sistema integrado, el cual está basado en un diseño distribuido de cada nivel de procesamiento. El procesamiento de bajo nivel de las imágenes captadas se realiza sobre una tarjeta con arquitectura

pipeline sobre bus VME controlada por una CPU con kernel de tiempo real. Esta CPU se encarga de la comunicación de ordenes de posicionamiento con el controlador que cierra los bucles de control de la cabeza estéreo. El procesamiento de alto nivel se realiza sobre una estación RISC con sistema operativo UNIX donde corren las tareas, a modo de cliente, de modelado gráfico, correspondencia 3D, reconocimiento y cálculo de trayectorias. El generador de trayectorias actúa sobre el servidor de control del robot para la operación sobre el entorno o la generación de iluminación láser sobre él.

4.3 Algoritmos de visión desarrollados.

Se han seleccionado las técnicas de visión estéreo pasiva, convergencia, enfoque y luz estructurada operando de forma cooperativa que permita obtener un mapa más preciso y robusto frente al ruido presente en la escena.

4.3.1 Extracción de primitivas de la imagen.

Los algoritmos de reconstrucción tridimensional de la escena se basan en una adecuada y precisa segmentación de características de la escena. Esta etapa es crítica ya que determinará tanto la precisión obtenida como la eficacia en el modelado. Se ha optado por la utilización de datos dispersos como los bordes de la imagen, al ser los bordes de la imagen suficientemente descriptivos para el tipo de entornos analizados y suponer un coste computacional muy inferior. Esta representación permite modelar fácilmente objetos poliédricos, pero tiene problemas con contornos curvos, para los cuales se debe utilizar aproximaciones polinomiales de orden superior para preservar la compacidad de la descripción y discriminar mejor la característica. Se ha profundizado especialmente en algoritmos de interpolación para la obtención de una precisión a nivel subpixel en la localización de los bordes de la escena.

4.3.2 Problema de correspondencia estéreo.

La siguiente etapa en la reconstrucción tridimensional de la escena es la puesta en correspondencia de datos característicos de la imagen. Esta etapa es crítica en la visión estéreo binocular debido a la indeterminación subyacente. Para solucionarlo se plantean un conjunto de restricciones tanto locales como globales procedentes tanto del sistema de formación de imagen (epipolaridad) como de los objetos visualizados (unicidad, continuidad de la disparidad, continuidad figural). El algoritmo desarrollado está basado en una aproximación jerárquica con un nivel local en la generación de correspondencias hipotéticas mediante las condiciones de epipolaridad, orientación,

longitud y gradiente de intensidad, y un nivel global con la criba de hipótesis en base a las restricciones globales indicadas anteriormente. Una parte importante del algoritmo de correspondencia es la técnica de "relajación" de los valores de probabilidad de correspondencia calculados en la primera etapa, y que permite reducir el número de correspondencias hipotéticas, disminuyendo el cálculo computacional necesario.

4.3.3 Control activo de enfoque y convergencia.

Uno de los principios de un sistema de visión activa es permitir el procesamiento de regiones significativas para el comportamiento global del sistema. Esta integración permite por ejemplo mantener el foco de atención sobre regiones de interés de la escena, la adaptación del sistema a cambios de iluminación, mantenimiento del enfoque de los objetos, etc.

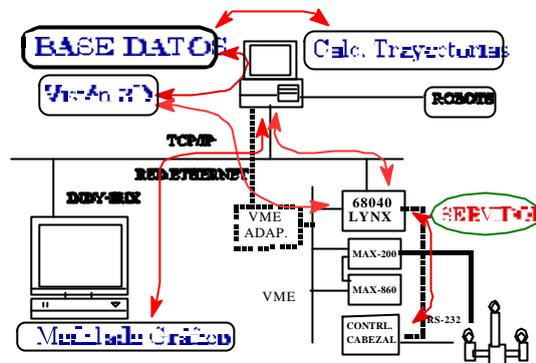


Figure 9 Arquitectura sistema de visión.

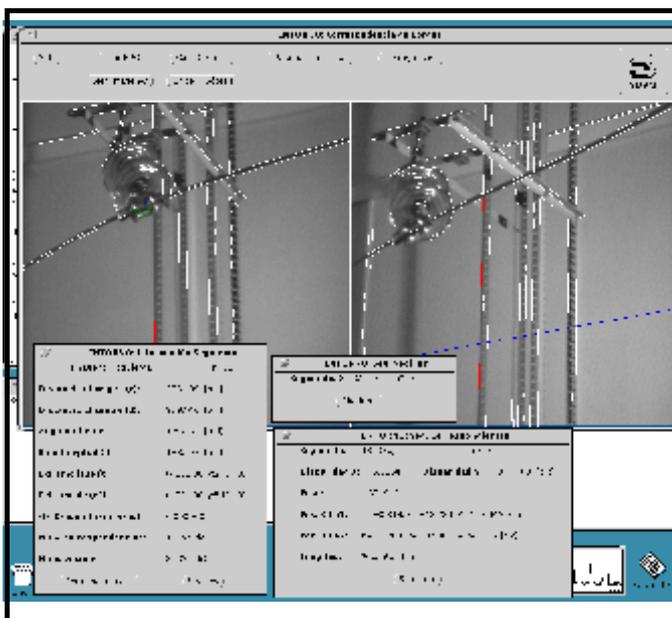


Figura 3 Cálculo de correspondencias estéreo.

Una tarea fundamental dentro de un sistema de visión activa es mantener enfocados los objetos de interés dentro de la superficie sensora. En fisiología se denomina "*Acomodación*" a la acción de las lentes oculares para enfocar el ojo. Este fenómeno de acomodación, mantiene enfocados los objetos de primer plano (objeto de interés) con el movimiento de la cabeza y constituye en sí mismo un fuente de información de la distancia de los objetos. Para ello se a evaluado una función de enfoque que relaciona la componente de alta frecuencia en un entorno de cada punto de la escena con la distancia de las superficies proyectadas.

El Objetivo de un sistema de convergencia es controlar la distancia desde las cámaras al punto de fijación en una orientación determinada, manteniéndolo próximo a la superficie del objeto. La convergencia permite al sistema concentrarse ("*foveate*") en áreas de interés para obtener una mayor precisión y confianza en la medida de profundidad, permitiendo un simplificación matemática del problema de correspondencia al situar los puntos del objeto de interés próximos al eje óptico de las cámaras facilitando la correspondencia estéreo. El estimador de convergencia está basado en la correlación cruzada y el filtro Cepstral permitiendo dar una estimación aproximada de la distancia entorno al eje focal.

4.3.4 Luz estructurada.

Las técnicas de luz estructurada permiten sustituir una de las cámaras por un patrón de luz proyectado conocido, pudiendo obtener la información 3D no solo en contornos de la escena sino también en las superficies de los objetos. En cambio esto requiere una calibración muy precisa de la geometría cámara-haz láser, para todo el ángulo de barrido de este haz. Se ha seleccionado la utilización de un plano láser como apoyo a la visión estéreo pasiva. La visión estéreo presenta graves problemas con objetos que presenten colinealidad con la epipolar, en estos casos se presentan indeterminaciones y grandes errores en la etapa de correspondencia. El plano láser permite generar un patrón de intersección claramente diferenciado en ambas cámaras. Mediante el giro del haz láser se ha conseguido una adecuada reconstrucción 3D de este tipos de objetos.

4.3.5 Fusión sensorial

Se ha seleccionado la integración de las técnicas de correspondencia estéreo, enfoque y convergencia en un algoritmo multinivel que permita ir obteniendo un mapa del entorno cada vez más refinado. De forma general se utiliza la convergencia para fijar el foco de atención lo cual disminuye el procesamiento necesario en la correspondencia y genera un mapa aproximado entorno a este punto, así mismo el algoritmo de enfoque mantiene la imagen enfocada, etapa crucial para la correcta extracción de bordes, ofreciendo un mapa aproximado que junto al anterior permite generar un mapa exacto mediante correspondencia con un coste computacional bajo. De esta forma se suplen las deficiencias de cada método: la correspondencia es difícil cuando la disparidad es elevada, pero esta disminuye si las cámaras son convergentes. Estéreo y convergencia no permiten calcular la profundidad para aquellos elementos no visibles en ambas cámaras, el enfoque sin embargo actúa monocularmente pudiendo dar estimaciones para objetos próximos que dificultan la correspondencia estéreo, evitando de este modo falsas correspondencias o "*ilusiones*".

4.3.6 Reconstrucción 3D de la escena e Integración con el planificador de trayectorias.

El modelado y representación de los objetos de la escena a partir de los primitivas tridimensionales calculadas, se realiza mediante la generación de envolventes poliédricas a partir de los contornos calculados, permitiendo la localización de entornos no estructurados para la posterior generación de trayectorias libres de colisión. El objetivo final del sistema es la integración de los datos del modelo del entorno con el planificador de trayectorias de robots. El planificador TOROS (fig. 1), desarrollado en el departamento, permite realizar el cálculo de trayectorias libres de colisión y la planificación a nivel de tarea.

5 CONCLUSIONES

En resumen, se ha acometido el problema de la integración de diferentes técnicas sensoriales para una eficiente reconstrucción tridimensional de la escena, que permita generar mapas visuales integrables dentro del conjunto robot-planificador de trayectorias, desempeñando tareas en entornos complejos y no estructurados. Se integran el control de convergencia del sistema estereoscópico, el control de enfoque y la correspondencia estéreo en un algoritmo cooperativo que permite simplificar el problema de correspondencia incrementando la fiabilidad de los datos reconstruidos. Se presenta la aplicación para la teleoperación inteligente en la reparación de líneas de distribución en carga mediante un sistema robotizado. El resultado de la reconstrucción actualiza una base de datos que permite implementar un control de colisiones de los brazos robotizados, una planificación de trayectorias para tareas estándar y la generación de información visual al operador con el objetivo de alcanzar una telepresencia suficiente en las tareas de reparación manual.

6 AGRADECIMIENTOS

Los autores del presente trabajo desean expresar su agradecimiento a la Comisión Interministerial de Ciencia y Tecnología (CICYT) por la financiación mediante el proyecto TAP92-0239 de la investigación realizada, así como a los socios participantes del proyecto ROBTET (IBERDROLA S.A. y COBRA S.A) por su colaboración en el desarrollo del prototipo para reparación de torres de distribución en carga. Este proyecto está subvencionado por OCIDE (PIE-132198).

BIBLIOGRAFÍA

- [Ahuj93] Narendra Ahuja, "ACTIVE STEREO: INTEGRATING DISPARITY, VERGENCE, FOCUS, APERTURE, AND CALIBRATION FOR SURFACE ESTIMATION", IEEE Trans. on Pattern Ana. & Mach. Intell. Vol. 15, No. 10, October 1993
- [Aracil-95] Aracil, R. et al. "3D DYNAMIC NON STRUCTURED ENVIRONMENT MODELLING BY MEANS ARTIFICIAL VISION, FOR ROBOT TRAJECTORY PLANNING". 1995 CICYT Technical Report.
- [KA-87] Kim, Y.C., and Aggarwal, J. K. "POSITIONING THREE-DIMENSIONAL OBJECTS USING STEREO IMAGES", IEEE J. Robotics & Automation, Vol. RA-3, No.4 August-1987 pp.361-373.
- [Aya-87] Ayache, N. and O. D. Faugeras, "BUILDING, REGISTRATING, AND FUSING NOISY VISUAL MAPS," Proc. Intl. Conf. Computer Vision, London, pp. 7382 (June, 1987).
- [Aya-91] Ayache, N., "ARTIFICIAL VISION FOR MOBILE ROBOTS", The MIT Press Cambridge, Massachusetts, 1991
- [Bajc-88] Bajcsy, R. "ACTIVE PERCEPTION", Proc. of the IEEE, vol. 76, n° 8, pag. 996-1005, Agosto 1988.
- [Bal-89] Ballard, D.H., "REFERENCE FRAMES FOR ANIMATE VISION", Proc. Inter. Joint Conf. on Artificial Intelligence, 1989.
- [Baker-81] H.H Baker, T.O Bindford, "DEPH FROM EDGE AND INTENSITY BASED STEREO", Proc. 7th Int Joint conf. AI, pp. 621-636, 1981
- [Barnard-82] Barbard S.T., Fischler M.A., "COMPUTATIONAL STEREO", ACM Computing surveys, vol. 14, pp. 553-572. Dec. 1982
- [Binford-87] H.S. Lim, T.O. Binford, "SETREO CORRESPONDENCE: A HIERARCHICAL APPROACH", Proc. Image understanding Workshop, pp. 234-241, Feb. 1987.
- [BL-84] M. Berthod and P. Long. "GRAPH MATCHING BY PARALLEL OPTIMIZATION METHODS: AN APPLICATION TO STEREO VISION". In Proc. International Conference on Pattern Recognition, pageJ 841-843, August 1984. 7th, Montreal, Canada.
- [Boyer-88] K.L. Boyer, A.C Kak, "STRUCTURAL STEREOPSIS FOR 3-D VISION", IEEE Trans. Pattern Anal. Machine Intell., Vol. 10, pp. 144-166, Mar. 1988
- [CocMed-89] S.D. Cochran, G. Medioni, "ACCURATE SURFACE DESCRIPTION FROM BINOCULAR STEREO", Proc. image Undersatnding Workshop, vol. 1, pp. 856-869, My 1989.
- [Coombs-92] Coombs, J. David. "REAL-TIME GAZE HOLDING IN BINOCULAR ROBOT VISION". Phd Thesis. University of Rochester. 1992
- [DA-89] Dhond U.R, Aggarwal J.K., "STRUCTURE FROM STEREO - A REVIEW", IEEE Trans. on Systems, Man. and Cybernetics, Vol. 19, No.6, Nov/Dec. 1989.
- [DC-87] R. Deriche , J-P. Cocquerez. "EXTRACTION DE COMPOSANTS CONEXES BASÉE SUR UNE DETECTION OPTIMALE DES CONTOURS" In Cognitivia, pages 1-9, CESTA, 1987.
- [DP-86] M. Drumheller and T. Poggio. "ON PARALLEL STEREO". In Proc. International Conf. on Robotics and Automation , pages 1439-1448, April 1986. San Francisco, CA, USA
- [Fau-91] O.D. Faugeras. "ARTIFICIAL 3D VISION". 1991
- [Fau-93] Faugeras, O. "THREE-DIMENSIONAL COMPUTER VISION". 1993 The MIT Press.
- [Gennery-77] D.B. Genery. "A STEREO VISION SYSTEM FOR AUTINOMOUS VEHICLE" Proc. 5th Int. Joint conf. Artificial Intelligence, Vol.2, pp 576-582, Aug. 1977.
- [Grim-81] Grimson, W. E. L., "FROM IMAGES TO SURFACES: A COMPUTATIONAL STUDY OF THE HUMAN EARLY VISION SYSTEM", MIT Press, Cambridge, Massachusetts (1981).
- [Grim-85] W.E.L. Grimson. "COMPUTATIONAL EXPERIMENTS WITH FEATURE BASED STEREO ALGORITHM", IEEE Trans. Pattern Anal. Machine Intell., Vol. PAMI-7, pp. 17-34, Jan. 1985.
- [Hor-86] Horn, B.K.P, "ROBOT VISION", MIT Press, Cambridge, Mass., 1986.
- [HoSk-89] R. Horaud, T. Skordas. "STEREO CORRESPONDENCE THROUGH FEATURE GROUPING AND MAXIMAL CLIQUES", IEEE Trans. Pattern anal. Machine Intell., Vol.11, pp. 1168-1180, Nov. 1989.
- [HZ-74] Hetch E., Zajac, A., "OPTICS", Addison-Wesley Massachusetts 1974.
- [Jiménez-91] Jimenez, L. M., Sebastian J.M, Aracil, R. "ANALISIS DE OBJETOS PRISMÁTICOS MEDIANTE UN PAR ESTEREOSCÓPICO DE CÁMARAS. 1991 INTERNAL REPORT DEP. OF SYSTEMS ENGINEERING AND AUTOMATION". Polytechnicl University of Madrid.
- [Jiménez-94] Jimenez, L.M. "MODELADO DE ENTORNOS 3D NO ESTRUCTURADOS MEDIANTE VISION ARTIFICIAL PARA PLANIFICACIIÓN DE TRAYECTORIAS DE ROBOTS". 1994 Technical Report Dep. of Systems Engineering and Automation. Polytechnic University of Madrid.
- [Jul-60] B. Julesz. "BINOCULAR DEPTH PERCEPTION OF COMPUTER-GENERATED PATTERNS", Bell syst. Tech.

- J., vol. 39, pp.1125-1162, 1960.
- [Jul-71] B. Julesz. "FOUNDATIONS OF CYCLOPEAN PERCEPTION", Univ of Chicago Press, 1971.
- [KA-87] Y.C. Kim, J.K. Aggarwal, "POSITIONING THREE-DIMENSIONAL OBJECTS USING STEREO IMAGES", IEEE J. Robotics Automat., Vol. RA-3, pp 361-373, Aug. 1987.
- [KH-75] Kuglin, C.D., Hines, D.C. "THE PHASE CORRELATION IMAGE ALIGNMENT METHOD", Proc. IEEE Int. Conf. on Cybernetics and Society 1975, pp. 163-165.
- [Krot-89] Krotkov, E.P., "ACTIVE COMPUTER VISION BY COOPERATIVE FOCUS AND STEREO", Springer-Verlag, New-York, 1989.
- [LG-86] P. Long-Limozin, G. Giraudon. "STEREO MATCHING USING CONTEXTUAL LINE REGION PRIMITIVES". In Proceedings of ICPR '86, Paris France, pages 974-977, IEEE, October 1986.
- [MF-81] J.E.W. Mayhew, J.P. Frisby, "PSYCHOPHYSICAL AND COMPUTATION STUDIES TOWARDS A THEORY OF HUMAN STEREOPSIS", Artificial Intelligence, 17:349-386, 1981
- [MN-85] G. Medioni, R. Nevatia, "SEGMENT-BASED STEREO MATCHING". Computer Vision, Graphics, and Image Processing, 31:2-18, 1985
- [MN-89] R. Mohan, G. Medioni, R. Nevatia, "STEREO ERROR DETECTION, CORRECTION, AND EVALUATION". IEEE Trans. Pattern. Anal. Machine Intell. Vol. 11, pp. 113-120, Feb 1989
- [Movarec-83] H.P. Movarec. "THE STANFORD CART AND THE CMU ROVER", Proc. IEEE, Vol. 71, pp.872-884, July 1983.
- [MP-76] D. Marr, T. Poggio. "COOPERATIVE COMPUTATION OF STEREO DISPARITY". Science, 194:283-287, 1976
- [MP-79] D. Marr, T. Poggio. "A COMPUTATIONAL THEORY OF HUMAN STEREO VISION" Proc. Royal Soc. London, Vol B 204, pp 301-328, 1979
- [Marr-82] D. Marr, "LA VISION", Ed. Freeman, 1982
- [MT-94] S. Marapane, M. Trivedi. "MULTI-PRIMITIVE HIERARCHICAL (MPH) STEREO ANALYSIS", IEEE Trans. on Pattern Ana. & Mach. Intell. Vol 16, No.3 March 1994.
- [Nev-82] R. Nevatia, "MACHINE PERCEPTION", Prentice-Hall, Englewood cliffs, N.J., 1982
- [Nish-84] H.K. Hishihara, "PRACTICAL REAL-TIME IMAGING STEREO MATCHER", Opt. Eng., Vol. 23, pp. 536-545, Sept. 1984.
- [NP-84] K. Nishihara, T. Poggio. "STEREO VISION FOR ROBOTICS". In R. Paul and M. Brady, editors, Robotics Research, The First International symposium, pages 489-505, MIT Press, 1984.
- [Otha-85] Y. Otha, T. Kanade, "STEREO BY INTRA- AND INTER-SCANLINE SEARCH USING DYNAMIC PROGRAMMING" IEEE Trans. Pattern anal. Machine Intell., Vol. PAMI-7, pp. 139-154, March 1985
- [Pent-85] Pentland, A., "A NEW SENSE FOR DEPTH OF FIELD," Proc. Intl. Joint Conf. Artificial Intelligence, Los Angeles, pp. 988-994 (August, 1985).
- [Person-77] Person J.J, Hines D.C, Golosman S, Kugli C.D., "VIDEO-RATE IMAGE CORRELATION PROCESSOR", SPIE v.119, Applications of Digital Image Processing, pp.197-205, San Diego, 1977.
- [PonHar-89] T. Pong, R.M. Haralick, L.G. Shapiro, "MATCHING TOPOGRAPHIC STRUCTURES IN STEREO VISION", Pattern Recognition Letters, Vol.9, pp. 856-869, may 1989.
- [PPMF-86] S. Pollard, J. Porrill, J. Mayhew, and J. Frisby. "DISPARITY GRADIENT, LIPSCHITZ CONTINUITY, AND COMPUTING BINOCULAR CORRESPONDENCE". Internal Report AIVRU-010, University of Sheffield, England, 1986.
- [Reinoso-95] Reinoso, O. "ANÁLISIS Y GENERALIZACIÓN DE ALGORITMOS DE VISIÓN ARTIFICIAL CON PRECISIÓN SUBPIXEL", Phd. Tesis Univ. Politécnica de Madrid, 1995.
- [Schl-83] Schlag, J. F., A. C.; Sanderson, C. P. Neumann, and F. C. Wimberly, "IMPLEMENTATION OF AUTOMATIC FOCUSING ALGORITHMS FOR A COMPUTER VISION SYSTEM WITH CAMERA CONTROL," Technical Report CMU-RI-TR 314, Carnegie Mellon University (August, 1983).
- [ShePel-90] D. Sherman, S. Peleg. "STEREO BY INCREMENTAL MATCHING OF CONTOURS", IEEE Trans. on Pattern Ana. & Mach. Intell. Vol 12, No.11 November 1990.
- [Sko90] P. Paget & T. Skordas. "AN OPTIMAL SOLUTION FOR MOBILE CAMERA CALIBRATION". 1990 IEEE Inter. Conf. on Robotics and Automation. Vol 1 pag. 34 May 13-18 1990
- [Ten-70] Tenenbaum, J. M., "ACCOMMODATION IN COMPUTER VISION," Ph.D. Dissertation, Stanford University (November, 1970).
- [Wan92] Y.F. Wang & David I. Cheng. "THREE-DIMENSIONAL SHAPE CONSTRUCTION AND RECOGNITION BY FUSING INTENSITY AND STRUCTURED LIGHTING ". Pattern Recognition Vol 25 No 12 Dec. 1992
- [YS-89] Yeshurun Y., Schwartz E. "CEPESTRAL FILTERING ON A COLUMNAR IMAGE ARCHITECTURE: A FAST ALGORITHM FOR BINOCULAR STEREO SEGMENTATION", IEEE Trans. on Pattern Analysis and Machine Intelligence, 11(7), July 1989