



UNIVERSIDAD NACIONAL DE COLOMBIA

# **Corrección del Error en el Proceso de Registro en los Sistemas de Realidad Aumentada Utilizando Técnicas Heurísticas**

**Gloria Elena Jaramillo Rojas**

Universidad Nacional de Colombia  
Facultad de Minas – Escuela de Sistemas  
Medellín, Colombia

2011

# **Corrección del Error en el Proceso de Registro en los Sistemas de Realidad Aumentada Utilizando Técnicas Heurísticas**

**Gloria Elena Jaramillo Rojas**

Tesis presentada como requisito parcial para optar al título de:

**Magíster en Ingeniería de Sistemas**

Director:

Ph.D., Jhon Willian Branch Bedoya

Línea de Investigación:

Inteligencia Artificial

Grupo de Investigación:

Grupo de Investigación y Desarrollo en Inteligencia Artificial (GIDIA)

Universidad Nacional de Colombia

Facultad de Minas – Escuela de Sistemas

Medellín, Colombia

2011



*La constancia es la virtual por la que todas  
las cosas dan su fruto.*

*Arturo Graf*

## Resumen

Desde la masificación de los primeros computadores en la década de los 90's su uso estaba restringido a desarrollar tareas específicas y cálculos repetitivos. En la actualidad, los computadores y sistemas informáticos están permeando cada vez más las actividades humanas, convirtiéndose en lo que Weiser define como Sistemas Ubicuos [Wei99]. Este nuevo paradigma de interacción humano-máquina implica disponer de interfaces naturales que permitan una comunicación efectiva entre el usuario y la máquina, presentando grandes retos para las ciencias de la computación en cuanto al diseño de sistemas, su modelamiento y, en particular, el desarrollo de las interfaces de usuario. En este contexto, las interfaces deben entrar en consonancia con el concepto de Amplificación de la Inteligencia [Bro96][Bro96a] tal que permita la creación de sistemas capaces de amplificar o mejorar las capacidades cognitivas humanas en lugar de imitarlas o reemplazarlas.

En esta búsqueda de interfaces cada vez más naturales que extiendan las capacidades perceptivas humanas, surge la creación de nuevos entornos y metáforas de visualización como la Realidad Virtual (RV) y la Realidad Aumentada (RA). En particular, la realidad aumentada aparece como un nuevo paradigma de visualización e interacción humano-máquina que permite al usuario obtener información adicional (información virtual) de su entorno e incluso manipular esta información conservando su relación con el ambiente real. La obtención de dicha información adicional, se logra mediante la superposición en el ambiente real, de información virtual generada por computador.

La creación de un sistema de realidad aumentada involucra varias etapas, a saber: calibración de dispositivos, extracción de características, seguimiento y registro. Esta última etapa presenta grandes retos para la creación de un ambiente de RA realista, ya que aquí se une tanto la información virtual como la real; si dicha alineación no es

correcta, el sistema será visualmente incoherente. Es por esto que el registro es una etapa crítica que actualmente limita las aplicaciones de realidad aumentada. En la presente tesis, se aborda por medio de técnicas heurísticas esta limitación, que ha sido recurrente en la literatura. Como aporte se propone el uso de técnicas heurísticas, las cuales hasta ahora no han sido abordadas en la literatura, para disminuir el error existente entre la información de posicionamiento obtenida en etapas anteriores al proceso de registro (información estimada) y la información real. Dicha disminución en el error se traduce en un alineamiento real-virtual (etapa de registro) mucho más preciso y coherente, obteniendo en consecuencia sistemas o ambientes que apoyen de manera efectiva los problemas de visualización y acceso a información en una mayor cantidad de aplicaciones.

**Palabras clave: realidad aumentada, error de registro, registro sin marcas, técnicas heurísticas**

## **Abstract**

Since the first computers appear in the early 90's, its use was restricted to perform specific tasks and repetitive calculations. Today, computers are getting mixed in human activities more and more, becoming what Weiser [Wei99] defined as ubiquitous systems. This new paradigm of human-machine interaction implies the availability of natural interfaces that allow an effective communication between the user and the machine. It presents great challenges for computer science related to the systems design and modeling, and in particular, the development of user interfaces. In this context, the interfaces must be in relation with the concept of Intelligence Amplification [Bro96][Bro96a], that allows the creation of systems that can amplify or improve human cognitive perceptions rather than imitate or replace them.

In this search for more and more natural interfaces that extend human cognitive capacities, new environments and visualization metaphors such as Virtual Reality (VR) and Augmented Reality (AR) rose. In particular, augmented reality appears as a new paradigm for visualization and human-machine interaction that allows the user to obtain

additional information (virtual information) from his/her environment and even manipulate this information while preserving their relationship with the real world. Obtaining such information is achieved by the overlap in the real environment, virtual information generated by computer.

The creation of an augmented reality system involves several steps, namely, device calibration, feature extraction, tracking and registration. This last stage presents great challenges for creating an realistic AR environment, since this stage binds both the virtual and the real information, if the alignment is not correct, the system will be visually incoherent. This is why the registration is a critical stage which currently limits the applications of augmented reality. In this thesis, this limitation is tackle by means of heuristics. Contributions include the use of heuristics to reduce the error between the position information obtained in previous stages to the registration process (information estimated) and real data. This reduction in error resulting in a real-virtual alignment (registration stage) much more accurate and consistent, thus gaining support to systems and environments to solve effectively visualization problems and access to information in a larger number of applications.

**Keywords: augmented reality, registration error, markerless registration, heuristic methods.**



# Contenido

	Pág.
<b>Resumen</b> .....	<b>V</b>
<b>Lista de figuras</b> .....	<b>XI</b>
<b>Lista de tablas</b> .....	<b>XIII</b>
<b>Introducción</b> .....	<b>1</b>
<b>Motivación</b> .....	<b>4</b>
<b>Objetivos</b> .....	<b>8</b>
<b>Organización</b> .....	<b>9</b>
<b>1. Capítulo 1</b> .....	<b>11</b>
1.1 Calibración.....	11
1.2 Identificación de marcas y extracción de características.....	16
1.3 Seguimiento.....	18
1.4 Registro .....	18
1.5 Visualización.....	19
1.5.1 Dispositivos de visualización basados en monitores .....	20
1.5.2 Head-Mounted Display (HMD) .....	22
1.6 Error de Registro .....	27
<b>Conclusión</b> .....	<b>28</b>
<b>2. Capítulo 2</b> .....	<b>30</b>
2.1 Identificación de Marcas .....	31
2.2 Extracción de características naturales– enfoque markerless.....	37
2.2.1 Extractor de puntos Shi-Tomasi .....	38
<b>Conclusión</b> .....	<b>42</b>
<b>3. Capítulo 3</b> .....	<b>43</b>
3.1 El seguimiento en sistemas de realidad aumentada .....	43
3.1.1 Flujo óptico.....	44
3.2 Determinación de la pose .....	49
3.2.1 Cálculo de la información tridimensional .....	50
<b>Conclusión</b> .....	<b>54</b>

<b>4. Capítulo 4 .....</b>	<b>56</b>
4.1 Colonia de Hormigas.....	59
<b>5. Capítulo 5 .....</b>	<b>64</b>
<b>6. Conclusiones .....</b>	<b>77</b>
<b>7. Referencias .....</b>	<b>79</b>

## Lista de figuras

	Pág.
<b>Figura 1.</b> Ambiente de realidad virtual.....	8
<b>Figura 2.</b> Sistema de realidad aumentada para videoconferencia.....	8
<b>Figura 3.</b> Continuo de Milgram.....	8
<b>Figura 4.</b> Aplicaciones de realidad aumentada en medicina.....	11
<b>Figura 5.</b> Aplicaciones de museología en los sistemas de realidad aumentada.....	12
<b>Figura 6.</b> Realidad aumentada en ambientes colaborativos.....	12
<b>Figura 7.</b> Esquema del proceso de creación de un sistema de realidad aumentada.....	15
<b>Figura 8.</b> Sistemas coordinados de un sistema de realidad aumentada.....	16
<b>Figura 9.</b> Patrones de calibración de cámaras.....	18
<b>Figura 10.</b> Marcas artificiales empleadas en los sistemas de realidad aumentada.....	21
<b>Figura 11.</b> Representación del error en el registro en los sistemas de RA.....	23
<b>Figura 12.</b> Dispositivos de visualización basados en monitores.....	24
<b>Figura 13.</b> Esquema de visualización basados en monitores.....	25
<b>Figura 14.</b> Esquema de visualización basados en HMD ópticos.....	26
<b>Figura 15.</b> Esquema de visualización basados en HMD de video.....	26
<b>Figura 16.</b> Sistemas de múltiples pantallas.....	29
<b>Figura 17.</b> Dispositivos de visualización basados en lentes de contacto.....	29
<b>Figura 18.</b> Proceso de creación de un sistema de realidad aumentada.....	31
<b>Figura 19.</b> Error dinámico de registro.....	32
<b>Figura 20.</b> Escenario aumentado mediante marcas.....	35
<b>Figura 21.</b> Secuencia del proceso de superposición de objetos virtuales utilizando marcas.....	37
<b>Figura 22.</b> Marcas a color propuestas por [Sta96].....	37
<b>Figura 23.</b> Marcas a color propuestas por [Lee08].....	38

---

<b>Figura 24.</b> Marcas jerárquicas propuestas por [Tat07].....	
<b>Figura 25.</b> Marcas binarias propuestas por [Hof96].....	39
<b>Figura 26.</b> Marcas utilizadas para detectar fuerzas.....	39
<b>Figura 27.</b> El problema de la apertura.....	48
<b>Figura 28.</b> Seguimiento piramidal .....	49
<b>Figura 29.</b> Relación entre el sistema coordinado de la cámara y el euclidiano.....	54
<b>Figura 30.</b> Configuración del experimento del doble puente.....	60
<b>Figura 31.</b> Elección estocástica del nodo a visitar.....	61
<b>Figura 32.</b> Características identificadas en el extractor Shi-Tomasi.....	65
<b>Figura 33.</b> Características identificadas en el extractor Shi-Tomasi.....	65
<b>Figura 34.</b> Imagen de control secuencia de video 1.....	66
<b>Figura 35.</b> Imagen de control secuencia de video 2.....	66
<b>Figura 36.</b> Puntos reproyectados método afín no optimizado.....	69
<b>Figura 37.</b> Resultados empleando ACO.....	71
<b>Figura 38.</b> Resultados empleando ACO.....	72
<b>Figura 39.</b> Resultados empleando ACO.....	73
<b>Figura 40.</b> Comparación entre el método propuesto optimizado y el método de Pang et al. [Pan06].....	74
<b>Figura 41.</b> Comparación entre el método propuesto optimizado y el método de Pang et al. [Pan06].....	74

## Lista de tablas

	<b>Pág.</b>
<b>Tabla 1:</b> Comparación entre tecnologías de visualización.....	39
<b>Tabla 2:</b> Parámetros del método propuesto .....	89



# Introducción

Los ambientes de Realidad Aumentada (RA) o *Augmented Reality* se definen como aquellos ambientes que mezclan simultáneamente tanto objetos virtuales como reales en un mismo entorno de visualización. La RA, tal como lo define Milgram [Mil94] hace parte de una categoría de estudio más amplia denominada Realidad Mixta (RM), línea a la que también pertenece la Virtualidad Aumentada (VA), y las cuales, afirma Milgram, son un caso particular de las tecnologías de Realidad Virtual (RV). La diferencia entre realidad virtual, virtualidad aumentada y realidad aumentada radica en la metáfora de visualización.

Por un lado, la realidad virtual se fundamenta en experiencias sensoriales virtuales por medio de la visualización de un ambiente completamente sintético. La característica más importante de los entornos de RV, y factor diferenciador de los demás ambientes de visualización, es la inmersión, lo cual implica el aislamiento del usuario del ambiente real. Un ejemplo de estos ambientes es la famosa comunidad virtual Second Life, en donde el usuario es representado por un avatar personalizado e interactúa con los demás usuarios en tiempo real en un mundo 3D generado por computador.

El gran auge que ha tenido la RV se debe principalmente a su capacidad de simular entornos cuyo análogo real sería imposible o difícil de visualizar o percibir. Para que dicha simulación sea posible, la realidad virtual está soportada en dos grandes pilares, a saber: interacción multimodal y presencia. Debido a que los seres humanos percibimos los estímulos externos como sonido, imágenes, sensaciones hápticas o propioceptivas y las transformamos en representaciones internas o modelos mentales, la inmersión, la sensación de “sentir” el ambiente virtual como real, está condicionado a factores como la riqueza y la fidelidad sensorial, el grado de interacción y factores psicológicos asociados al diseño del mundo virtual, entre los que se cuentan la previsibilidad, consistencia, el uso

de objetos auxiliares y la incorporación de contenido emocional [Gut08]. La Figura 1 presenta un ejemplo de un ambiente virtual.

Por otro lado, la RA tiene como principal objetivo aumentar o mejorar la información visual que percibe el usuario de su entorno, mediante la visualización de objetos generados por computador, uniendo por tanto, el mundo real y el virtual, pero predominando el mundo real. En 1997, Azuma [Azu97] define los sistemas de realidad aumentada mediante tres características:

1. Un ambiente de realidad aumentada combina información real y virtual
2. Es interactivo en tiempo real
3. Registra información en 3-D

Con esta última característica no limita a los sistemas a aumentar sólo gráficos 3-D, ya que muchas aplicaciones renderizan etiquetas 2-D u otro tipo de información, por ejemplo auditiva; sin embargo, el sistema si debe considerar posiciones tridimensionales para poder superponer los gráficos virtuales coherentemente sobre los objetos reales 3-D.

Entre las ventajas de los entornos de realidad aumentada se pueden mencionar:

- Debido a que la superposición de los gráficos generados por computador se realiza en el ambiente real, el usuario conserva las herramientas tradicionales de interacción con los objetos del entorno, a la vez que puede manipular tanto los objetos reales como los virtuales.
- La realidad aumentada provee una forma no invasiva de presentar información adicional al usuario “mezclándola” directamente en su entorno.
- A diferencia de la realidad virtual, la realidad aumentada no separa al usuario de su entorno real, por lo que se logra un mayor nivel de realismo en la percepción.
- Desde un punto de vista tecnológico, la realidad aumentada es una solución viable y económica para el despliegue de información que en el ambiente real sería difícil de acceder o visualizar.
- La superposición de gráficos virtuales en el entorno real abre grandes oportunidades de interacción humano - máquina.

Dentro de la misma categoría de ambientes mixtos a la que pertenece la RA, se encuentra la Virtualidad Aumentada (VA), la cual tiene como principio la unión de ambos ambientes, pero en este caso predominando el ambiente virtual, en la VA se hace un mapeo (generalmente de texturas) de objetos reales hacia el ambiente virtual. La Figura

2 muestra un caso de realidad aumentada y virtualidad aumentada para una aplicación de video conferencia.

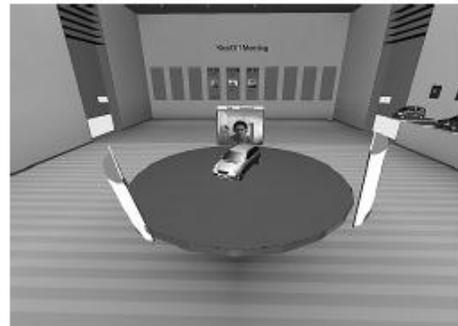
**Figura 1.** Ambiente de Realidad Virtual [Gut08]



**Figura 2.** (a) Sistema de RA para videoconferencia [Kat99]. (b) Sistema de VA para videoconferencia [Reg03]



(a)



(b)

La Figura 3 presenta el esquema de la taxonomía propuesta por Milgram [Mil94]. En este continuo se presenta el ambiente real en el extremo opuesto de los ambientes virtuales, debido a que estos últimos reemplazan o eliminan completamente el ambiente real. En la parte central del continuo se presentan los ambientes mixtos, en la cual aparece una unión de ambos mundos, la RA aparece más cerca al mundo real y la VA más cerca a los ambientes virtuales, ya que para este último caso predomina la información sintética.

**Figura 3.** Continuo de Milgram



Un caso particular de la realidad aumentada, que no aparece en el continuo propuesto por Milgram, es la Realidad Disminuida o Diminished Reality, cuyo principio se basa no en la superposición de objetos virtuales en escenarios reales, sino en la “eliminación” de objetos reales presentes en la escena. Trabajos en esta área se pueden encontrar en [Owe03].

## Motivación

Aunque la realidad aumentada es un área de investigación relativamente nueva, muchos trabajos anteriores, tanto desde el punto de vista conceptual como tecnológico, motivaron su desarrollo. La aparición de la realidad aumentada se puede atribuir a avances tanto teóricos como tecnológicos en áreas como realidad virtual, computación gráfica, inteligencia artificial, y particularmente, visión por computador. Sin embargo, estas áreas sólo permitieron el desarrollo de sistemas más precisos haciendo posible el uso de técnicas bien estudiadas en esas áreas como reconocimiento de patrones, correspondencia de imágenes, procesamiento digital de imágenes, calibración de cámaras, visualización de gráficos 3D, tracking, entre otros. En la presente tesis se retoman algunos trabajos previos a la aparición de estas áreas, específicamente teorizaciones sobre el futuro de la interacción humano-máquina y las interfaces de usuario, para argumentar y justificar la necesidad de disponer de ambientes de realidad aumentada.

Desde la década de los 80's, tomando como base la cibernética y la computación ubicua ya se afirmaba que las máquinas serían utilizadas de manera transparente y se mimetizarían en la vida cotidiana de las personas de una manera natural, tal que el usuario no se percataría que está interactuando con una máquina; lo cual no solamente sería una consecuencia natural de la evolución tecnológica, sino que también es visto como un aspecto psicológico en el cual la visión tradicional de la tecnología crea barreras mentales en cuanto a su acceso y uso, lo cual llevaría tarde o temprano a crear

interfaces que fueran mucho más transparentes y naturales. Esta idea de sistemas ubicuos fue afirmada por trabajos como el de Weiser [Wei91] y Brooks [Bro96][Bro96a]. Este último, asociando la computación ubicua al segundo término con el que se argumenta y justifica la necesidad de disponer ambientes de realidad aumentada: la amplificación de la inteligencia.

El término amplificación de la inteligencia o Intelligence Amplification (IA) es un término tan antiguo como la cibernética, sin embargo, menos conocido. Brooks [Bro96] propone la tesis en la cual

$$IA > AI$$

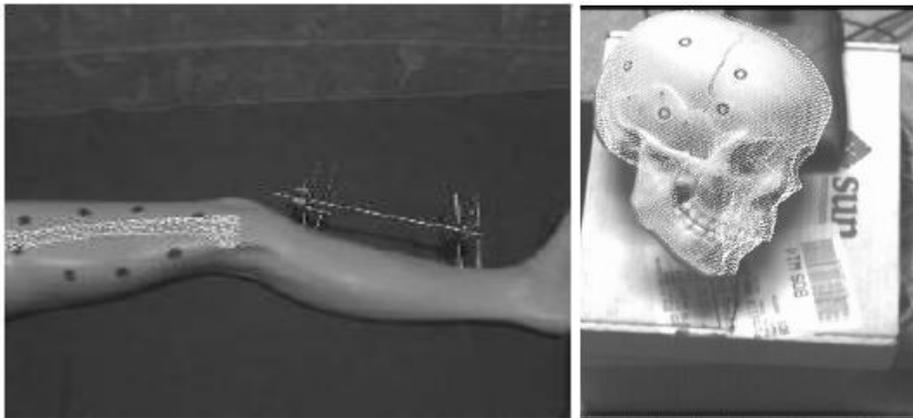
donde IA es la Amplificación de la Inteligencia y AI es Inteligencia Artificial (por sus siglas en inglés, Artificial Intelligence). La tesis propuesta por Brooks se fundamenta en que el enfoque tradicional y “estricto” de la inteligencia artificial llevaría a las ciencias de la computación por la dirección equivocada. Según Brooks, no es posible que una máquina supere las capacidades cognitivas de quien la construyó, lo cual lo ejemplifica haciendo un símil con el proceso bíblico de la creación. Por otro lado, propone Brooks, las máquinas (con cada vez más capacidades) deben ser utilizadas como una herramienta (*toolsmith*) que ayude a la solución efectiva de los problemas de la vida cotidiana, ya que aunque no es posible que las máquinas superen al hombre en inteligencia, si lo pueden hacer en aspectos como velocidad de procesamiento y capacidad de almacenamiento.

Tomando como fundamento los principios de la ubicuidad de los sistemas informáticos y la amplificación de la inteligencia, se puede argumentar la importancia de disponer de ambientes de realidad aumentada como una herramienta que brinde una solución efectiva a los problemas cotidianos, ya no mediante la interacción tradicional humano-máquina sino por medio de interfaces de visualización mucho más naturales y transparentes; entendiendo en este caso “problemas” como acceso y visualización de información. Adicionalmente, dado el uso cada vez mayor de los computadores y sistemas informáticos, el argumento anterior se convierte no sólo en una necesidad sino en una motivación para el desarrollo de sistemas de realidad aumentada en aplicaciones tanto especializadas como cotidianas.

Es por lo anterior, que cada día son mucho más numerosas las aplicaciones de realidad aumentada. Por ejemplo, imagine que usted está en una ciudad desconocida y que necesita ir a un lugar específico; mediante la realidad aumentada se podría simplemente, mediante algún dispositivo de visualización, como unos lentes especiales o la pantalla de un celular, obtener información adicional sobre el sitio actual dónde se encuentra y sobre

cómo llegar a su destino. El sistema de realidad aumentada estaría en capacidad de reconocer el lugar en el que se encuentra el usuario y superponer información virtual textual, por ejemplo, superponer sobre los edificios la dirección de las calles, o sobreponer información gráfica como flechas que indiquen hacia donde debe dirigirse. Aplicaciones de realidad aumentada surgen cada vez en muchas más áreas y con más fuerza. Aplicaciones de medicina, donde se superponen imágenes de tomografía computarizada o de resonancia magnética son cada vez más comunes, reduciendo la necesidad, en algunos casos, de cirugías invasivas en los pacientes [Mel95][Baj92][Obh92][Sta94] (ver Figura 4).

**Figura 4.** Aplicaciones de realidad aumentada en medicina



En museología, la visualización de estilos de vida y artefactos de civilizaciones antiguas ha llamado particularmente la atención para museólogos y curadores, para lo cual la realidad aumentada aparece como una solución [Whi04][Hil02][Vla01] (ver Figura 5). Aunque para este caso, otras aproximaciones, tales como la realidad virtual, han sido ampliamente utilizadas, ésta podría presentar problemas en la interacción al alejar al usuario de su entorno real; adicionalmente, en muchos casos, requiere el uso de dispositivos especiales como cascos para proporcionar la sensación de inmersión. Es por ello, que actualmente el uso de la realidad virtual, principalmente para museología y entretenimiento (videojuegos), se concentra en Internet, particularmente en museología, para la creación de visitas virtuales.

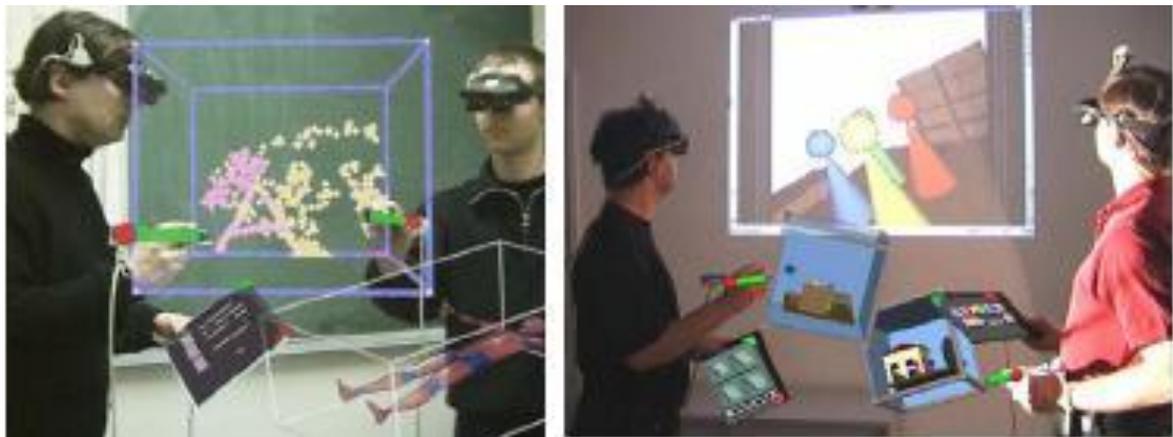
Otra de las aplicaciones que ha llamado fuertemente la atención es el uso de RA para la anotación y visualización. Muchos trabajos encontrados en este campo de aplicación, se concentran en sistemas de guía, reconocimiento o ubicación en edificaciones. En

[Kim03][Ros05][Ros95] se presentan algunos trabajos realizados en el campo de la anotación textual. En la Figura 6 se presenta una aplicación de RA para solucionar problemas de visualización en sistemas colaborativos.

**Figura 5.** Aplicaciones de museología en los sistemas de realidad aumentada



**Figura 6.** Realidad aumentada en ambientes colaborativos



Recientemente, se ha propuesto como nueva motivación el uso de la realidad aumentada para el apoyo a la toma de decisiones, uniendo tanto la visualización como el acceso a la información. Pattie Maes del MIT Media Lab es una de las más conocidas investigadoras en esta área [Mis09][Mer07]. En su proyecto SixthSense, Maes propone apoyar la toma de decisiones cotidianas mediante información adicional, por ejemplo, en una librería el usuario podría seleccionar un libro y recibir información adicional auditiva, como críticas del libro, o algún tipo de información visual indicándole al usuario el precio del artículo.

Uno de los mayores aportes de este trabajo es la capacidad de aumentar la información visual en cualquier escenario (escenarios no estructurados), ofreciendo por tanto, una solución tecnológica que se acerca cada vez más a que la RA se emplee de manera eficiente y eficaz a la solución de problemas cotidianos. Todo lo anterior fundamentado en los principios y tecnologías de la realidad aumentada.

Finalmente, reconociendo la importancia e impacto en la vida cotidiana del uso y masificación de los sistemas de realidad aumentada, resumidos en esta sección, como su uso para solucionar problemas de visualización, acceso a información y como apoyo para la toma de decisiones, en la presente tesis, se plantea como problema a resolver el ajuste de las coordenadas reales y virtuales mediante la optimización de las coordenadas estimadas; lo anterior soportado en la hipótesis que el error en el registro se puede atribuir en su mayoría a errores en las mediciones arrojadas por el tracker y la imprecisión en la obtención de las coordenadas bidimensionales donde se superponen los objetos virtuales cuando se aborda un enfoque markerless, por lo que su optimización en tiempo real podría disminuir el error en el alineamiento real-virtual.

## Objetivos

### Objetivo general

Proponer un método basado en visión por computador y técnicas heurísticas para reducir el error en el alineamiento real-virtual en el proceso de registro mediante la reducción de la diferencia entre la información de posicionamiento estimada y la información real.

### Objetivos específicos

- ❖ Determinar un método de visión por computador que permita identificar características representativas de una escena y seguirlas a lo largo de una secuencia de video.
- ❖ Obtener información bidimensional y tridimensional sin el uso de marcas para propósitos de cálculo de la pose en una escena.

- ❖ Diseñar un método basado en técnicas heurísticas que minimice el error entre las coordenadas bidimensionales estimadas y las reales que forman el espacio euclidiano sobre el cual se superpone el objeto virtual.
- ❖ Validar las técnicas y método heurístico propuesto, mediante su implementación en una plataforma computacional.

## Organización

El presente documento está organizado según la siguiente estructura temática. En el capítulo 1 se desarrolla el marco teórico, en el cual se contextualiza al lector sobre el proceso en general de la creación de un sistema de realidad aumentada. Posteriormente, siguiendo el orden lógico para la creación de este tipo de sistemas, se describe en el capítulo 2 el método implementado para la identificación de rasgos característicos de la escena, para a continuación presentar y justificar en el capítulo siguiente la técnica de tracking seleccionada. En el capítulo 4 se describen algunas técnicas heurísticas tradicionalmente utilizadas en aplicaciones de visión por computador, y particularmente, las utilizadas en el proceso de registro en reconstrucción tridimensional, algunas de las cuales, por su similitud al proceso de registro de los sistemas de realidad aumentada, han sido implementadas para estos propósitos. Finalmente, haciendo uso de las técnicas seleccionadas en las etapas de extracción de características y tracking, se presenta el método heurístico propuesto junto con los resultados obtenidos, para concluir el resultado de la investigación de la presente tesis en el capítulo 5.

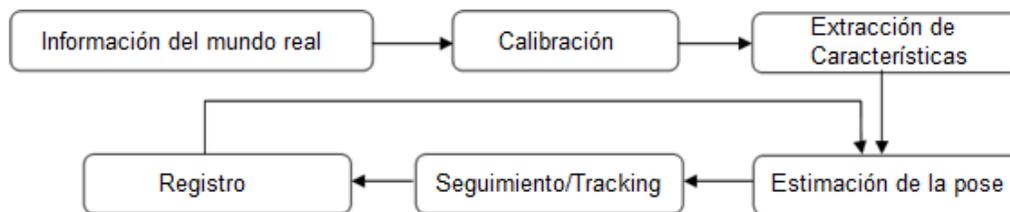


# 1. Capítulo 1

## PROCESO DE CREACIÓN DE UN SISTEMA DE REALIDAD AUMENTADA

El proceso de creación de un sistema de realidad aumentada basado en visión por computador involucra, en general, varias etapas, a saber: calibración, identificación de marcas/patrones, extracción de características, seguimiento (tracking) de características y registro (ver Figura 7). En el presente capítulo, se describe la creación de un sistema de realidad aumentada, considerando las etapas mencionadas, así como las tecnologías y dispositivos empleados en cada una de ellas

**Figura 7** Esquema del proceso de creación de un sistema de RA



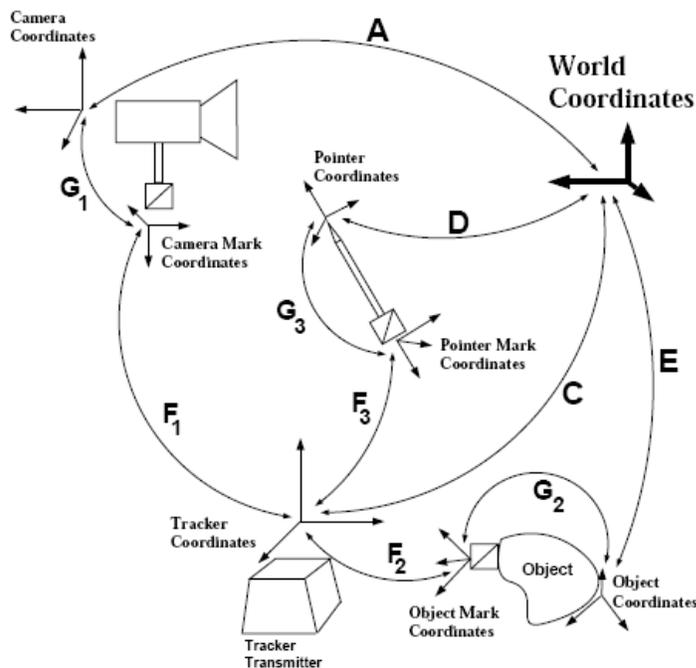
## 1.1 Calibración

Como se definió anteriormente, un ambiente de realidad aumentada consiste en la unión de información virtual y real. Este proceso requiere conocer o modelar, bien sea a priori o en tiempo real, el escenario en el cual se trabaja. Específicamente, la etapa de calibración tiene como objetivo conocer los parámetros con los cuales se está adquiriendo la información del mundo real.

Los sistemas de realidad aumentada utilizan varios dispositivos para adquirir la información visual y/o espacial del ambiente real, tales como: sensores magnéticos, sensores inerciales, sensores de ultrasonidos, GPS y/o cámaras; tales dispositivos

modelan el comportamiento del mundo real, en términos de posiciones o formas. Sin embargo los resultados de posicionamiento que arrojan los dispositivos están dados por su propio sistema de referencia. El proceso de calibración, consiste en conocer los parámetros de los dispositivos para modelar su comportamiento y mapear o transformar de manera precisa el sistema coordenado del dispositivo al sistema coordenado del mundo (ver Figura 8).

**Figura 8.** Sistemas coordenados de un sistema de realidad aumentada y sus relaciones [Tuc95]



En la literatura es posible encontrar varios métodos para la calibración de cámaras. En [Bal82][Gan84] se propone calibrar utilizando una matriz de transformación perspectiva, en donde las distorsiones lineales de los lentes no son modeladas. Tsai [Tsa85][Len88] propone uno de los métodos más ampliamente utilizados por la comunidad de visión por computador, en el cual los parámetros extrínsecos son estimados en una etapa inicial, y la distancia focal y los coeficientes de distorsión son calculados en una etapa posterior. Las distorsiones tangenciales y radiales de lentes son abordadas en trabajos como el de Weng de 1992 [Wen92]. En general, todos estos métodos buscan hallar los parámetros intrínsecos (longitud focal, coeficientes de distorsión) y la matriz de parámetros

extrínsecos que modelan el movimiento de rotación y traslación. Matemáticamente, la calibración consiste en hallar  $f_x$ ,  $f_y$ ,  $y_0$ ,  $x_0$ ,  $s$  y los coeficientes de las matrices  $R$  y  $T$ , tal que

$$P_c = R(P_w - T) \quad (1)$$

representado en matrices homogéneas:

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \\ 0 & 0 & 0 \end{bmatrix} -RT \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (2)$$

donde los subíndices  $w$  y  $c$  representan, respectivamente, los puntos en el sistema coordenado del mundo y de la cámara;  $T$  describe la posición del origen del sistema coordenado de la cámara respecto al sistema coordenado del mundo (información de traslación), y  $R$  describe la rotación que alinea el sistema coordenado de la cámara con el sistema coordenado del mundo. Ambas matrices,  $R$  y  $T$ , representan los parámetros extrínsecos.

Por otra parte, los parámetros intrínsecos ( $f_x$ ,  $f_y$ ,  $y_0$ ,  $x_0$ ,  $s$ ) deben cumplir la siguiente ecuación homogénea

$$\begin{bmatrix} x \\ y \\ w \end{bmatrix} = \begin{bmatrix} f_x & s & x_0 & 0 \\ 0 & f_y & y_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} \quad (3)$$

donde  $x_0$  y  $y_0$  son las coordenadas de los puntos principales de la imagen,  $f_x$  y  $f_y$  son las distancias focales y  $s$  representa la potencial no ortogonalidad de las filas y columnas de las celdas fotosensibles de la cámara. Cabe notar que en la mayoría de los modelos el parámetro  $s$  se omite, substituyéndolo por el valor cero.

Gráficamente, la ecuación (3) de parámetros intrínsecos surge de la relación de proyección de los puntos 3-D en el plano de la imagen, como se muestra en la Figura 9. Asimismo, la ecuación (2) surge de la relación de movimiento entre los frames ilustrado en la Figura 10, lo cual indica la rotación y traslación de los objetos de mundo real frente a movimientos relativos de los mismos.

Transformando los parámetros intrínsecos y extrínsecos en un sólo modelo, se tiene que la proyección  $p(u,v)$  de un punto  $P(X, Y, Z)$  en 3-D está dada por una matriz  $M$  tal que

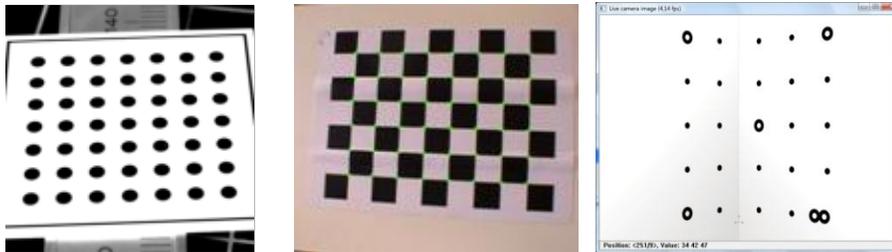
$$p = MP = M_{\text{int}} M_{\text{ext}} P \quad (4)$$

matricialmente

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & s & x_0 \\ 0 & f_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (5)$$

De la ecuación (5), la matriz M tiene 11 grados de libertad (5 parámetros intrínsecos, 3 parámetros extrínsecos de rotación y 3 parámetros extrínsecos de traslación). Para resolver este sistema y hallar los once parámetros de la calibración, es necesario conocer para un conjunto de puntos sus coordenadas en 3-D y sus proyecciones 2-D. En la calibración off-line (calibración usualmente empleada en aplicaciones de visión por computador) las coordenadas de los puntos 3-D se hallan utilizando un patrón calibrado con dimensiones conocidas como los mostrados en la Figura 9; mientras que las proyecciones de estos puntos en 2-D se hallan mediante un procesamiento digital de la imagen. Una vez se tienen estos dos valores, se obtendrá un sistema de ecuaciones no lineal.

**Figura 9.** Patrones para calibración de cámaras



Como se mencionó anteriormente, varios métodos han sido propuestos para romper la no linealidad de la matriz M y calcular los 11 coeficientes. Métodos tradicionales para resolver este sistema de ecuaciones han sido la eliminación Gaussiana, descomposición LU, QR, y métodos iterativos como el gradiente descendente o conjugado; estos últimos, dada su precisión han sido utilizados con frecuencia para refinar los resultados arrojados por otros métodos. Trabajos más recientes como los propuestos en [Bay02][Bay01] utilizan el cálculo de las ecuaciones de Kruppa y el teorema de Pascal para hallar los parámetros intrínsecos; los cuales, específicamente para aplicaciones de realidad

aumentada, son asumidos como información a priori en la mayoría de las aplicaciones [Yua05][Yua04][Zau03][Fru03][Tom00][Kat99].

La etapa de calibración no es una tarea trivial, ya que como consecuencia de la ecuación (5) se deduce que una vez se han obtenido los parámetros de la matriz  $M$ , se puede relacionar mediante este sistema de ecuaciones matriciales la información 3-D del mundo real (y su movimiento, mediante la matriz de parámetros extrínsecos) con sus proyecciones bidimensionales, por lo que el cálculo incorrecto de estos parámetros conllevaría a que la información con la que trabaja el sistema no represente de manera fiel el mundo real. En [Kum94] se describe tanto cuantitativa como cualitativamente los efectos de una incorrecta calibración en aplicaciones de visión por computador.

Considerando que un sistema de RA debe trabajar en tiempo real y con el menor retardo posible para que sea visualmente realista y a la vez redunde en una mayor eficiencia, algunos investigadores [Seo00][Che99][Kut98][Kut96] han retomado trabajos de visión por computador de “self-calibration” o calibración en línea [Fau92][Fau92a] incorporándolos a los sistemas de realidad aumentada. El principio subyacente de estos métodos consiste en la calibración de la cámara en tiempo real mediante el seguimiento de rasgos característicos o puntos de un frame al siguiente. Aunque estos métodos evitan el pre-procesamiento offline de la calibración, han mostrado no ser estables.

Similarmente, debido a que un sistema de RA puede disponer de varios dispositivos, no sólo cámaras sino también lentes especiales y una gran variedad de trackers, es necesario calibrarlos análogamente a como se hace con las cámaras. En [Tuc95] se presenta el sistema GRASP para realidad aumentada, el cual consta de un dispositivo señalador de seis grados de libertad, una cámara análoga y un tracker magnético, igualmente de seis grados de libertad. La principal característica de GRASP es la implementación de un gran conjunto de técnicas para la calibración de la imagen, el transmisor tracker, el señalador y la calibración del objeto basado en el señalador.

En general, la obtención de información del mundo real mediante sensores (sensores ultrasónicos, sensores inerciales y/o sensores magnéticos) consta de dos dispositivos: un transmisor y un receptor (ver Figura 8). El transmisor envía al receptor información de posicionamiento 3-D de los objetos en la escena; el receptor, ubicado generalmente en el usuario, a la altura de la cabeza, manipula esta información para superponer los objetos virtuales en las posiciones tridimensionales enviadas por el transmisor, y actualiza la información de posicionamiento conforme el usuario mueve la cabeza o se desplaza por

la escena. En lo que respecta a la calibración, este tipo de sistemas requieren además de establecer la relación del mundo real con la cámara, la relación entre las coordenadas del mundo real con el transmisor y las coordenadas del transmisor respecto al receptor.

Aunque este tipo de configuraciones basadas en sensores han sido ampliamente abordadas para el desarrollo de sistemas de realidad aumentada [Ble08][Caa03] [Fei93] y presentan como gran ventaja el uso directo de información 3-D de posicionamiento (en oposición el enfoque basado en visión por computador que utiliza relaciones matriciales para hallar información 3-D a partir de posiciones 2-D), el uso de sensores está siendo desplazado por técnicas de visión por computador al no requerir dispositivos adicionales y trabajar directamente con la información que está siendo capturada por la cámara.

Entre las principales desventajas de los sensores se encuentran: los sensores magnéticos se ven restringidos a ambientes controlados, debido al alto grado de imprecisión obtenido frente a la presencia de objetos metálicos en el ambiente. Por su parte, los sensores mecánicos, entre los que se encuentran los sensores inerciales, suelen ser anti-ergonómicos para los usuarios debido a su peso, ya que incorporan varios dispositivos como acelerómetros y giroscopios. En la literatura se encuentran varios trabajos que presentan como principal limitación la poca aplicabilidad del uso de sensores bajo condiciones no controladas y en espacios abiertos para sistemas de RA [Cha02] [Sat01] [You99][Stri98].

Por otro lado, los enfoques basados en visión por computador surgen como una alternativa a los problemas asociados a los sensores presentando como principal ventaja que trabaja con la misma información que está siendo capturada por la cámara. Adicionalmente, debido a que no requiere de dispositivos adicionales, el proceso de calibración y el costo computacional asociado a dicho proceso disminuye.

## **1.2 Identificación de marcas y extracción de características**

Una vez se ha realizado el proceso de calibración y se conoce como se comporta el mundo real respecto a sus proyecciones bidimensionales, es necesario saber sobre qué objeto se superpondrá la información virtual. Debido a que una de las mayores

consideraciones en los sistemas de realidad aumentada consiste en qué información superponer y sobre cuales objetos del mundo real suponerlos, y en general, debido a que sería tecnológicamente imposible contar con una base de datos que albergue todos los posibles objetos que se puedan encontrar en una escena para poder identificar y superponer el gráfico virtual correspondiente a un determinado objeto, los sistemas de realidad aumentada suelen simplificar este proceso, almacenando a priori los modelos a renderizar y ubicando marcas artificiales en la escena, tal que, el sistema pueda reconocerlas y compararlas con marcas almacenadas previamente para relacionar un objeto virtual con una marca específica. Para que las marcas sean fácilmente identificadas en el sistema se utilizan marcas binarias como las mostradas en la Figura 10, sin embargo, en la literatura se encuentran trabajos que relajan esta restricción incorporando marcas a color.

**Figura 10.** Marcas artificiales empleadas en los sistemas de realidad aumentada



Luego de detectarse las marcas a partir de algún método de reconocimiento de patrones, se extraen características o *features* a partir de las cuales se realiza el seguimiento de la(s) marca(s) en el ambiente real (considerando rotaciones o translaciones de la misma). En consecuencia, las características a reconocer deben ser únicas o casi únicas tal que permitan ser identificadas y seguidas con precisión de un frame al siguiente. Tradicionalmente, las características a detectar en los sistemas de RA han sido formas, tales como bordes y puntos de alta curvatura. En trabajos como [Baj95][Azu94] se emplean LEDs rojos como características.

## 1.3 Seguimiento

El seguimiento o *tracking* consiste en hallar la pose (posicionamiento y orientación) de las características durante toda la secuencia de video. Debido a que la información de posicionamiento de los objetos reales está dada en coordenadas tridimensionales en el sistema coordinado del mundo, el tracker debe brindar igualmente información en 3-D, tal que al superponer la información virtual en el objeto real sea visualmente coherente. En este aspecto, las configuraciones basadas en sensores presentan la ventaja de obtener y manipular información 3-D directamente del escenario real, mientras que los sistemas basados en métodos de visión por computador sólo tienen información 2-D de la escena y la información 3-D es estimada mediante relaciones geométricas o de proyección.

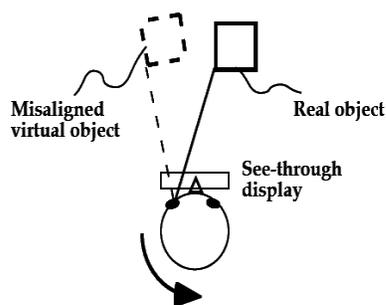
El seguimiento está fuertemente ligado a la extracción de características, tal que un mal seguimiento puede producirse tanto por una incorrecta identificación de los rasgos característicos durante la secuencia de video, bien sea por problemas de oclusión o por cambios de iluminación, o porque las características a seguir no son las adecuadas. En consecuencia, esta etapa presenta como grandes retos no sólo la estimación precisa de las coordenadas tridimensionales frente a movimientos relativos de los objetos o características en la secuencia de video, sino también se debe considerar que dicho cálculo debe hacerse con el menor costo computacional, tal que los resultados correspondan en tiempo real a la posición del usuario o la cámara en un determinado momento.

## 1.4 Registro

En esta etapa se alinea la información virtual sobre el objeto real una vez se ha identificado su pose a partir de la etapa de seguimiento. Esta etapa es crucial en los sistemas de realidad aumentada ya que un registro incorrecto implica profundas e incluso graves repercusiones en la visualización. En medicina, por ejemplo, donde se han venido desarrollando aplicaciones de visualización de imágenes de resonancia magnética (IRM) y tomografía computarizada (TC) directamente sobre el cuerpo del paciente, un error de

incluso 2 píxeles estaría por debajo de los niveles de tolerancia permitidos y podría llegar a ser fatal para el paciente. Como consecuencia, el error de alineación debe reducirse al máximo para evitar efectos negativos en la visualización y en la tarea en general que se pretenda resolver mediante ésta. Gráficamente, un error en la alineación de la información real y virtual se muestra en la Figura 13, donde el cuadro punteado representa el objeto virtual, el cual debería aparecer en la misma posición donde está ubicado el objeto real, representado por la línea sólida.

**Figura 13.** Representación del error en el registro en los sistemas de RA [Pas99]



Finalmente, una vez se ha obtenido la información de posicionamiento, los objetos virtuales son renderizados en algún tipo de dispositivo de visualización. Según la metáfora de presencia que pueda ser lograda en el entorno de realidad aumentada, se distinguen varios dispositivos de visualización, a saber: HMD, monitores y dispositivos de inmersión espacial.

## 1.5 Visualización

Recientemente en una de las revistas del Institute of Electrical and Electronics Engineers - IEEE [ISpe09] se ha publicado una lista de los retos actuales en el campo de la ingeniería, entre los que se destacan mejoras en la realidad virtual y sus tecnologías asociadas; esto deja entrever no sólo la motivación de la comunidad académica mundial en estas formas de interacción, y en su impacto, sino la importancia y los retos

tecnológicos asociados a los dispositivos de visualización que logren crear sensaciones de percepción mucho más reales.

Para visualizar un ambiente aumentado es necesario disponer de dispositivos gráficos que permitan la combinación entre el mundo real y el virtual. En general, tres configuraciones han sido identificadas: WoW o ventana al mundo, HMD (Head-Mounted Displays) y dispositivos de inmersión espaciales. Actualmente, una cuarta configuración basada en lentes ha aparecido como una prometedora tecnología de visualización, sin embargo, se encuentra todavía en etapas de desarrollo inicial y pasarán años antes de que se convierta en una tecnología comercial y ampliamente utilizada para estos propósitos.

### **1.5.1 Dispositivos de visualización basados en monitores**

Este tipo de visualización es también conocido como Ventana al Mundo (Window of the World - WoW). Dicha implementación consiste en la visualización de la realidad aumentada por medio de un monitor tradicional o “ventana”. La interacción es realizada por medio de los dispositivos tradicionales de interacción humano-máquina, principalmente, eventos de mouse y teclado. Este enfoque presenta una subdivisión asociada al modo de visualización, a saber: lentes mágicos (magic lens) y espejos mágicos (magic mirror). Los lentes mágicos [Bie93] permiten una visualización de la escena real y virtual a través de dispositivos equipados con cámara y una pantalla, entre los que se incluyen PC, PDA y teléfonos celulares. Por otra parte, los espejos mágicos requieren que cada usuario disponga de su propio hardware, debido a que es sobre el usuario sobre el que se aumentan los objetos virtuales, es por ello que esta configuración requiere identificar en cada usuario puntos característicos, los cuales son utilizados en la etapa de tracking, similarmente, también es posible ubicar directamente sobre el usuario marcas artificiales, como las mostradas en la Figura 14 (a), sin embargo, esto podría no ser cómodo u ergonómico para el usuario dependiendo de la aplicación.

La Figura 12 muestra aplicaciones de sistemas de realidad aumentada que emplean espejos mágicos y lentes mágicos respectivamente.

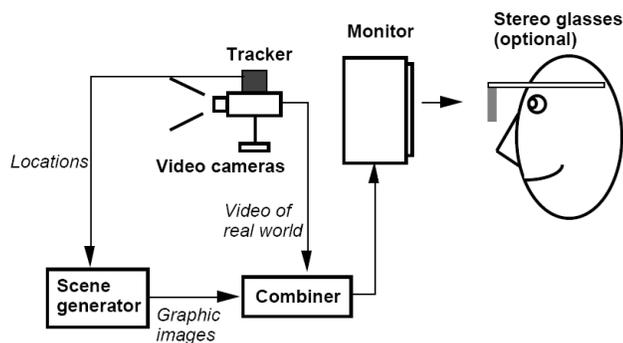
Como una característica de inmersión de esta configuración es que el usuario no es el centro del universo aumentado sino un simple espectador, por lo que la percepción de presencia en el ambiente mixto es mínimo, sin embargo, son configuraciones apropiadas para aplicaciones de telepresencia.

**Figura 12.** Dispositivos de visualización basados en monitores (a) espejos mágicos (b) lentes mágicos <sup>1</sup>



Un esquema general de un sistema de realidad aumentada que utiliza este tipo de visualización es mostrado en la Figura 13, en el cual se observa al usuario libre de dispositivos adicionales como cascos o guantes. En esta visualización, los parámetros que definen como deben orientarse o trasladarse los objetos para su correcta visualización son capturados por la cámara, la cual puede ser manipulada libremente por el usuario. La escena aumentada se presenta mezclando los gráficos generados por computador con el video capturado por la cámara en tiempo real. El escenario aumentado es presentado en un monitor convencional.

**Figura 13.** Esquema de visualización basada en monitores [Azu97]



<sup>1</sup> Tomado de: (a) <http://www.artag.net/> (b) <http://zaycloud.com/page/2/>

## 1.5.2 Head-Mounted Display (HMD)

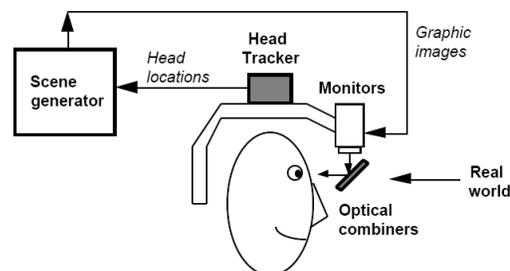
El uso de HMD ha sido tradicionalmente usado tanto en RA como en RV. Aunque los registros no son muy claros en cuanto a la fecha, muchas publicaciones coinciden en que el primer HMD del cual se tenga registro aparece en 1966 como un diseño de Iván Sutherland, cuyo propósito de visualización radicaba en las primitivas tecnologías de gráficos virtuales. Tal vez es por esta unión tecnológica de los dispositivos de visualización de la realidad aumentada y la realidad virtual, que muchos investigadores a nivel mundial consideran la realidad aumentada como un caso particular de la RV.

Los HMD son una solución viable en lo que respecta a sistemas que requieren movilidad por parte del usuario. En general, hay dos tipos de HMD: ópticos y video.

### HMD ópticos

Los HMD ópticos son una combinación de lentes transmisivos y reflectivos, que dejan pasar la luz de la escena real (similares a los lentes convencionales) a la vez que permite visualizar los gráficos por computador, reduciendo para ello, la cantidad de luz que percibe el usuario. La Figura 14 presenta un esquema general de un HMD óptico, dónde el objeto generado por computador se visualiza directamente en los lentes, al tiempo que el usuario observa el mundo real. En este caso, los parámetros de posicionamiento son capturados mediante la estimación de los movimientos de la cabeza utilizando un tracker ubicado en el HMD.

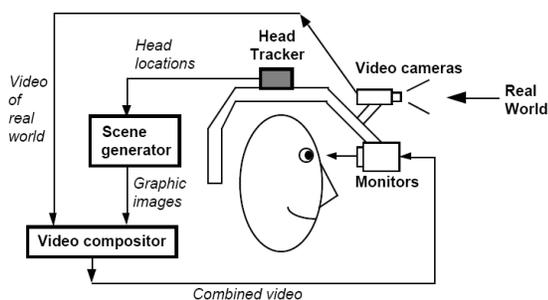
**Figura 14.** Esquema de visualización basado en un HMD óptico [Azu97]



## HMD de video

Los HMD de video a diferencia de los ópticos utilizan una o varias cámaras para capturar la información de la escena. Es en el video presentado al usuario por medio de los lentes, que se aumenta el escenario superponiendo los gráficos generados por computador (ver Figura 15). En este modo de visualización, se debe disponer de un dispositivo que “mezcle” la escena real y virtual utilizando alguna técnica como chroma keying o información de profundidad.

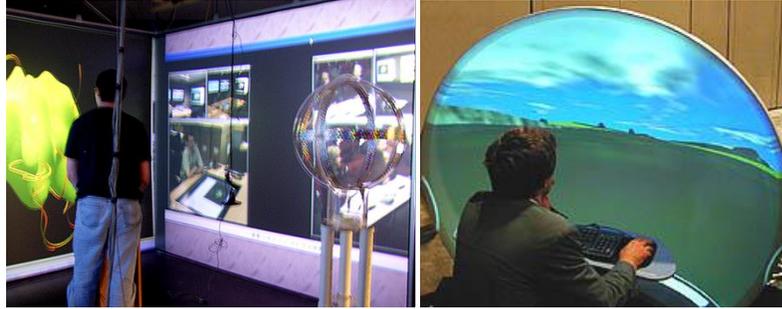
**Figura 15.** Esquema de visualización basado en un HMD de video [Azu97]



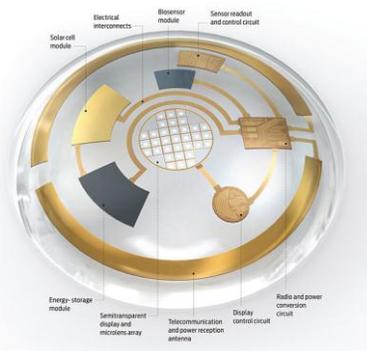
Una comparación detallada entre estas técnicas de visualización se puede encontrar en [Cat07] [Azu97].

## Dispositivos de Inmersión Espaciales (DIE)

Los dispositivos de inmersión espaciales rodean al usuario de múltiples pantallas de proyección, creando una simulación de inmersión. Este tipo de visualización, al igual que los HMD de video, pueden ser vistos como un caso especial de ventana al mundo, sin embargo, debido a que el usuario está “dentro” del ambiente mixto, se logra un mayor grado de inmersión.

**Figura 16.** Sistema de múltiples pantallas**Dispositivos de visualización de última generación**

Recientemente, la Universidad de Washington ha publicado resultados parciales de una investigación que comenzó aproximadamente en el 2006 orientada a la superposición de información directamente en el campo de visión del observador, por medio de un dispositivo de visualización que consta de unos lentes de contacto, similares a los convencionales, compuestos de pequeños circuitos de silicón y LED (ver Figura 17).

**Figura 17.** Dispositivos de visualización basados en lentes de contacto

Estos dispositivos, que funcionan vía comunicación inalámbrica, además de poder ofrecer una mejora visual por medio de texto u anotaciones en el campo de visión del ambiente real del usuario, podrían ser utilizados como un sistema de monitoreo de salud no invasivo, detectando por ejemplo, los niveles de glucosa del organismo. Esta tecnología aunque promisoría está todavía en una etapa inicial de investigación y desarrollo ya que aspectos como el movimiento relativo alrededor de la pupila y las limitaciones propias de visualización de dispositivos tan pequeños (el área de visualización de los lentes de contacto es aproximadamente de 1.2 milímetros de

diámetro) deben ser considerados y refinados, por lo que su masificación y desarrollo potencial tardará todavía unos 10 años más, como mínimo, según publicaciones de la universidad de Washington. Sin embargo, estos dispositivos de visualización están destinados a convertirse en una interfaz mucho más transparente, no invasiva y natural para el usuario.

En la Tabla 1 se presenta una comparación entre las diferentes tecnologías de visualización de ambientes mixtos disponibles hasta el momento.

**Tabla 1.** Comparación entre tecnologías de visualización

<b>HMD ópticos</b>	<b>HMD de video</b>	<b>WoW</b>	<b>DIE</b>
Mayor grado de movilidad	Mayor grado de movilidad	Mayor grado de movilidad, especialmente si se utilizan celulares, TablePC o pantallas similares como mecanismo de visualización	Movilidad restringida. Sólo hay movilidad dentro del espacio que conforman las pantallas de visualización
Más económicos y simples que los HMD de video	Retraso de decenas milisegundos	Económicos y utilizan pantallas tradicionales	Degrada la resolución
Retraso de nanosegundos	Limita la resolución de lo que usuario percibe	Mucho más seguros que los HMD ya que el usuario no es el centro del ambiente mixto	
La vista del usuario no es degradada	Menos seguros que los HMD ópticos ya que si se desconectan, el usuario queda completamente	Degrada la resolución	

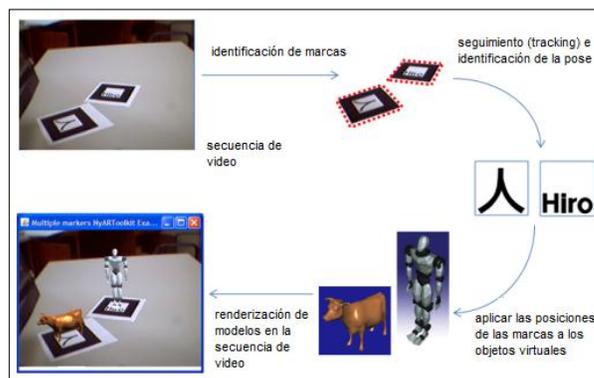
	aislado del mundo real		
Los objetos virtuales no son completamente ocluidos por los virtuales			

Como plantea Azuma [Azu97] la elección de una determinada configuración deberá obedecer a la evaluación de criterios tales como simplicidad, resolución, seguridad, distancia entre la ubicación de la cámara y la posición real de los ojos, campo de visión, retardo, estrategias de registro adicionales que puedan ser implementadas, realismo (en cuanto a facilidad de imitar aspectos como brillos y sombras) y portabilidad.

Aunque la elección de una determinada configuración dependerá en gran medida de estos aspectos, los cuales deben ser evaluados para cada aplicación en particular, el uso de HMD (basados en video u ópticos) podría decirse que es actualmente la tecnología de visualización más utilizada para aplicaciones generales, motivada principalmente por su soporte para aplicaciones en el exterior (outdoor), las cuales son uno de los retos y objetivos de las aplicaciones de realidad aumentada: la visualización de información virtual en ambientes no estructurados.

La Figura 18 resume el proceso anteriormente descrito en cada uno de las etapas.

**Figura 18.** Proceso de creación de un sistema de realidad aumentada



## 1.6 Error de Registro

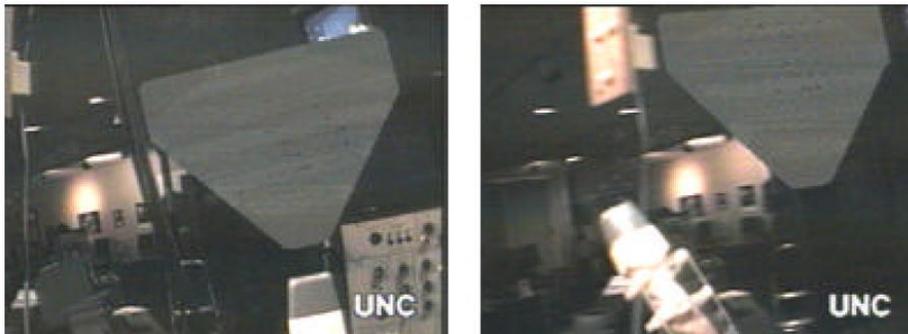
Aunque el proceso anteriormente descrito compuesto por la etapa de calibración, identificación de marcas, extracción de características, seguimiento, registro y visualización, se realice con mucha precisión, el error de registro es casi inevitable y se atribuye no sólo a imprecisiones de una etapa en particular sino que corresponde a una serie de errores que se propagan de una etapa a la siguiente.

Azuma [Azu97] atribuye el error en el registro a factores estáticos y dinámicos. Las fuentes de error estáticas son aquellas que causan errores aun cuando no hay movimiento relativo entre el usuario y los objetos reales. Este tipo de errores se deben principalmente a:

- ❖ Distorsiones ópticas
- ❖ Errores de tracking/seguimiento
- ❖ Desalineaciones mecánicas
- ❖ Parámetros incorrectos de visualización (campo de visión, pose del tracker-cámara, distancia interpupilaria)

Por otra parte, las fuentes de error dinámicas se ocasionan en presencia de movimiento relativo, y son causadas principalmente por el retardo del sistema. En estas fuentes de error se trata de minimizar el retardo end-to-end, el cual consiste en la diferencia entre el momento en que el tracker mide las coordenadas de posicionamiento y el momento en el cual se visualiza el ambiente aumentado correspondiente a esa posición. Por lo tanto, este tipo de error sólo se producirá cuando ocurre algún movimiento, si el punto de vista del usuario y el escenario permanecen constantes, el retardo no será percibido y el error dinámico será cero. La Figura 19 presenta un ejemplo de error dinámico en el registro.

**Figura 19.** Error dinámico de registro [Azu97]



Considerando la importancia de una correcta alineación entre la información real y la virtual, y que el registro continúa siendo una etapa crítica en la creación de un sistema de realidad aumentada [Pri02][Seo00][Kut98], en la presente tesis se propone disminuir el error en esta alineación, empleando un framework que no pretenda abordar una sola etapa sino que considere los efectos propagados en las demás etapas con el fin de proponer una solución que considere tanto la extracción de características y el seguimiento como el cálculo de la pose.

## Conclusión

Los enfoques de realidad aumentada que trabajan con aproximaciones de visión por computador han sido ampliamente utilizados en la literatura para determinar la posición y orientación de los objetos en el espacio, así como el seguimiento de los mismos. Para ello, se vale de técnicas ampliamente estudiadas en este campo como son el reconocimiento e identificación de patrones, análisis de escenas y seguimiento.

Tal como afirman varios autores, el registro es una etapa crucial que actualmente limita las aplicaciones de realidad aumentada, por lo que un trabajo que aborde la disminución del error en esta etapa se convierte en una gran contribución para la comunidad académica en general. Adicionalmente, tal como se afirma en [Baj95] construir un

sistema de realidad aumentada con un registro preciso es difícil y adicionalmente requiere intervenir en casi todas las etapas del proceso y no sólo en etapas individuales como se ha propuesto tradicionalmente en la literatura.

## 2. Capítulo 2

### IDENTIFICACIÓN DE MARCAS Y EXTRACCIÓN DE CARACTERÍSTICAS

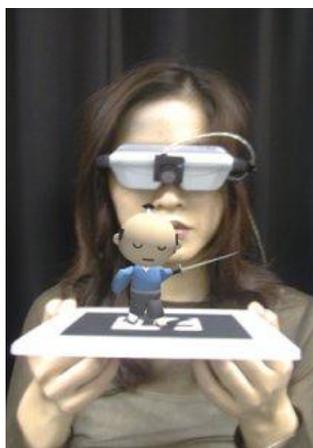
Como se ha mencionado en el capítulo anterior, los desarrollos y avances en realidad aumentada se han valido de una serie de técnicas ampliamente investigadas en otras áreas de estudio como soporte para el desarrollo de este tipo de ambientes, lo cual no implica de ninguna manera que la RA no plantee sus propios retos. Saber exactamente cuáles objetos virtuales aumentar en un determinado escenario y en qué posición exacta hacerlo, tal que sea visualmente coherente es uno de esos retos particulares de la RA; para ello, se emplean técnicas de identificación de patrones y extracción de características, ampliamente investigadas en otras áreas como análisis de imágenes, modelación de escenas y correspondencia de imágenes.

Particularmente, para el caso de realidad aumentada, los rasgos característicos que sirven de guía para obtener la información de pose (posición y orientación) de los objetos reales pueden ser obtenidos de dos maneras: mediante el uso de marcas o mediante la extracción de rasgos característicos naturales presentes en la escena, también denominados natural features. En ambos enfoques, dichas características a detectar bien pueden ser geométricas o cromáticas; considerando que en los enfoques basados en marcas se suele utilizar en primera instancia la extracción de información basada en color o características cromáticas. Por su parte, la extracción de rasgos naturales se fundamentos principalmente en información geométrica o de forma.

## 2.1 Identificación de Marcas

La ubicación de marcas artificiales en la escena real fue el enfoque que más atención ganó en los primeros desarrollos de ambientes aumentados. La razón, radica en la simplificación del problema de medir la posición y orientación del punto de vista del usuario, a partir de un proceso de segmentación de imagen que facilite la extracción de características por medio de una rápida identificación de la marca. En este enfoque los objetos virtuales aparecen sobre la marca y debido a que la forma y el tamaño de ésta son reconocidos por el sistema, la posición en la cual aparezca frente a la cámara indicará la posición relativa de la marca frente a la cámara (y por ende frente al usuario). La simplificación por tanto, podría ser expresada de la siguiente manera: si se conoce la posición de la marca respecto a la cámara, se conocerá la posición del usuario, dicha información de posicionamiento es utilizada para transformar, en términos de rotación y traslación, el objeto virtual tal que aparezca alineado con la posición de la marca. Por lo tanto, el principal reto de este enfoque radica en la correcta identificación de características de la marca a través de toda la secuencia de video.

**Figura 20** Escenario aumentado mediante el uso de marcas<sup>2</sup>



En la literatura, diferentes diseños de marcas han sido reportados, siendo más frecuente el uso de marcas planas cuadradas como las mostradas en la Figura 10, consistentes de

---

<sup>2</sup> Tomado de <http://www.hitl.washington.edu/artoolkit/>. Visitada el 02 de agosto de 2010

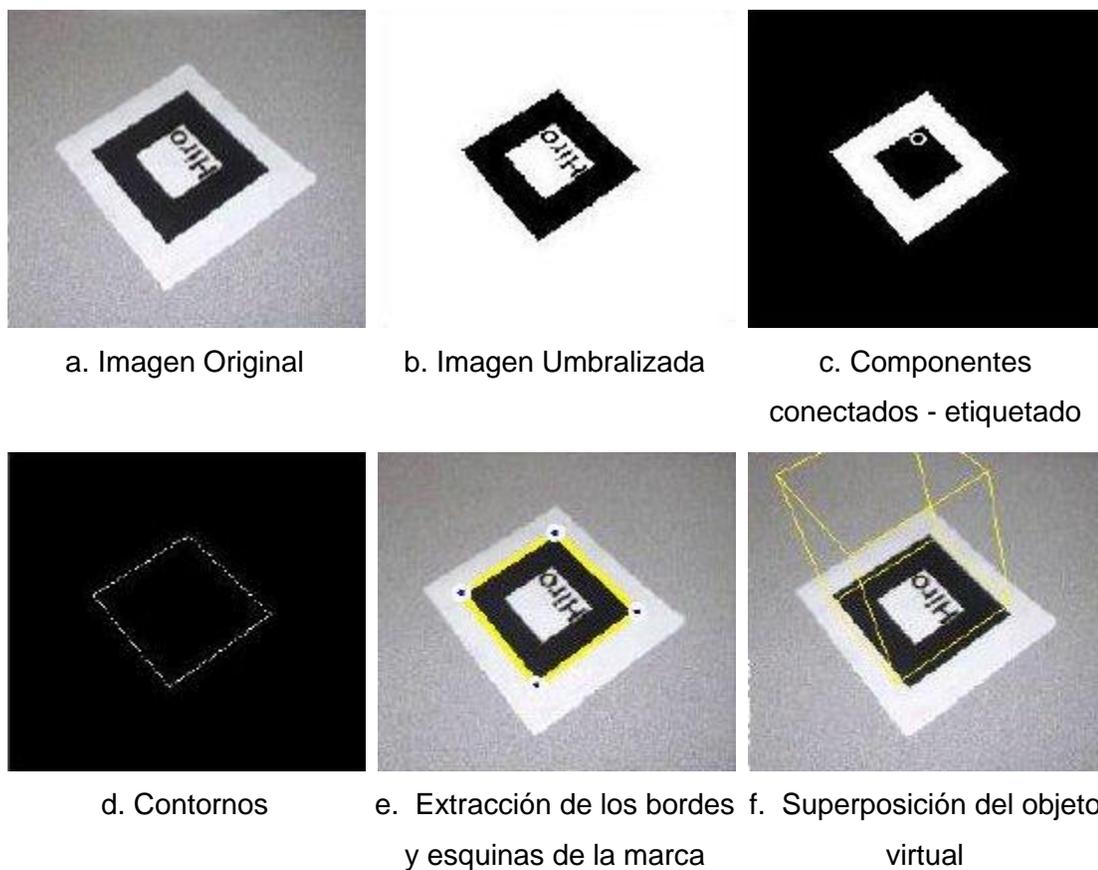
un marco negro grueso que encierra un cuadrado blanco en cuyo interior se encuentra una imagen o patrón que es reconocido por el sistema. En la documentación de la librería ARToolkit [Kat99][Kat00], probablemente la más utilizada basada en el registro utilizando marcas, se resume el proceso de registro como se presenta en el Figura 21. En general, el sistema binariza en tiempo real la información capturada por la cámara para permitir mediante segmentación el reconocimiento de regiones de alto contraste (regiones blancas y negras); debido a que el sistema conoce la forma de la marca, busca regiones con formas similares, las cuales indican la presencia de una marca. Una vez se ha identificado, se realiza un procesamiento digital a la imagen para extraer su contorno y a partir de ella los cuatro puntos donde se interceptan cada dos semirectas, dichos vértices son utilizados en el proceso de seguimiento para identificar la posición de la marca. Finalmente, la imagen interior de la marca se extrae y se compara con las imágenes almacenadas en el sistema y se le asocia el objeto virtual correspondiente.

Aunque el uso de marcas ha mostrado ser poco práctico para la mayoría de las aplicaciones de propósito general, otras aplicaciones por el contrario las han incorporado de manera exitosa, especialmente en aplicaciones indoor donde el campo de visión y los movimientos del usuario no cambian drásticamente.

Basado en el uso de marcas, varias técnicas de registro han sido propuestas. State [Sta96], argumentando que el uso de marcas a color simplifica la evaluación de píxeles a bajo nivel, redundando en una rápida inspección, propone un diseño de marcas basada en círculos concéntricos coloreados con pinturas fluorescentes. Dicho diseño hace mucho más robusto al sistema frente a posibles falsos positivos en la detección, lo cual suele ser común con el uso de marcas cuadradas debido a que otros objetos en la escena podrían presentar la misma forma y el sistema podría confundirse fácilmente; con el diseño propuesto, la probabilidad que se detecte en la escena la misma forma de círculos concéntricos con los mismos colores resulta ser muy pequeña. Para este tipo de marcas a color, como es de esperarse el proceso de identificación cambia respecto al presentado en la Figura 21, dependiendo mayormente del diseño de la marca. Trabajos como los reportados en [Lee08][Lee09] abordan específicamente el problema de la detección de marcas a color, en particular para dispositivos móviles. Considerando que si bien el color incrementa la información contenida en la marca, la resolución de la cámara incorporada en estos equipos contradice las ventajas de utilizar marcas a color. Para superar tal limitación en la detección, propone un método de corrección de color en

tiempo real consistente en reducir el ruido causado por la iluminación, identificando en primer lugar las regiones con patrones de color predefinidos y a partir de ello las estimaciones de los puntos blanco y negro de la marca (ver Figura 22).

Además del ruido, en [Fou94] se analizan otros problemas asociados a la iluminación en la creación de sistemas de RA.



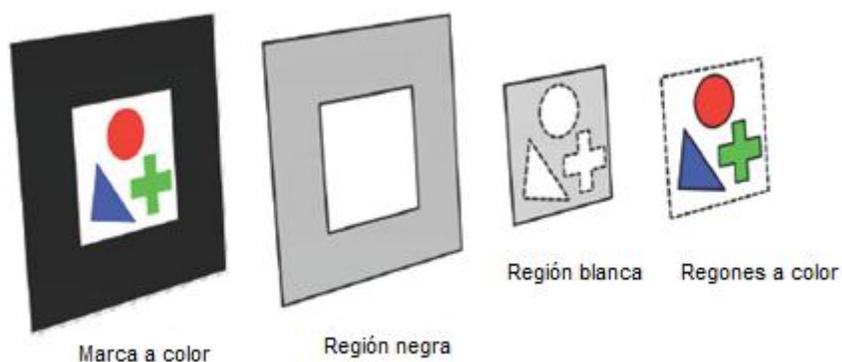
**Figura 21.** Secuencia del proceso de superposición de objetos virtuales utilizando marcas

Otras aproximaciones al diseño, proponen el uso de marcas “invisibles” como las reportadas en [Nak05] consistentes en retroreflectores traslúcidos iluminados a partir de LEDs infrarrojos. La reflexión de la luz de dichos retroreflectores es capturada por una cámara infrarroja ubicada cerca a los LEDs.

**Figura 22.** Marcas a color propuestas por [Sta96]



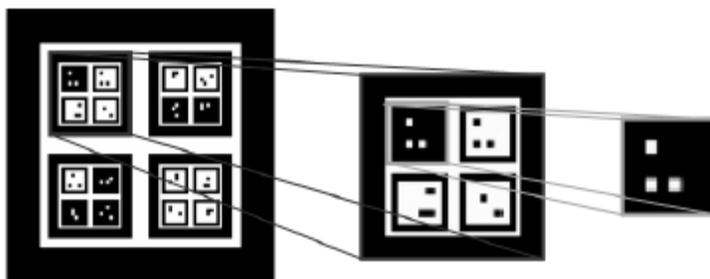
**Figura 23.** Marcas a color propuestas por [Lee08]



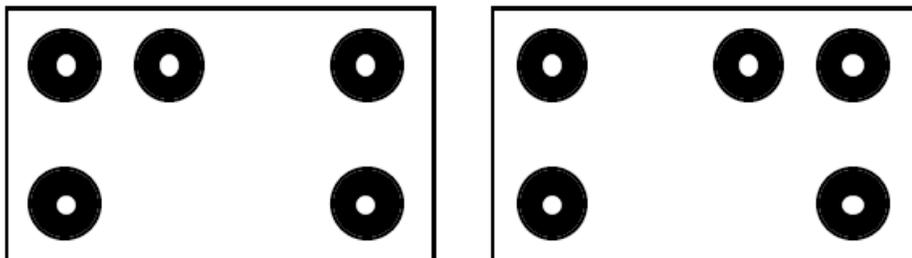
Como se anotó anteriormente, la identificación de bordes es un paso intermedio para la extracción de características (vértices del cuadrado) utilizadas en el proceso de seguimiento. Sin embargo, si la marca está parcialmente ocluida, los bordes, y por lo tanto, las características, no serán detectadas, lo cual lo convierte en una de las mayores limitaciones del uso de marcas artificiales en la escena. [Tat07] propone abordar tal limitación sin la necesidad de apoyarse en la detección de marcas adicionales en la escena como se propone en [Uml02]. Si se considera que la precisión del registro depende de la información que sea capturada de la marca, el tamaño y diseño de la misma requieren especial atención. Basado en esto, el autor propone una estructura jerárquica consistente en varias marcas anidadas (ver Figura 24) que permite una calibración de la cámara mucho más precisa. Este tipo de diseño resulta ser apropiado para observar características detalladas de los modelos 3-D virtuales, al disminuir la distancia de la marca y la cámara.

En la literatura es posible encontrar otros trabajos que corroboran la tesis anterior apoyando el diseño cuidadoso de la marca, como el presentado en [Hof96], en donde los autores proponen marcas basadas en círculos binarios concéntricos, los cuales por su forma especial (ver Figura 25) permiten determinar con mayor facilidad la posición y orientación, obteniéndose una mayor precisión en la identificación y por tanto en el registro. A diferencia del enfoque anterior propuesto por [Tat07], se plantea el uso de varias marcas individuales, por lo que este trabajo presenta como limitación mantener todas las marcas dentro del campo de visión de la cámara del HMD.

**Figura 24.** Marcas jerárquicas anidadas propuestas por [Tat07]



**Figura 25.** Marcas binarias propuestas por [Hof96]



En los últimos cinco años se han refinado cada vez más las técnicas de registro sobre marcas a la vez que surgen nuevas aplicaciones. En [Jua05] se propone un sistema de realidad aumentada multimodal para el tratamiento de fobias a insectos pequeños como cucarachas y arañas. En el artículo se destaca el uso de la RA frente a la RV por el sentimiento de presencia y la interacción en un ambiente real. En una segunda versión de este trabajo [Jua06] las marcas binarias son reemplazadas por marcas diseñadas con tintas especiales visibles a una cámara infrarroja.

En [Jo08] se presenta un novedoso sistema denominado ARForce mediante el cual se usan marcas no sólo para la determinación de la *pose* del usuario sino para calcular vectores de fuerza que permiten manipular los objetos virtuales mediante movimientos de los dedos; logrando de esta manera una interfaz mucho más natural acercándose a manipular los objetos virtuales como se hace con los reales. Para tal fin, en este proyecto se utiliza un sensor consistente de un cuerpo elástico transparente compuesto de dos capas de marcas de colores, el cual puede ser manipulado con seis grados de libertad (Figura 26), a diferencia de trabajos anteriores que sólo permitía movimientos en un plano horizontal [Sat07][Kam05]. Al igual que en el enfoque anterior la información de posicionamiento es capturada mediante una cámara infrarroja.

**Figura 26.** Marcas utilizadas para detectar fuerza (a) Prototipo inicial [Kam05][Sat07] (b) Marca propuesta por [Jo08]



En general, de los trabajos reportados en la literatura se evidencia la ventaja, en términos de precisión en el registro, del uso de marcas artificiales atribuido por la simplificación del proceso de medición de la posición y orientación del usuario. Sin embargo, con el costo asociado a la pérdida de practicidad al requerir modificar el ambiente real pre-ubicando marcas para la detección. Adicionalmente, aspectos como posibles oclusiones y falsos positivos en la detección deberán ser controlados y evaluados para evitar fallas en el sistema.

Para superar las limitaciones de la ubicación de marcas en el mundo real, denominado en [Sch05] como contaminación visual del ambiente, se ha propuesto como enfoque

---

alternativo capturar la información de pose a partir de los rasgos propios de la escena, extrayendo información directamente de la imagen.

## **2.2 Extracción de características naturales– enfoque markerless**

El registro basado en visión considerando un enfoque que no utiliza marcas, aborda la identificación de características representativas que brinden información sobre el contenido y comportamiento de la escena. Este enfoque plantea una serie de retos asociados a ¿Qué identificar como característica representativa? ¿Cómo seguir dichas características a lo largo de toda la secuencia de video? ¿En qué posición superponer los objetos virtuales? ¿A partir de qué información calcular la pose de la cámara? Todo esto ha llevado a que la extracción de características se convierta en una aproximación mucho más compleja y muchos más esfuerzos e investigaciones habrán de realizarse para obtener la precisión que se lograría con una aproximación basada en marcas o con un registro basado en sensores. A pesar de ello, esta metodología ha encabezado los estudios recientes en RA por sus potenciales aplicaciones y practicidad.

En el campo de visión por computador, trabajos interesantes sobre métodos efectivos de extracción de características han sido propuestos [Low87][Dho89][DeM93][Kum94] aplicados principalmente a recuperación de estructuras 3-D, seguimiento, correspondencia de imágenes y reconocimiento de escenas u objetos. Para tales fines, se han extraído como características representativas elipses, círculos, contornos y segmentos de línea. Sin embargo, aunque reportan resultados eficientes en la identificación, muchas de estas técnicas no han sido ampliamente implementadas en aplicaciones de realidad aumentada debido a que presentan como gran desventaja su desempeño en tiempo real. Es por ello, que además de la efectividad, la eficiencia, asociada al tiempo de ejecución, se vuelve un factor crítico en la selección del método de extracción de características.

En 1999, Neumann [Neu99] publica uno de los primeros trabajos de los cuales se tenga registro sobre el proceso de seguimiento a partir de características propias de la escena para aplicaciones de realidad aumentada. El método propuesto consta de una configuración a ciclo cerrado en donde se identifican como características

representativas puntos y regiones para posiciones desconocidas a priori. Neumann argumenta, tal como se afirma en la presente tesis, la fuerte relación entre la extracción de características y el seguimiento, al asegurar que para garantizar la precisión, las características deben ser estables y confiables durante toda la secuencia de video, por lo que opta por detectar los puntos a partir de la distribución de los valores propios de una matriz de covarianza. A su vez, las regiones son definidas a partir de ventanas 31x31 píxeles que contienen la mayor cantidad de puntos característicos, incluyéndoles un modelo afín para tolerar posibles distorsiones en el proceso de seguimiento.

Otros autores por su parte, le han apuntado al reconocimiento de características más elaboradas, entre las que se encuentran planos, bordes, líneas, elipses y cilindros [Sim00][Dru02][Com03] [Skr04]. Recientemente, se ha propuesto como característica distintiva los descriptores de forma, entre los que se destaca SIFT [Yua06][Azz08][Lee08a][Lee09a][Bre09]. Aunque con una gran variedad de posibles características a detectar, el trabajo de Neumann inspiró la aparición de muchas publicaciones posteriores [Moo09][Pan06][Fre06][Yua05][Fer01] en donde todos concuerdan en el uso de puntos como característica representativa. Estos puntos sin embargo, deben cumplir características especiales tal que garanticen la estabilidad y confiabilidad a la que refiere Neumann, tal como lo propone Shi y Tomasi [Shi94].

### **2.2.1 Extractor de puntos Shi-Tomasi**

En 1993, inspirado en el trabajo de Lucas-Kanade [Luc81], Jianbo Shi y Carlo Tomasi proponen uno de los criterios más ampliamente utilizados para extraer puntos para propósitos de seguimiento. El método propuesto subyace en establecer una estrecha relación entre las características a extraer y el seguimiento. Los autores proponen que no siempre una región con una alta información de textura es un buen indicador de presencia de características representativas, por lo que propone cuantificar una medida de disimilitud que relacione la apariencia de una característica en particular en el primer cuadro de la secuencia de video y su apariencia de cualquier cuadro posterior; para ello, demuestran que un modelo solamente traslacional como el propuesto por Lucas-Kanade [Luc81] no define completamente el movimiento de una característica (incluyendo su

deformación), por lo que proponen representar dicha disimilitud a partir de dos modelos: uno traslacional y otro afín.

### Modelos de desplazamiento

Los cambios de intensidad de dos cuadros de una secuencia de video (de ahora en adelante considerados como imágenes) separados por un tiempo  $\tau$  está dado por el desplazamiento de los puntos  $x = (x, y)$ , referidos como la cantidad de movimiento  $\delta = (\xi, \eta)$ , así

$$I(x, y, t + \tau) = I(x - \xi(x, y, t, \tau), y - \eta(x, y, t, \tau)) \quad (1)$$

representando un modelo traslacional. Sin embargo, una representación mucho más realista del modelo incluirá un campo de movimiento afín que permita modelar deformaciones, causadas principalmente por rotaciones

$$\delta = A x + d \quad (2)$$

donde

$$A = \begin{bmatrix} d_{xx} & d_{xy} \\ d_{yx} & d_{yy} \end{bmatrix} \quad (3)$$

Considerando por lo tanto una deformación  $D$  de los puntos  $x$  junto con una traslación  $d$  respecto al centro de una ventana que contiene  $x$ . Los seis parámetros de  $A$  y  $d$  deberán ser calculados en el proceso de seguimiento y su calidad dependerán de la riqueza de la textura de la imagen, el tamaño de la ventana y la distancia entre las dos imágenes. Dichos parámetros deberán ser calculados tal que minimicen la ecuación (4) sobre una ventana  $W$  para un par de imágenes  $I$  y  $J$ .

$$\epsilon = \int_W [J(Ax + d) - I(x)]^2 w(x) dx \quad (4)$$

donde  $w(x)$  es una función de peso asignada usualmente a 1.

Para hallar tal minimización se hace necesario derivar respecto a A y d, y posteriormente linealizar el sistema resultante mediante una expansión de Taylor (para mayor detalle ver [Shi93]), resultando en

$$T z = a \quad (5)$$

donde

$$z^T = [d_{xx} \ d_{yx} \ d_{xy} \ d_{yy} \ d_x \ d_y] \quad (6)$$

a es el vector de error representando la diferencia entre las dos imágenes.

$$a = \iint_W [I(x) - J(x)] \begin{bmatrix} x g_x \\ x g_y \\ y g_x \\ y g_y \\ g_x \\ g_y \end{bmatrix} w \, dx \quad (7)$$

y

$$T = \iint_W \begin{bmatrix} U & V \\ V^T & Z \end{bmatrix} w \, dx \quad (8)$$

con

$$U = \begin{bmatrix} x^2 g_x^2 & x^2 g_x g_y & x y g_x^2 & x y g_x g_y \\ x^2 g_x g_y & x^2 g_y^2 & x y g_x g_y & x y g_y^2 \\ x y g_x^2 & x y g_x g_y & y^2 g_x^2 & y^2 g_x g_y \\ x y g_x g_y & x y g_y^2 & y^2 g_x g_y & y^2 g_y^2 \end{bmatrix} \quad (9)$$

$$V^T = \begin{bmatrix} x g_x^2 & x g_x g_y & y g_x^2 & y g_x g_y \\ x g_x g_y & x g_y^2 & y g_x g_y & y g_y^2 \end{bmatrix} \quad (10)$$

$$Z = \begin{bmatrix} g_x^2 & g_x g_y \\ g_x g_y & g_y^2 \end{bmatrix} \quad (11)$$

Debido a que durante el seguimiento la deformación D es tan pequeña y errores en el cálculo de ésta conllevarían a errores en el cálculo del modelo traslacional, sólo se considera el modelo de desplazamiento traslacional en el seguimiento y el modelo afín es

resuelto para comparar disimilitudes entre el primer cuadro y otro en un momento  $t + \tau$  para una ventana en particular. Por lo que el modelo se reduce a

$$Z d = e \quad (12)$$

donde  $e$  contiene los dos últimos componentes del vector  $a$ .

Las características representativas serán obtenidas a partir de la matriz  $Z$  bajo las siguientes premisas:

1. Dos valores propios pequeños representan una intensidad constante en la ventana. No se garantiza que la matriz  $Z$  este por encima de los límites del ruido de la imagen.
2. Un valor propio grande y otro pequeño representan un patrón con textura unidireccional.
3. Dos valores propios grandes representan la presencia de patrones apropiados para ser seguidos, tales como esquinas.

Si el menor valor propio es mayor a un umbral predefinido, se dirá que la matriz  $Z$  está bien condicionada y por encima de los niveles de ruido de la imagen. En conclusión, una ventana de características representativas será aceptada si:

$$\min(\lambda_1, \lambda_2) > \lambda \quad (13)$$

con  $\lambda_1, \lambda_2$  los valores propios de  $Z$  y  $\lambda$  un umbral predefinido.

Más de 15 años de trabajos reportando los resultados obtenidos utilizando los tres criterios anteriormente descritos confirman la efectividad y eficacia del método, no sólo en aplicaciones de realidad aumentada sino en un amplio campo de aplicaciones que incluyen coincidencia de imágenes, localización, mapeado, recuperación y análisis de imágenes; superando y refinando aproximaciones iniciales como las propuestas por Moravec[Mor79], [För87] y Harris[Har88].

En la presente tesis, apoyada en la premisa que una buena característica es aquella que permite ser seguida correctamente, se define como característica representativa, puntos, los cuales son extraídos utilizando el método Shi-Tomasi descrito anteriormente.

## Conclusión

Debido a las limitaciones explícitas del registro utilizando marcas, los trabajos actuales en esta línea apuntan en su mayoría a mejorar la robustez de tal aproximación, específicamente en la identificación de marcas frente a cambios y oclusiones en la escena, tal que conlleven a concentrarse en las características que deberían identificarse en la marca. En general, futuras investigaciones deberán direccionarse hacia la coherencia en la alineación real-virtual considerando variaciones en las condiciones del ambiente real, incluyendo robustez a oclusión, cambios de las condiciones de la iluminación y seguimiento.

La revisión de la literatura evidencia que a pesar de los esfuerzos de la comunidad y la multitud de trabajos adelantados en torno al registro basado en marcas, enfocándose cada vez más en el diseño cuidadoso de la misma, su aplicación a espacios restringidos, oclusión o su ubicación por fuera del campo de visión conllevan a que la tendencia apunte al registro utilizando la menor cantidad de información a priori como sea posible; valiéndose para ello de técnicas de extracción de características, las cuales aunque numerosas se ven restringidas en su desempeño (asociado al tiempo) y su capacidad de ser seguidas correctamente a lo largo de la secuencia de video. Se considerará, por lo tanto, que la identificación apropiada de las características representativas será el primer paso para garantizar el correcto alineamiento en la etapa de registro.

## **3. Capítulo 3**

### **SEGUIMIENTO DE CARACTERÍSTICAS**

Debido a que el registro propuesto en la presente tesis está basado sólo en técnicas de visión por computador, se acudirá a la definición de Neumann [Neu99], quien define el término seguimiento en dos sentidos, el primero asociado al proceso mediante el cual se obtienen mediciones de pose con 6 grados de libertad (6 DOF), el segundo enfocado hacia la correspondencia entre características 2-D en secuencias de imágenes. Lo anterior indica que el seguimiento presenta dos componentes principales, a saber: modelación e identificación; éste último en estrecha relación con la extracción de características descritas en el capítulo anterior.

En el presente capítulo se hará referencia a algunas técnicas tradicionalmente utilizadas para el seguimiento, para posteriormente describir el método implementado para detectar correspondencias en las características representativas seleccionadas. Finalmente, se aborda la determinación de la pose.

### **3.1 El seguimiento en sistemas de realidad aumentada**

Estimar el campo de desplazamiento entre dos imágenes no está restringido a la estimación del movimiento, otras aplicaciones incluyen reconstrucción 3-D y el registro de imágenes, como es el caso de los sistemas de realidad aumentada basados en visión, ya que para esta aplicación en particular, determinar el desplazamiento no es un fin en sí

mismo mas se convierte en una etapa intermedia para el ajuste de la información real y virtual llevada a cabo en el registro.

Como ha sido descrito a lo largo del documento, el seguimiento es otro de los factores fundamentales para asegurar la precisión en el registro. Han sido varios los autores que atribuyen el error del registro a la obtención incorrecta de las posiciones en las cuales se deben superponer los objetos virtuales [Tat07][Hol95][Sta96], dichas posiciones deberán ser obtenidas de los puntos extraídos en la etapa inicial del registro y seguidos en la secuencia de video.

Aunque el flujo óptico ha sido la técnica más ampliamente utilizada para estos propósitos, otras aproximaciones han sido propuestas, tal es el caso de las denominadas partículas [San08], las cuales exhiben propiedades similares a las características representativas descritas en el capítulo anterior, probando un buena efectividad en video con geometrías de escenas complejas, varios tipos de oclusiones, regiones con poca textura y deformaciones no rígidas. Sin embargo, ésta técnica se concentra en la robustez frente a la oclusión, por lo que el método suele fallar frente a cambios en la apariencia de una partícula, adicionalmente, su tiempo de ejecución es de más de 100 segundos para cada par de imágenes, lo cual no es tolerable para una aplicación de realidad aumentada, ya que conllevaría a una gran cantidad de errores dinámicos en el registro, como ha sido mencionado en el capítulo 1.

### **3.1.1 Flujo óptico**

El método de flujo óptico como estrategia para medir la velocidad fue primero propuesto en 1950 por Gibson [Gib50][Gib66][Gib77]. Sin embargo, no fue sino hasta 1981 que Horn y Schunck [Hor81] formalizaron su formulación matemática proponiendo un método que incluía la restricción de brillo constante (válida en la actualidad) que conllevaría a resolver los componentes de la velocidad en dos dimensiones.

El flujo óptico es probablemente la técnica más utilizada para evaluar el movimiento entre dos imágenes (o una secuencia de imágenes) utilizando para ello, sólo características propias de éstas. Este problema puede ser redefinido como el problema análogo de encontrar un vector velocidad (o su desplazamiento) tal que represente la distancia de un

píxel entre un par de imágenes. En este punto existen dos aproximaciones propuestas en la literatura, a saber: flujo óptico denso y escaso, cada una con sus ventajas y desventajas. Mientras el método denso utiliza mayor cantidad de puntos, haciéndolo más robusto a oclusiones, el método escaso es capaz de tolerar grandes movimientos de la cámara.

### **Flujo óptico denso**

El flujo óptico denso considera el seguimiento de todos los píxeles de la imagen, entre los métodos más conocidos se encuentra el de Horn y Schunck [Hor81]. Para aplicaciones de RA, de los trabajos reportados no se ha encontrado evidencia del uso de este tipo de técnicas ya que como el lector podrá inferir implica un tiempo computacional considerable que irá en contravía con el requisito de desempeño en tiempo real que quieren los sistemas de RA. Por otra parte, será mucho más complejo el seguimiento de regiones que presenten poca textura; si se considera por ejemplo, una región con una intensidad constante en un tiempo  $t$ , esta misma región podrá coincidir (considerando un pequeño desplazamiento) con otra región de intensidad constante y la correspondencia entre los píxeles de ambas regiones será difícil de determinar, teniendo que hacer uso de métodos de interpolación que implicarán un tiempo computacional adicional.

### **Flujo óptico escaso**

El principio de los métodos de flujo óptico escaso subyace en especificar un conjunto de puntos que serán seguidos en la secuencia de imágenes. Esta aproximación será la utilizada en el presente tesis considerando el conjunto de puntos extraídos en la etapa anterior, para ello se implementa un método de seguimiento piramidal como se describe a continuación.

El método piramidal implementado trabaja utilizando información local de la imagen, considerada en una ventana que encierra una característica representativa, la cual es capaz de tolerar posibles movimientos por fuera de la ventana en inspección (debido a grandes desplazamientos de la cámara).

Los métodos de seguimiento escaso se basan en tres supuestos fundamentales que soportan el método:

1. *Consistencia en el brillo*: un píxel de una imagen no cambia su apariencia al moverse de una imagen a la siguiente. Para una imagen a escala de grises se supondrá que

$$I(x + u, y + v, t + 1) = I(x, y, t) \quad (14)$$

por lo que

$$\frac{\partial f(x)}{\partial t} = 0 \quad (15)$$

2. *Persistencia temporal*: el movimiento de la imagen en una ventana no cambia rápidamente en el tiempo, lo cual significa que los objetos no se mueven mucho de imagen a imagen. Considerando la ecuación (14) se obtendría

$$\frac{\partial I}{\partial x} \Big|_t \left( \frac{\partial x}{\partial t} \right) + \frac{\partial I}{\partial t} \Big|_{x(t)} = 0 \quad (16)$$

lo cual será válido para el caso 1-D. Adicionalmente, deberá considerarse que los dos supuestos anteriores no siempre se cumplen completamente, sin embargo si se está lo suficientemente cerca de la solución se podrá iterar (por ejemplo, con el método de Newton) hasta llegar a una velocidad que sea lo más exacta posible, considerando la velocidad estimada como punto de partida para la siguiente iteración.

Para el caso bidimensional se tendrá una componente  $v$  y otra  $u$  para el vector velocidad, representando el componente  $y$  y  $x$  respectivamente.

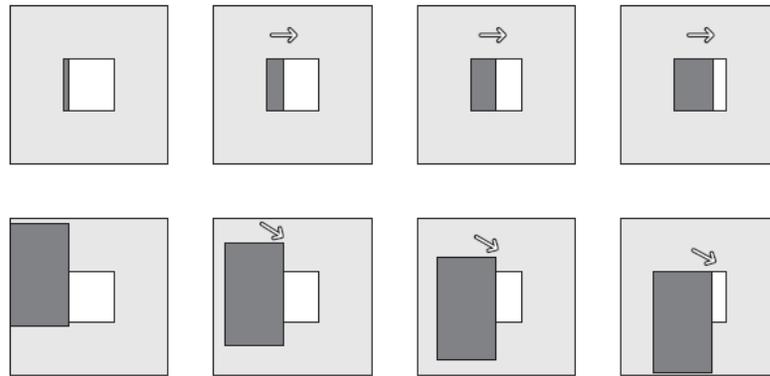
$$I_x u + I_y v + I_t = 0 \quad (17)$$

3. *Coherencia espacial*: puntos vecinos en una escena pertenecientes a la misma superficie, tienen movimiento similar y se proyectan en puntos cercanos al plano de la imagen.

Con estos tres supuestos, es fácil determinar que para cada píxel en la ventana, se tienen dos variables desconocidas, por lo que no será posible obtener una sola solución

para el movimiento bidimensional, éste es conocido como el *problema de apertura*; el cual afirma que cuando se desea calcular el movimiento en una ventana, con frecuencia se ven sólo bordes no esquinas (ver Figura 27), lo cual es insuficiente para determinar exactamente la dirección de movimiento del objeto.

**Figura 27.** El problema de apertura en el seguimiento [Gar08]



Para superar esta limitación, se hace uso del último supuesto, ya que si todos los píxeles cercanos al píxel central se mueven coherentemente a éste, podrán utilizarse las ecuaciones que describen los movimientos de todos los puntos en la ventana. Obteniéndose por tanto el sistema de ecuaciones siguiente (para el caso de una ventana 5 x5)

$$\left[ \begin{array}{ccc} I_x(p_1) & \cdots & I_y(p_1) \\ \vdots & \ddots & \vdots \\ I_x(p_{25}) & \cdots & I_y(p_{25}) \end{array} \right]_{25 \times 2} \begin{bmatrix} u \\ v \end{bmatrix}_{2 \times 1} = \begin{bmatrix} I_t(p_1) \\ \vdots \\ I_t(p_{25}) \end{bmatrix}_{25 \times 1} \quad (18)$$

o equivalentemente

$$Ad = b \quad (19)$$

Al resolver este sistema con un método de minimización de mínimos cuadrados se obtiene

$$A^T A \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} \sum I_x I_t \\ \sum I_y I_t \end{bmatrix} \equiv A^T A \begin{bmatrix} u \\ v \end{bmatrix} = -A^T b \quad (20)$$

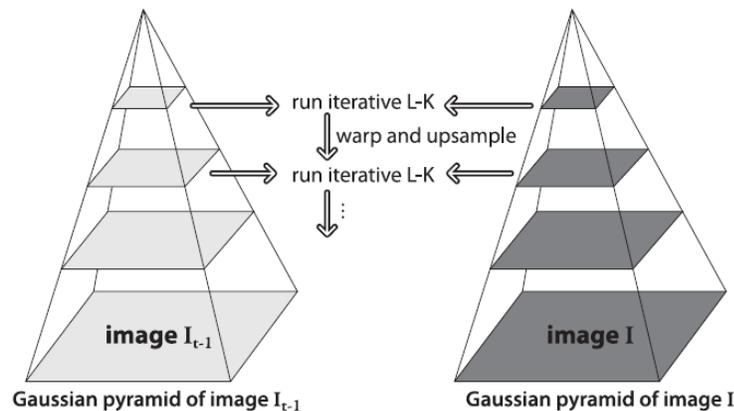
Las componentes del movimiento en x y y serán

$$\begin{bmatrix} u \\ v \end{bmatrix} = (A^T A)^{-1} A^T b \quad (21)$$

lo cual será válido si la matriz A es invertible.

Como se anotó anteriormente, la implementación de un método piramidal aumenta la confiabilidad del flujo óptico al soportar grandes desplazamientos, refinando la aproximación inicial arrojada por el método anteriormente descrito, comenzando desde los niveles más altos hasta el final de la pirámide, minimizando las violaciones a los tres supuestos básicos del flujo óptico.

**Figura 28.** Seguimiento piramidal [Gar08]



Si bien el método descrito anteriormente soporta modelar tanto el movimiento relativo de los objetos como del observador (se considerará que el observador se encuentra en la misma posición donde está la cámara), brindando información importante sobre la ubicación espacial de los objetos, es importante resaltar que un método de flujo óptico no modela el movimiento tridimensional de los objetos, sólo los cambios en la imagen que resultan de este movimiento. Considerando esto y el hecho que los objetos virtuales deberán ser ubicados espacialmente (3-D), algún método deberá considerarse para

obtener estas coordenadas tridimensionales, se podrá decir por lo tanto que la segunda y tercera causa de un registro incoherente pueden ser atribuidas respectivamente a un seguimiento incorrecto de las características representativas de la imagen y a una determinación poco precisa de la información 3-D a partir de sus proyecciones bidimensionales.

### **3.2 Determinación de la pose**

La pose es entendida como el cálculo de la posición y orientación de un objeto en la imagen, o análogamente la posición y orientación de la cámara frente a la escena. Tal como propone Schall [Sch05] el uso de marcas para propósitos de seguimiento es solamente pragmático, un tracker basado en visión será más apropiado para RA. El uso de marcas no solamente facilita la identificación y seguimiento de características sino también la determinación de su pose. Otra de las técnicas que han sido ampliamente reportadas en la literatura que combina la filosofía del uso de marcas con la practicidad de los rasgos característicos naturales en la escena, es el propuesto mediante el seguimiento basado en modelos 3-D [Bre09][Yu07][Com03][Pas99]; técnica consistente en comparar las posiciones de los rasgos característicos estimados con las posiciones reales obtenidas en un modelo tridimensional pre-construido de la escena, refinando iterativamente la medición de la pose. Esta técnica si bien permite describir y predecir movimientos frente a oclusión obteniendo una determinación mucho más precisa de la pose, comparte la desventaja de la falta de practicidad, atribuida a la cantidad de información a priori necesaria, para implementar tal aproximación.

En general, no hay manera de obtener la geometría 3-D de un objeto a partir de una imagen, por lo que habrá que recurrir a extraer la información a partir de múltiples vistas, para ello varias técnicas han sido propuestas. Enfoques numéricos e iterativos, entre los que se encuentran métodos lineales como mínimos cuadrados, métodos no lineales como virtual visual servoing [Com03] o iterativos como Newton-Raphson [Uen94][Fis03], Levenberg-Marquardt o POSIT (Pose from Orthography and Scaling with Iteration) [Vac04][Yua06]; sin dejar de lado los enfoques geométricos.

Después de la correspondencia de puntos, la pose puede ser calculada. Sin embargo, para ello será necesario el cálculo de las coordenadas 3-D de las características, y por

supuesto, de los puntos sobre los cuales se superpondrán los objetos virtuales. Para ello, se propone en la presente tesis el uso de propiedades afín, ampliamente reportadas en la literatura por su capacidad de reconstrucción y simplicidad en la implementación.

### 3.2.1 Cálculo de la información tridimensional

El sistema coordinado afín para propósitos de RA ha sido reportado en [Koe91] [Kut98][Wei95][Pan06] con dos propiedades distintivas, a saber: reproyección y reconstrucción.

Considerando un espacio afín definido a partir de cuatro puntos no coplanares  $\{P_0, P_1, P_2, P_3\} \in R^3$  ( $P_0$  representado el centro del espacio afín y los otros tres puntos, las bases) y un conjunto de características extraídas en la etapa anterior al seguimiento, las coordenadas tridimensionales  $[X \ Y \ Z]^T$  del conjunto de características pueden ser determinadas en el espacio afín si se tienen la proyección de cada punto en al menos dos vistas diferentes ( $[u^j \ v^j]^T$ ;  $j = 1, 2$ ), al igual que las proyecciones de los cuatro puntos que definen el espacio afín; ésta es conocida como la propiedad de reconstrucción, esto es

$$\begin{bmatrix} u^1 \\ v^1 \\ u^2 \\ v^2 \end{bmatrix} = \begin{bmatrix} u_{b1}^1 - u_{bo}^1 & u_{b2}^1 - u_{bo}^1 & u_{b3}^1 - u_{bo}^1 & u_{bo}^1 \\ v_{b1}^1 - v_{bo}^1 & v_{b2}^1 - v_{bo}^1 & v_{b3}^1 - v_{bo}^1 & v_{bo}^1 \\ u_{b1}^2 - u_{bo}^2 & u_{b2}^2 - u_{bo}^2 & u_{b3}^2 - u_{bo}^2 & u_{bo}^2 \\ v_{b1}^2 - v_{bo}^2 & v_{b2}^2 - v_{bo}^2 & v_{b3}^2 - v_{bo}^2 & v_{bo}^2 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (22)$$

con  $[u_{bi}^j \ v_{bi}^j]^T$  ( $j = 1, 2$ ;  $i = 0, 1, 2, 3$ ) las proyecciones del origen afín  $P_0$  y los puntos bases  $P_1, P_2$  y  $P_3$  en las dos imágenes de control.

Por su parte, la propiedad de reproyección permite calcula la proyección  $[u \ v]$  en una imagen  $k$  de un punto a partir de su coordenada 3D afín y la proyección de la base y el origen afín.

$$\begin{bmatrix} u_k \\ v_k \\ 1 \end{bmatrix} = \begin{bmatrix} u_{b1} - u_{bo} & u_{b2} - u_{bo} & u_{b3} - u_{bo} & u_{bo} \\ v_{b1} - v_{bo} & v_{b2} - v_{bo} & v_{b3} - v_{bo} & v_{bo} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_k \\ Y_k \\ Z_k \\ 1 \end{bmatrix} \quad (23)$$

Como ya el lector habrá notado, una característica especial del registro a partir de características representativas consiste en que si bien no se tiene información a priori de la escena, tampoco se tendrá información sobre la posición en la cual se superpondrán los objetos virtuales. Es allí, donde se requerirá la intervención explícita del usuario para señalar dicha posición. Varios trabajos han sido presentados en esta línea [Bre09][Rei07][Pan06], en donde el enfoque general gira en torno a la selección manual de puntos que definen el espacio sobre el cual se renderizan los objetos virtuales. Dichos puntos difieren de los cuatro puntos previamente seleccionados para definir el espacio afín. En la presente tesis, el método propuesto para la corrección del registro asume que los objetos virtuales se superpondrán sobre superficies planas; por lo que, la restricción de no coplanaridad de las bases y el origen afín implicará que estos mismos puntos no pueden ser utilizados para tales propósitos. Lo anterior, por otra parte, no se convierte en una desventaja del método, por el contrario lo hará más robusto frente a posibles oclusiones, ya que en el caso de una oclusión de algunos de los puntos sobre los cuales se superponen los objetos virtuales, se podrá hallar sus coordenadas estimadas a partir de la propiedad de reproyección. Similar al trabajo de [Pan06] sólo será necesario que el usuario defina cuatro puntos, los cuales conformarán un espacio euclidiano, en lugar de seis, cinco u ocho como reportado en otras aproximaciones [Yua04][Seo00].

En este punto será necesario por lo tanto, definir apropiadamente las bases y el origen del sistema coordenado afín, para la cual existen dos aproximaciones, la primera implicando nuevamente la intervención explícita del usuario que seleccione cuatro puntos que sean visualmente no coplanares; la segunda, detectar dichos puntos de manera automática. En la presente tesis se ha optado por la segunda aproximación, la cual plantea un cuestionamiento adicional: por definición todos los puntos en una imagen tienen la misma profundidad, ¿Cómo determinar puntos no coplanares a partir de información 2-D? este es un problema mal condicionado cuando los objetos son distantes respecto al tamaño. En [Tom92] se propone un método de factorización capaz de recuperar la forma y el movimiento sin la necesidad de calcular la profundidad como paso

intermedio; en lugar de ello, se propone calcular una matriz  $A_{4 \times n}$  centrada en el centro de masa del vector de rasgos característicos y calcular su rango.

Basados en la propiedad de reproyección, es posible afirmar que el centro de masa de un conjunto de puntos afín  $\{P_1, P_2, \dots, P_n\} \in R^3$ , denotado  $P_c [X_c Y_c Z_c 1]^T$ , es proyectado en el centro de las proyecciones de estos puntos, denotado  $p_c [u_c v_c]^T \in R^2$ , es decir  $p_c$  representa la proyección de  $P_c$ . En consecuencia, dadas las proyecciones de los puntos característicos en dos imágenes, la matriz  $A_{4 \times n}$  puede ser obtenida mediante la ecuación 24.

$$A_{4 \times n} = \begin{bmatrix} u_{p_1}^1 - u_c^1 & \dots & u_{p_n}^1 - u_c^1 \\ v_{p_1}^1 - v_c^1 & \dots & v_{p_n}^1 - v_c^1 \\ u_{p_1}^2 - u_c^2 & \dots & u_{p_n}^2 - u_c^2 \\ v_{p_1}^2 - v_c^2 & \dots & v_{p_n}^2 - v_c^2 \end{bmatrix} \quad (24)$$

donde  $[u_{p_i}^j, v_{p_i}^j]^T$  ( $j = 1, 2; i = 1, 2, \dots, n$ ) representa la proyección de los puntos 3-D afín y  $[u_c^j, v_c^j]^T$  ( $j = 1, 2$ ) es el centro de masa de estas proyecciones. Las bases del espacio afín serán obtenidas a partir del cálculo de las bases de los tres componentes linealmente independientes de la matriz  $A$ , para ello se hace uso de la descomposición SVD, mediante la cual se obtiene:

$$A_{4 \times n} = U_{4 \times 4} D_{4 \times n} V_{n \times n}^T \quad (25)$$

Siendo  $U_{4 \times 3}$ ,  $D_{3 \times 3}$  y  $V_{n \times 3}$  las submatrices superiores  $4 \times 3$ ,  $3 \times 3$  y  $n \times 3$  de  $U$ ,  $D$ , y  $V$  respectivamente, las matrices de reproyección y reconstrucción pueden ser reformuladas mediante las ecuaciones 26 y 27

$$\begin{bmatrix} u_{b_1}^1 - u_c^1 & u_{b_2}^1 - u_c^1 & u_{b_3}^1 - u_c^1 \\ v_{b_1}^1 - v_c^1 & v_{b_2}^1 - v_c^1 & v_{b_3}^1 - v_c^1 \\ u_{b_1}^2 - u_c^2 & u_{b_2}^2 - u_c^2 & u_{b_3}^2 - u_c^2 \\ v_{b_1}^2 - v_c^2 & v_{b_2}^2 - v_c^2 & v_{b_3}^2 - v_c^2 \end{bmatrix} = U_{4 \times 3} D_{3 \times 3}^{1/2} \quad (26)$$

$$M_{3 \times 4}^k = \begin{bmatrix} u_{b_1}^k - u_c^k & u_{b_2}^k - u_c^k & u_{b_3}^k - u_c^k & u_c^k \\ v_{b_1}^k - v_c^k & v_{b_2}^k - v_c^k & v_{b_3}^k - v_c^k & v_c^k \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (27)$$

donde  $[u_{b_i}^k, v_{b_i}^k]^T$  ( $k = 1, 2; i = 1, 2, 3$ ) son las proyecciones de los puntos base afín del conjunto  $\{P_1, P_2, \dots, P_n\}$  y  $[u_c^k, v_c^k]^T$  representa la proyección del origen afín (la

proyección del centroide del conjunto de puntos 3-D afín). Finalmente, los puntos 3-D afín serán obtenidos con el cálculo de la ecuación 28

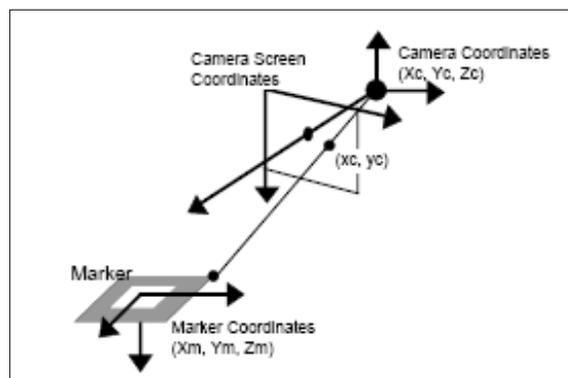
$$[P_1 \quad \dots \quad P_n] = \begin{bmatrix} (D_{3 \times 3})^{1/2} (V_{n \times 3})^T \\ 1 \end{bmatrix} \quad (28)$$

Una vez se ha obtenido la estimación del posicionamiento en el espacio afín de los puntos que forman el espacio euclidiano utilizando para ello los cuatro puntos definidos manualmente por el usuario, se procede a calcular la pose, la cual basada en la suposición que dicho espacio fue definido en una región plana, se implementa utilizando el método geométrico propuesto en [Kat99].

Mediante procesamiento digital de imágenes, Kato [Kat99] propone relacionar las coordenadas de la cámara con las de la marca, que en este caso están definidas mediante el espacio euclidiano definido por el usuario (ver Figura 29), por medio de la ecuación 29.

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} V_{11} & V_{12} & V_{13} & W_x \\ V_{21} & V_{22} & V_{23} & W_y \\ V_{31} & V_{32} & V_{33} & W_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_m \\ Y_m \\ Z_m \\ 1 \end{bmatrix} \equiv \begin{bmatrix} 0 & V_{3 \times 3} & 0 & W_{3 \times 3} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_m \\ Y_m \\ Z_m \\ 1 \end{bmatrix} \equiv T_{cm} \begin{bmatrix} X_m \\ Y_m \\ Z_m \\ 1 \end{bmatrix} \quad (29)$$

**Figura 29.** Relación entre el sistema coordenado de la cámara y el sistema coordenado euclidiano [Kat99]



Dos líneas paralelas en el cuadrado que conforma en espacio euclidiano estarán representadas por las ecuaciones 30 y 31. Considerando, que se tiene una cámara calibrada, lo cual es una suposición válida en la mayoría de los sistemas de RA, se conocerán sus parámetros intrínsecos representados en la matriz P de la ecuación 32.

$$a_1x + b_1y + c_1 = 0 \quad (30)$$

$$a_2x + b_2y + c_2 = 0 \quad (31)$$

$$\begin{bmatrix} hx_c \\ hy_c \\ h \\ 1 \end{bmatrix} = \begin{bmatrix} P_{11} & P_{12} & P_{13} & 0 \\ 0 & P_{22} & P_{23} & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} \quad (32)$$

Reemplazando  $x_c$ ,  $y_c$  y  $h$  de la ecuación 32 en las ecuaciones 30 y 31, se podrán representar los planos que contienen estas líneas, en el sistema coordenado de la cámara mediante las ecuaciones 33 y 34.

$$a_1P_{11}X_c + (a_1P_{12} + b_1P_{22})Y_c + (a_1P_{13} + b_1P_{23} + c_1)Z_c = 0 \quad (33)$$

$$a_2P_{11}X_c + (a_2P_{12} + b_2P_{22})Y_c + (a_2P_{13} + b_2P_{23} + c_2)Z_c = 0 \quad (34)$$

## Conclusión

Relacionando los objetivos propuestos en la presente tesis con el método descrito anteriormente se observa que se persigue una coherencia entre todas las etapas del proceso, pretendiendo reducir, si bien no es posible eliminar, los errores en la identificación y seguimiento. El método propuesto por Shi-Tomasi resulta ser un excelente insumo para la etapa de registro ya que se concentra específicamente en la detección de esquinas y no líneas o bordes, lo cual presenta dos grandes ventajas, la primera asociada a la identificación correcta de estos puntos en cualquier frame y la segunda su generalización para cualquier escenario, esto debido a que es mucho más probable encontrar esquinas que formas complejas en escenarios no estructurados, lo cual fundamenta el objetivo que el método propuesto pueda ser utilizado en escenarios desconocidos.

Por otra parte, las propiedades afín explotan las ventajas del proceso de reconstrucción y reproyección a la vez que lo simplifican. Utilizando los métodos descritos en esta sección

será posible cumplir con el objetivo de obtener coordenadas 3-D en una secuencia de video, con un costo computacional bajo, a la vez que permite relacionarlos para cualquier instante de tiempo con sus proyecciones bidimensionales.

## 4. Capítulo 4

### REGISTRO

Retomando la clasificación de errores propuesta por Azuma [Azu94], los errores más importantes en los sistemas de RA son producidos en la etapa de registro [Azu94][Baj95], argumentando que los errores en un contexto estático pueden ser reducidos al máximo, pero no ocurre lo mismo con los errores dinámicos.

Para que la información aumentada sea significativa, los modelos deben estar posicionados de forma tal que se “mezclen” en el mundo real, en términos de alineación, perspectiva, iluminación, etc. En cuanto a la alineación, varios métodos han sido propuestos, los cuales, en general, buscan disminuir el error en la alineación abordando una etapa del proceso en particular. En [Seo00][Che99][Tuc95] se proponen técnicas de calibración precisas que permiten modelar con el menor error posible la relación entre la información capturada por la cámara y el mundo real. Otras aproximaciones, por el contrario, abordan el diseño cuidadoso de la marca o de la extracción de características tal que permita una correcta identificación [Tat07][Baj95][Azu94]. Similarmente, han sido propuestos métodos que mejoran el cálculo de información de posición y orientación para reducir el error en la alineación [Yu07][Aba04]. Sin embargo, como se ha argumentado en la presente tesis, la corrección en el error del registro implicará abordar cada una de las etapas que intervienen en el proceso. Si se considera por ejemplo que se tiene un método preciso para el seguimiento de las características pero éstas vienen mal identificadas desde una etapa anterior, hay una fuerte probabilidad de que en un momento particular del proceso, éste falle debido a que puntos característicos que se estaban siguiendo no eran los apropiados, siendo susceptibles a perderse o confundirse

---

fácilmente con la apariencia de otras características. Similarmente, si se dispone de un método preciso para relacionar las coordenadas 2-D y 3-D pero alguna de éstas está mal calculada, la matriz resultante será incorrecta y no podrá representar apropiadamente el posicionamiento y la orientación.

Por su similitud con la etapa de registro en el proceso de reconstrucción 3-D algunos métodos han sido extraídos de este campo e implementados en los sistemas de realidad aumentada. Para efectos de contextualización se podrá sintetizar la etapa de registro en la reconstrucción tridimensional como aquella mediante la cual se halla la mejor transformación para ajustar varias vistas parcialmente solapadas de un mismo objeto; en cuyo caso se dispondrá de información tridimensional para realizar el ajuste. Najafi [Naj04] propone el uso del algoritmo ICP, ampliamente utilizado en reconstrucción, para reducir el error en el alineamiento de la información real-virtual. El algoritmo propuesto por Najafi consta de un refinamiento sucesivo de la pose inicial calculada en el primer frame incorporando para ello dos restricciones de colinearidad que permitan la eliminación de valores extremos.

En cuanto al uso de técnicas heurísticas, recientemente ha sido propuesto [Yu07] el uso de algoritmos genéticos para determinar la mejor pose resultante de la minimización entre los puntos 2-D de una imagen y la reproyección de puntos tridimensionales conocidos. En el algoritmo propuesto, los cromosomas están compuestos por dos segmentos, el primero representando los componentes traslacionales y rotacionales de la pose; el segundo, representa las características 3-D que minimizan la reproyección de dichos puntos utilizando los componentes del primer segmento del cromosoma. Este trabajo presenta resultados interesantes en cuanto a tiempo de computación, dado por el tiempo de convergencia del algoritmo, y minimización del error; sin embargo no presenta resultados de desempeño en escenarios reales, sólo simulados. Por otro lado, completamente diferente al trabajo propuesto en la presente tesis, el trabajo de Yu requiere del modelado a priori de los objetos reales para obtener con precisión la información tridimensional.

La realización de la presente tesis ha sido motivada por los trabajos de Yuan [Yua04][Yua05] y Pang [Pan06], quienes proponen un registro basado en rasgos característicos extraídos de la escena; Yuan utiliza el método de Harris para realizar dicha extracción de puntos, mientras que Pang propone el uso de Shi-Tomasi, junto con un seguimiento realizado con el método Kanade-Lucas-Tomasi (KLT). La reconstrucción

de los puntos tridimensionales se realiza en los trabajos propuestos por Yuan mediante propiedades de reconstrucción proyectivas, completamente diferentes a las propiedades afín propuestas por Yuan. La principal diferencia del presente trabajo y los anteriormente mencionados, radica en la etapa de registro. En los resultados reportados por Yuan en [Yua05] se describe el ajuste de la matriz proyectiva que relaciona los puntos tridimensionales y bidimensionales a partir de una matriz general que pueda ser aplicable a cualquier par de imágenes la cual es ajustada aleatoriamente. Por otra parte, similar el trabajo de Pang, en la presente tesis se utilizan las propiedades afín, sin embargo en el desarrollo del presente trabajo se propone ajustar la matriz de reproyección mediante métodos heurísticos que determinen la mejor relación 2-D/3-D que minimice la diferencia entre los puntos seguidos con el tracker y los puntos estimados a partir de dicha matriz. En la presente tesis se propone el uso de Colonia de Hormigas (Ant Colony Optimization - ACO) transformando el problema del registro en un problema análogo de optimización combinatoria. En el contexto de un sistema que no utilice marcas para el registro no es posible seguir los puntos que define el usuario para superponer los objetos virtuales ya que éstos o bien podrían ocluirse en algún momento a lo largo de la secuencia de video, o podrían no presentar las características necesarias para que pueden ser seguidos correctamente por el tracker. En este orden de ideas, será necesario calcular la matriz de reproyección a partir de los puntos tridimensionales de las características extraídas en la escena y sus proyecciones bidimensionales. Bajo el precepto que dicha matriz podrá ser aplicable no sólo a los puntos característicos de la escena sino a cualquier punto de la imagen, podrán obtenerse las coordenadas de los puntos bidimensionales donde se superpondrán los objetos virtuales, si se conocen sus coordenadas tridimensionales, es decir, utilizando las propiedades de reconstrucción y reproyección afín, se obtienen las siguientes relaciones para  $n$  características:

$$\text{CaracterísticasBidimensionales}_{3 \times n} = M_{3 \times 4} \text{CaracterísticasTridimensionales}_{4 \times n}$$

La matriz de reproyección ( $M_{3 \times 4}$ ) que relaciona las coordenadas tridimensionales y bidimensionales de las características representativas de la imagen podrá ser utilizada para calcular los puntos bidimensionales donde se superpondrán los objetos virtuales si sus coordenadas tridimensionales ya han sido obtenidas a partir de la propiedad de reconstrucción utilizando la matriz descrita en la ecuación (26) de la sección anterior.

Por definición, el registro es un problema de optimización que consiste en hallar la mejor transformación que minimice el error entre dos conjuntos de datos. Dicho problema se convierte en un problema combinatorio, al considerar que la matriz  $M_{3 \times 4}$  requiere la selección de sólo cuatro puntos para que el sistema pueda ser resuelto, por lo que habrá de encontrarse la mejor combinación de pares de puntos 2-D/3-D entre todos los rasgos característicos obtenidos, tal que minimicen el error de reproyección.

El método de Colonia de Hormigas (ACO) está inspirado en el comportamiento de algunas especies de hormigas que depositan una sustancia química llamada feromona indicando que algún camino en particular debe ser seguido por otros miembros de la colonia; dicho principio ha sido explotado por los algoritmos de ACO para resolver problemas de optimización como el presentado anteriormente.

## 4.1 Colonia de Hormigas

En 1959 el entomólogo Pierre-Paul Grassé observó que algunos tipos de termitas reaccionadas a ciertos estímulos cuyos efectos podían actuar como nuevos estímulos tanto para el insecto que la produjo como para los de la colonia. Grassé utilizó el término estigmergia (stigmergy) para describir esta comunicación entre las especies y la describió mediante dos características:

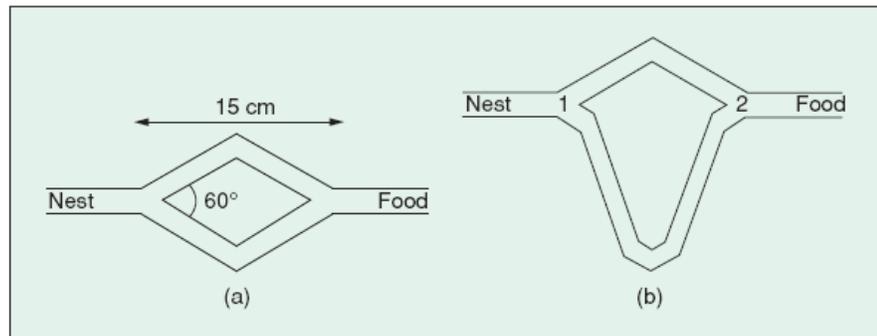
1. La estigmergia es una forma de comunicación indirecta y no simbólica mediada por el ambiente, es decir, los insectos intercambian información modificando su ambiente.
2. La estigmergia es local: la información sólo puede ser accedida por los insectos que se encuentren dentro del lugar.

Ejemplos de estigmergia pueden ser observados en colonias de hormigas cuando se desplazan desde y hacia una fuente de comida depositando feromonas a lo largo del camino, cuando otras hormigas perciben rastros fuertes de feromonas tienden a seguir dicho rastro. En 1990, Deneubourg et al. [Den90] investigó el efecto de las feromonas en las hormigas demostrándolo mediante lo que se denominó el “el experimento del doble puente”. El experimento consistía en conectar el hormiguero de una colonia de hormigas argentinas a una fuente de comida mediante dos puentes de igual longitud (ver Figura

30). Al comienzo del experimento las hormigas exhibían un patrón aleatorio para llegar a la fuente de comida. Sin embargo, después de un tiempo y debido a la concentración de feromonas, las hormigas cruzan por un solo camino.

Un experimento similar al anterior es adelantado por Goss et al. [Gos89] quien propone un modelo matemático que indica la probabilidad de seleccionar uno u otro camino, realizando simulaciones MonteCarlo, al variar la configuración en la longitud de los caminos. Los resultados demuestran el mismo efecto causado por las feromonas. Ambos trabajos motivaron la aparición de los métodos de ACO.

**Figura 30.** Configuración del experimento del doble puente [Dor06] (a) Experimento adelantado por Deneubourg et al. [Den90], (b) variación realizada por Goss et al. [Gos89]

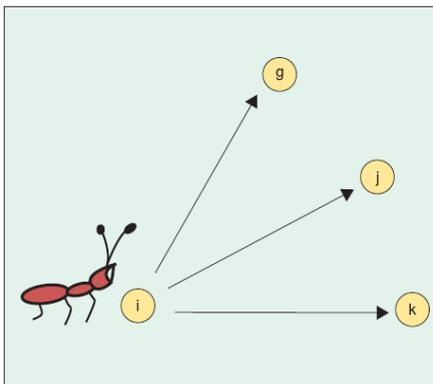


En ACO, el problema es abordado simulando un número de hormigas artificiales que se mueven en un grafo que contiene el espacio de búsqueda del problema. En [Dor06] la optimización basada en colonia de hormigas es ilustrada mediante el problema del vendedor viajero como sigue: el problema del vendedor viajero consiste en encontrar el camino más corto que permita a un vendedor visitar todas las ciudades, cada una de ellas sólo una vez, dado el grupo de ciudades y las distancias entre ellas. Para dicho problema, los vértices del grafo representan las ciudades y las aristas, la conexión entre dos ciudades. La feromona en este caso será distribuida a lo largo de cada arista que conecta cada par de ciudades.

El ACO es un método iterativo que considera el trayecto de cada hormiga como una posible solución al problema. La hormiga camina de vértice a vértice construyendo la solución, con la restricción que no podrá visitar un vértice dos veces. Para una iteración en particular, la hormiga selecciona su siguiente vértice dependiendo de la cantidad de feromona de los vértices susceptibles a ser visitados (ver Figura 31). Al finalizar cada

iteración, los valores de las feromonas son modificados según la calidad de las soluciones construidas por las hormigas de dicha iteración.

**Figura 31.** Elección estocástica del siguiente nodo a visitar



La optimización basada en hormigas fue formalizada como heurística por Dorigo et al. [Dor99] describiendo los siguientes componentes para un modelo  $P = (S, \Omega, f)$  de un problema de optimización combinatorial:

- Un espacio de búsqueda  $S$  definido sobre un grupo finito de variables de decisión discretas  $X_i, i = 1, \dots, n$ ;
- Un grupo de restricciones  $\Omega$  sobre las variables
- Una función objetivo  $f: S \rightarrow \mathbb{R}_0^+$  a ser minimizado.

La variable genérica  $X_i$  toma valores en  $D_i = \{v_i^1, \dots, v_i^{|D_i|}\}$ . Una solución factible  $s \in S$  es una asignación completa de valores de las variables que satisface todas las restricciones  $\Omega$ . Una solución  $s^* \in S$  es llamado el óptimo global si y sólo si  $f(s^*) \leq f(s) \forall s \in S$ .

El algoritmo de un ACO es definido como meta-heurística según el Algoritmo 1.

---

**Algoritmo 1.** La meta-heurística para ACO

---

Inicialización: numeroHormigas, numeroMaxIteraciones, rastroFeromona

**Mientras** no se cumpla la condición de terminación

    ConstruirSolucionHormigas()

    AplicarBusquedaLocal // (opcional)

    ActualizarFeromonas

**Fin\_Mientras**

---

En la literatura han sido propuestos varios algoritmos basados en ACO, tales como: Sistema de Hormigas (AS – Ant System) [Dor91][Dor92][Dor96], Sistema de hormigas MAX-MIN (MMAS) [Stü00], Sistema de Colonia de Hormigas (ACS) [Gam96] los cuales varían principalmente por la forma en la cual se calcula la cantidad de feromona que es dejada por cada hormiga.

Considerando las características de la Optimización Basada en Colonia de Hormigas, en la presente tesis se sigue el algoritmo presentado a continuación para obtener la matriz de reproyección que minimice el error entre las coordenadas estimadas y reales, permitiendo obtener de manera precisa las posiciones bidimensionales de puntos donde se superpondrán los objetos virtuales. Para ello se establece el error como la medida resultante de la ecuación (35) donde  $\widehat{m}_{k_i}$  representa las reproyecciones obtenidas con la matriz calculada y  $m_{k_i}$  las proyecciones seguidas con el tracker.

$$error = \frac{\sum_{i=1}^n ||m_{k_i} - \widehat{m}_{k_i}||}{n} \quad (35)$$

$$Siguiete\_characteristica_{caracteristica}^{S_j} = \left\{ \frac{FeromonaTotal_i^\alpha}{\sum_{g \in S_j} FeromonaTotal_g^\alpha} \right\} \quad (36)$$

---

**Algoritmo 2.** ACO aplicado al cálculo de la matriz de reproyección con el menor error

---

Iniciar: numIteraciones, numHormigas, rastroFeromona, FeromonaTotal, mejorHormi, mejoresIniciales

**Para** i = 1 Hasta numHormigas

Asignar aleatoriamente cuatro características a la hormiga  $S_i$

Estimar el error de cada  $S_i$  mediante la ecuación (35)

**Fin Para**

**Para** j = 1 Hasta numHormigas

Ordenar el vector de Errores de menor a mayor

**Fin Para**

**Para** k = 1 Hasta mejoresHormigas

$$\text{rastroFeromona}[\text{caracteristica}] = \begin{cases} \frac{g:1:\text{mejorHormi}(\max(E_g) - E_j)}{\max_{g,t:1:\text{mejorHormi}}((\max(E_g)) - E_t)}, & \text{caracteristica} \in S_k \\ 0, & \text{de otra manera} \end{cases}$$

$$\text{FeromonaTotal}[\text{caracteristica}] = \text{cte} * \text{FeromonaTotal}[\text{caracteristica}] + \text{rastroFeromona}[\text{caracteristica}]$$

**Fin Para**

**Mientras** IteracionActual < numIteraciones

**Para** u = 1 Hasta numHormigas

A partir de las mejorHormi generar mejoresIniciales cantidad de Hormigas y asignar a  $S_u$

**Fin Para**

**Para** r = mejoresIniciales + 1 Hasta 4

**Para** h = 1 Hasta numHormigas

Dado  $S_j$  elegir la siguiente característica como aquella con mayor posibilidad de ser escogida según la ecuación (36)

**Fin Para**

**Fin Para**

**Para** j = 1 Hasta numHormigas

Estimar el error de cada  $S_j$  mediante la ecuación (35)

**Fin Para**

**Para** j = 1 Hasta numHormigas

Ordenar el vector de Errores de menor a mayor

**Fin Para**

**Para** k = 1 Hasta mejoresHormigas

$$\text{rastroFeromona}[\text{caracteristica}] = \begin{cases} \frac{g:1:\text{mejorHormi}(\max(E_g) - E_j)}{\max_{g,t:1:\text{mejorHormi}}((\max(E_g)) - E_t)}, & \text{caracteristica} \in S_k \\ 0, & \text{de otra manera} \end{cases}$$

$$\text{FeromonaTotal}[\text{caracteristica}] = \text{cte} * \text{FeromonaTotal}[\text{caracteristica}] + \text{rastroFeromona}[\text{caracteristica}]$$

**Fin Para**

**Fin Mientras**

---

## 5. Capítulo 5

### RESULTADOS

Tal como se describe a lo largo de la presente tesis, se argumenta que la corrección del error implicará tanto la selección adecuada de las relaciones 2-D/3-D, así como la intervención en las etapas anteriores a ésta, a saber: identificación de características y seguimiento. A continuación se presentan los resultados obtenidos en cada una de estas etapas y la comparación con otros trabajos reportados en la literatura. Para ello se considerará que las pruebas se dividieron en dos grupos, el primero de ellos, considerando los mismos parámetros reportados en otros trabajos; en el segundo grupo sólo se evaluó el sistema propuesto con el fin de establecer el desempeño frente a diferentes condiciones.

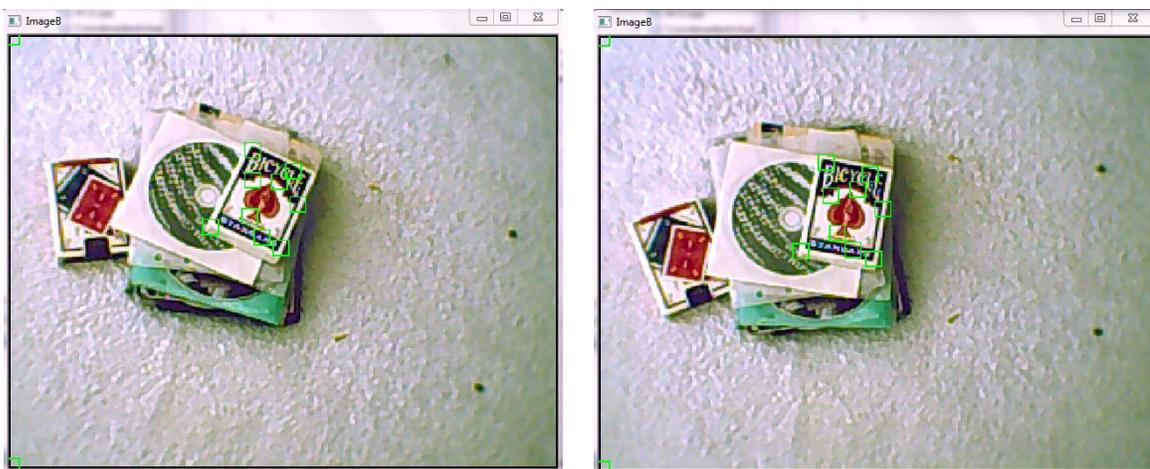
Las pruebas son realizadas utilizando un computador con procesador AMD Turion II Dual-Core 2.2 GHz, memoria de 4.0 GB corriendo bajo el sistema operativo Microsoft Windows 7. Las implementaciones de los métodos fueron realizadas en C++ haciendo uso de las librerías OpenCV y OpenGL para la visualización. Para la adquisición del video se utilizó una cámara web Genius i-Look 110, capturando las imágenes a 30fps.

Tal como se describe en el capítulo 2, se utilizó el extractor de Shi-Tomasi para capturar los rasgos característicos que serán seguidos a lo largo de la secuencia de video. Las Figuras 35 y 36 presentan los resultados obtenidos en esta etapa para un número predefinido de 12 características a identificar (cuadros verdes en las Figuras 35 y 36), este número ha sido seleccionado para realizar las pruebas del primer grupo ya que ha sido la cantidad de puntos reportada en los trabajos de Pang [Pan06]. Con dicha cantidad de puntos, el porcentaje que no se alcanzan a seguir correctamente a lo largo de la

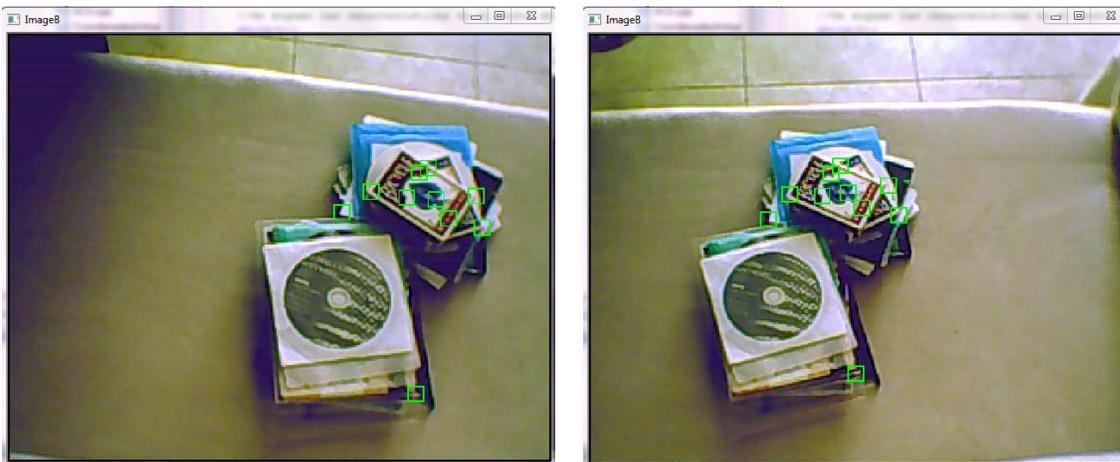
secuencia de video es 0% para la secuencia presentada en la Figura 32. Las pruebas se realizaron con 9 diferentes grupos para los cuales se obtuvo un porcentaje de 0,5% de características perdidas a lo largo de las secuencias de video, las cuales estaban compuestas entre 1.710 (para la secuencia más corta) y 3.510 frames (para la secuencia más larga).

Las Figuras 32 y 33 dejan ver que tal como se esperaba, las características identificadas representan esquinas de los objetos presentes en la escena o de las figuras dibujadas el interior de estos.

**Figura 32.** Características identificadas con el extractor Shi-Tomasi (a) frame 72 (b) frame 137



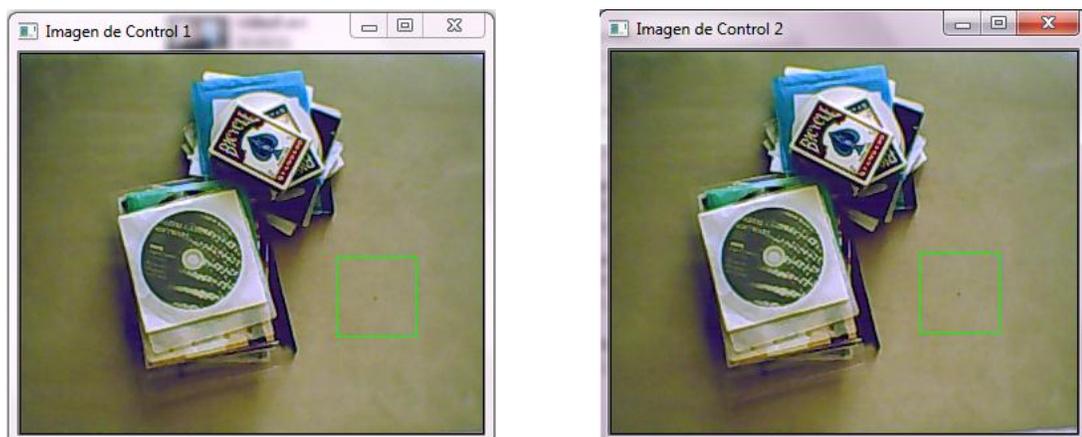
**Figura 33.** Características identificadas con el extractor Shi-Tomasi (a) frame 73 (b) frame 158



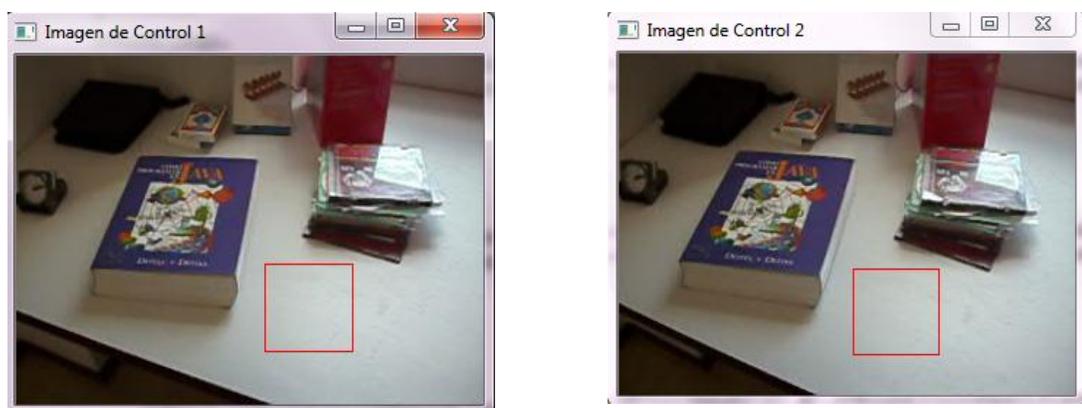
Cabe anotar que estas características son extraídas sólo una vez, para el primer frame, y son seguidas utilizando el tracker KLT. Las Figuras 32(b) y 33(b) presentan los resultados del seguimiento de estas características donde sólo se siguen aquellos puntos que coinciden en ambas imágenes con un error inferior a un umbral, considerando que las características son identificadas a nivel de sub-píxeles, y son buscadas en una ventana de 10 x 10 en el siguiente frame. Si la apariencia de la característica es mayor al umbral es descartada para ser identificada en el siguiente frame. El umbral ha sido definido empíricamente en 650.

El siguiente paso, una vez se tienen las posiciones de todas las características representativas en el frame actual, consiste construir el espacio Euclidiano a partir de los cuatro puntos seleccionados manualmente por el usuario. En este mismo momento se crean las imágenes de control que sirven para definir el espacio afín (ACS – Affine Coordinate System). Los resultados de esta etapa son presentados en las Figuras 34 y 35 para dos secuencias de video diferentes.

**Figura 34.** Imágenes de control secuencia de video 1



**Figura 35.** Imágenes de control secuencia de video 2



Para la selección de los puntos que forman el ACS el usuario puede valerse de alguna marca que facilite la correcta identificación de los cuatro puntos en ambas imágenes. Dicho patrón podría ser removido de la escena una vez haya seleccionado los puntos.

Es importante mencionar que la creación del ACS considerando las mismas imágenes en las cuales el usuario selecciona los puntos que forman el espacio euclidiano, asegura que las bases y el origen del sistema coordenado afín no están siendo ocluidas hasta ese momento, el no haber sido calculado en una etapa anterior evita posibles pérdidas de los puntos principales del ACS.

Una vez que se ha calculado el espacio afín, se han obtenido también las matrices de reconstrucción, reproyección y las coordenadas 3D de los puntos característicos. Las coordenadas tridimensionales son almacenadas junto con un índice que representa una

etiqueta de la coordenada 2-D que generó dicha coordenada. Lo anterior, asegura que si en un momento  $k$  de la secuencia de video algunos de los puntos característicos se han perdido, no habrá confusiones entre los puntos 2D actuales y su correspondencia en las coordenadas 3-D almacenadas.

Con la matriz de reproyección calculada es posible obtener las coordenadas 3-D de los puntos que forman el espacio euclidiano, relacionando dicha matriz con las coordenadas bidimensionales seleccionadas por el usuario.

Para un frame posterior en cualquier instante  $k$  no será necesario volver a calcular la matriz de reproyección como se define en la ecuación (27), será necesario sólo resolver el sistema que relaciona los puntos 3-D de las características previamente almacenadas con sus correspondientes proyecciones bidimensionales calculadas por el tracker.

En la Figura 36 se presentan los resultados siguiendo la metodología anteriormente descrita como es propuesta en [Pan06]. Las Figuras 36(a) y 36 (b) presentan los puntos seleccionados por el usuario.

**Figura 36.** Puntos seleccionados por el usuario para formar el espacio euclidiano (a) Imagen de control 1. (b) Imagen de control 2. Puntos reproyectados (c) frame 126 (d) frame 135 (e) frame 315 (f) frame 453 (g) frame 480 (h) frame 492



(a)



(b)



(c)



(d)



(e)



(f)



(g)



(h)

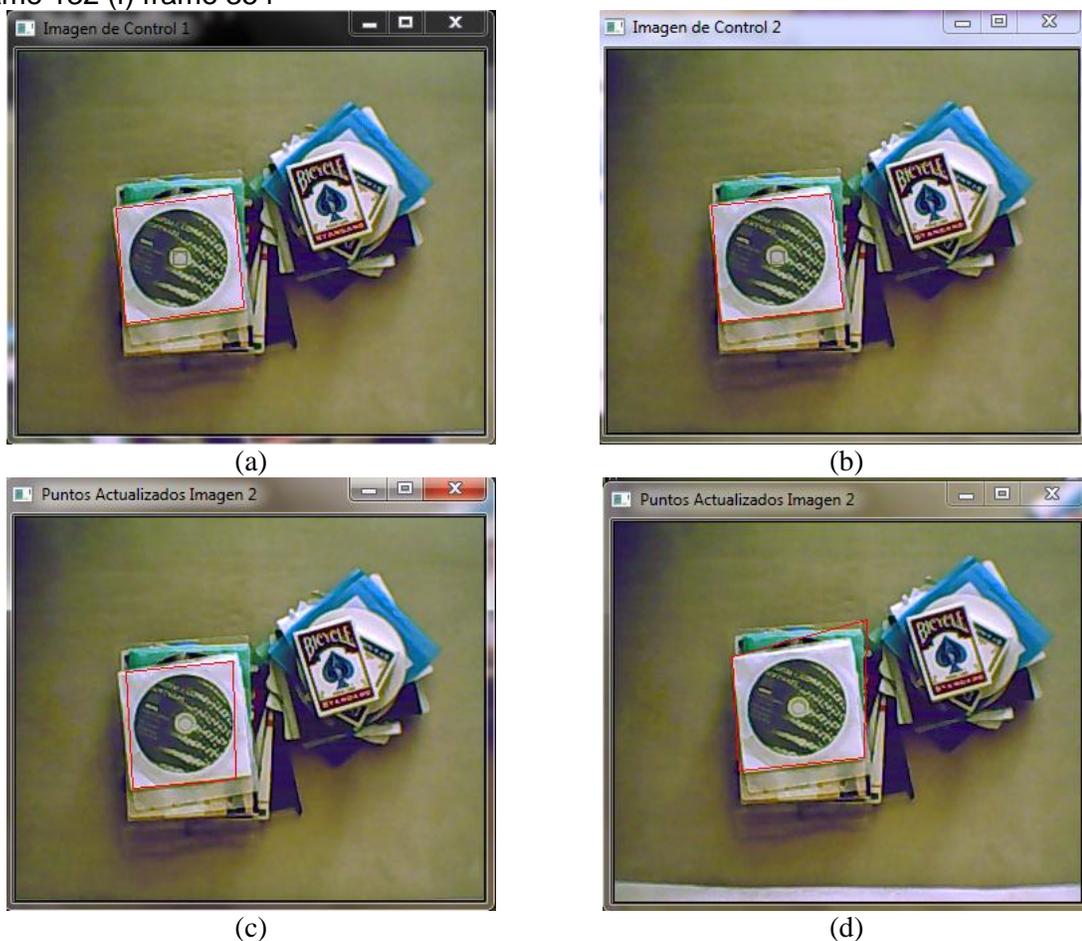
El método propuesto por Pang presenta muy buenos resultados bajo las siguientes condiciones: movimientos lentos de la cámara, diferencias pequeñas entre el punto donde se definieron las bases y el origen del sistema coordenado afín y sus reproyecciones, ausencia de variaciones angulares sobre el eje x. Lo anterior, explica los errores en el reproyección mostrados en la Figura 36. Los cuales se atribuyen en su mayoría a errores en el seguimiento, como se explica a continuación:

- ❖ Recordemos que para un frame  $k$  en un instante  $i$  de la secuencia de video, los puntos bidimensionales son obtenidos a partir del seguimiento de los mismos, lo cual implica un error acumulado de las detecciones en frames anteriores. Aunque dichos errores son muy pequeños de frame a frame, el error acumulado y movimientos grandes la cámara podría ocasionar que la reproyección de los puntos bases y el origen afín no coincidan con lo que serían sus posiciones reales, lo cual se traduciría en un cambio en el sistema coordenado en el cual se definieron los puntos tridimensionales. El cálculo de la matriz de reproyección con el cual se obtienen los puntos que representan el espacio euclidiano presentaría errores debido a que los puntos tridimensionales fueron definidos en un sistema coordenado que no coincide con el sistema coordenado actual.
- ❖ Por otra parte, estos mismos movimientos de la cámara, principalmente ante la presencia de rotaciones, ocasionan que el modelo afín que incorpora el tracker KLT no soporte los cambios de apariencia de las características representativas que deben ser seguidas y terminen por perderse al no poder ser identificadas de un frame al siguiente.

En la presente tesis, se abordan dichas limitaciones, en primer lugar, primero asegurando que la cantidad de puntos total que se no son identificados no sean superiores a un umbral dado. En cuyo caso, sin necesidad de la intervención del usuario, se realiza un proceso de re-inicialización, calculando nuevamente el origen del espacio afín y de sus bases, las coordenadas 3-D de las características y de los puntos previamente almacenados que habían sido seleccionados por el usuario para formar el espacio euclidiano. En segundo lugar, se propone no utilizar características aleatorias para hallar la matriz de reproyección, sólo aquellas cuatro que presenten el menor error como ha sido descrito en la sección anterior. Los resultados de la implementación de la optimización basada en colonia de hormigas son presentados en las Figuras 37, 38 y 39.

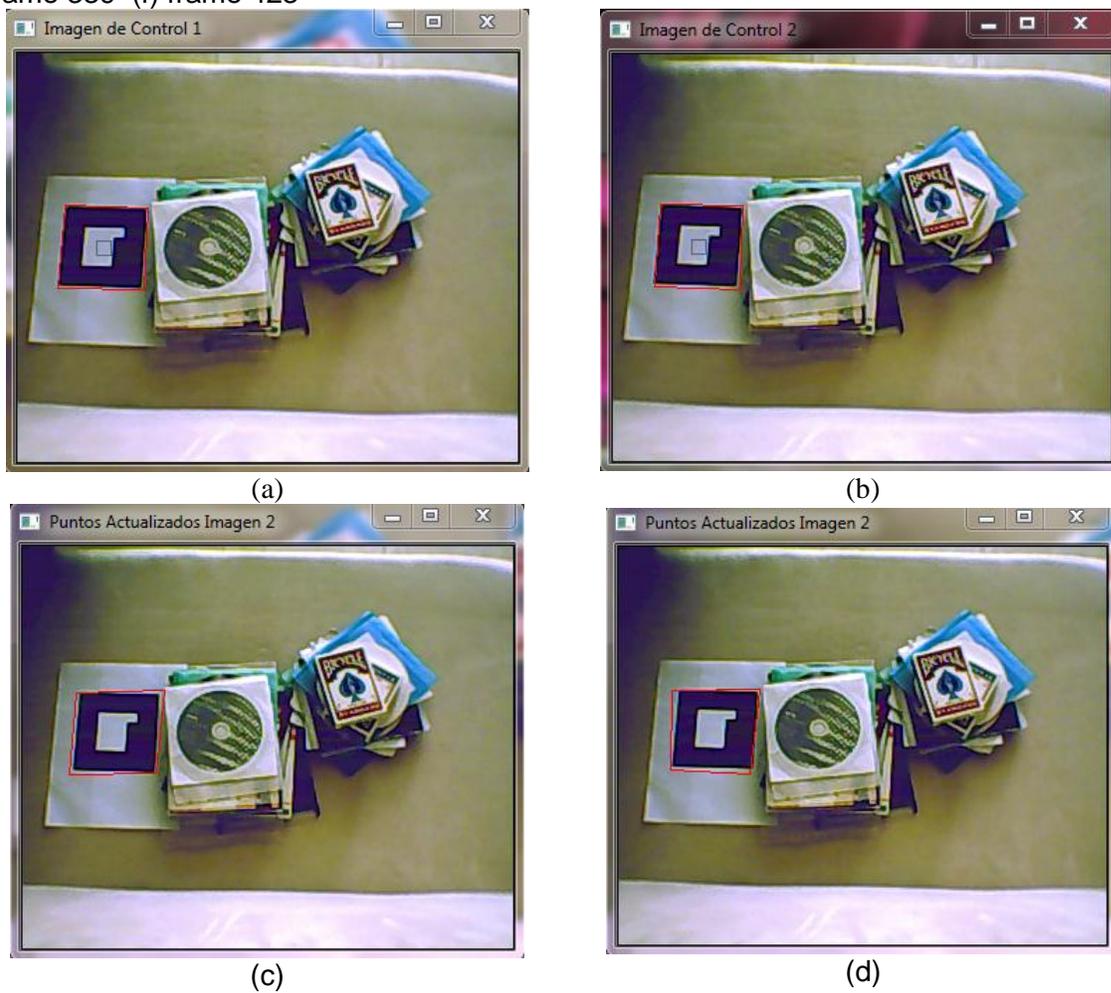
Para mejorar la visualización de la posición correcta donde deberían aparecer los puntos re proyectados del espacio euclidiano, en las Figuras 38 y 39 se ha optado por incorporar un patrón que facilite visualizar la diferencia entre los puntos reales y los estimados, sin embargo, como se presenta al final de estas Figuras, los patrones pueden ser removidos de la escena sin consecuencias de la reproyección.

**Figura 37.** Resultados empleando ACO. Puntos seleccionados por el usuario (a) Imagen de control 1 (b) Imagen de control 2. Reproyecciones en (c) frame 57 (d) frame 94 (e) frame 132 (f) frame 354



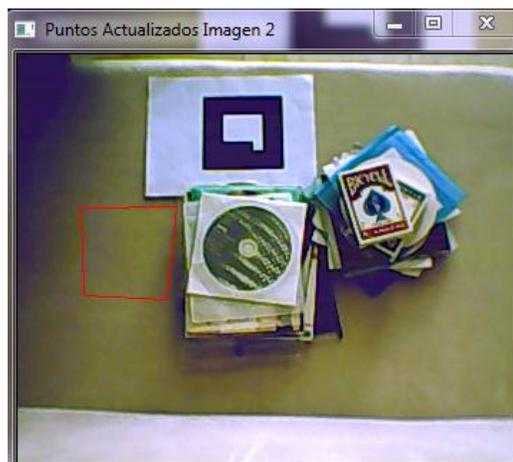


**Figura 38.** Resultados empleando ACO. Puntos seleccionados por el usuario (a) Imagen de control 1 (b) Imagen de control 2. Reproyecciones en (c) frame 312 (d) frame 321 (e) frame 339 (f) frame 423



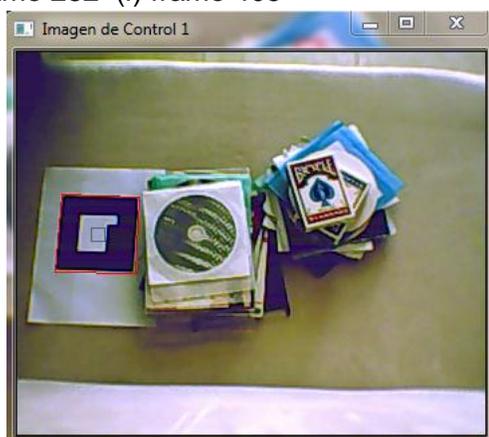


(e)



(f)

**Figura 39.** Resultados empleando ACO. Puntos seleccionados por el usuario (a) Imagen de control 1 (b) Imagen de control 2. Reproyecciones en (c) frame 219 (d) frame 237 (e) frame 282 (f) frame 405



(a)



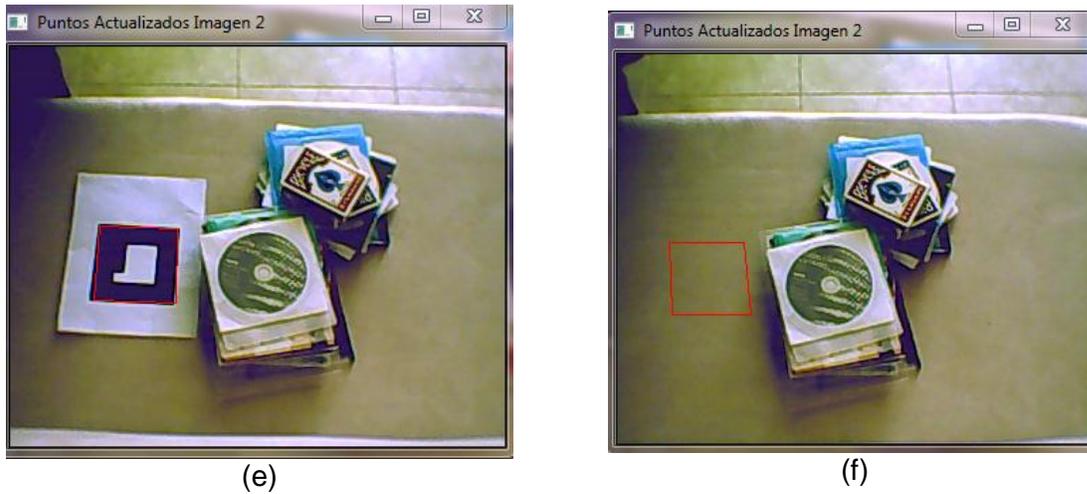
(b)



(c)

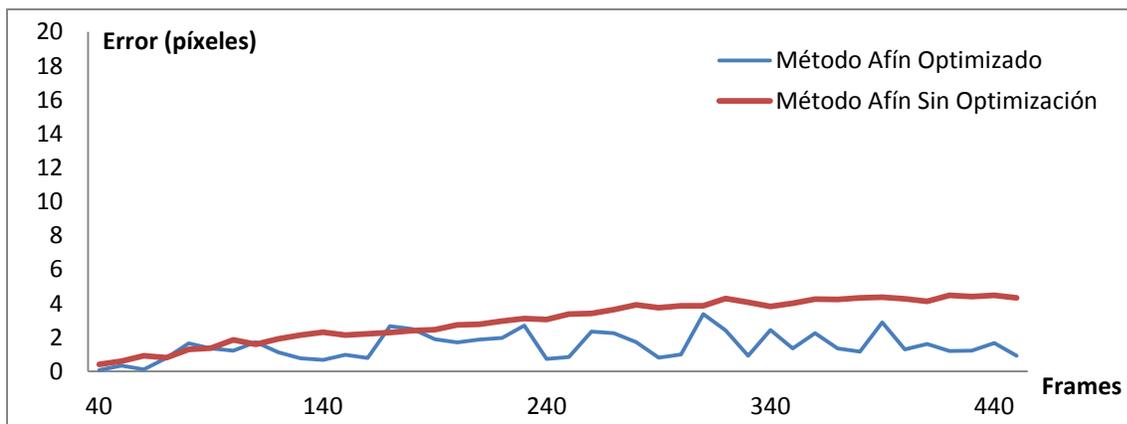


(d)

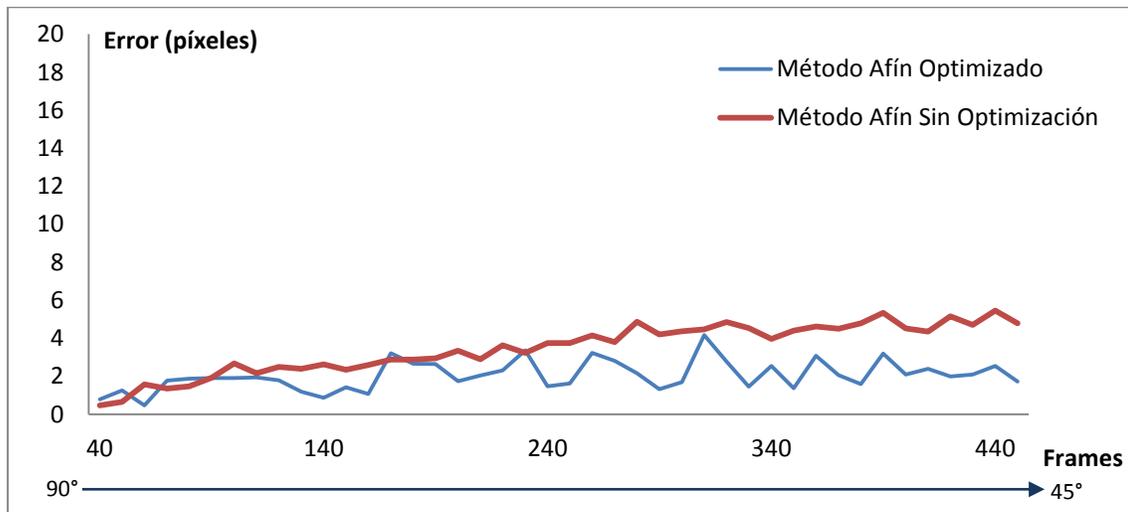


Para la realización de las pruebas del primer grupo se consideran 12 características a identificar, con secuencias de video superiores a 350 frames para la implementación del método afín propuesto por Pang [Pan06]; para la implementación del método afín optimizado propuesto se han seleccionado 200 hormigas y un número de 12 iteraciones. Los resultados son mostrados en la Figura 40. La Figura 41 presenta los resultados comparativos entre el método propuesto por Pang al mover la cámara desde un ángulo de  $90^\circ$  (el eje óptico de la cámara siendo ortogonal a la escena) hasta un ángulo de  $45^\circ$ .

**Figura 40.** Comparación entre el método propuesto optimizado y el método de Pang [Pan06]



**Figura 41.** Comparación entre el método propuesto optimizado y el método de Pang et al. [Pan06] con variaciones desde 90° hasta 45°



Los resultados muestran que el método de Pang es más estable que el método propuesto. Sin embargo el método basado en colonia de hormigas no presenta problemas de errores acumulados en el cálculo de la pose ya que no utiliza resultados de pose de frames anteriores.

La Tabla 2 presenta las pruebas realizadas en el grupo 12, las cuales pretendían determinar el valor óptimo de los parámetros del método propuesto.

**Tabla 2.** Parámetros del método propuesto

Número de características	15
Número de hormigas	200
Número de iteraciones (ACO)	5
$\alpha$ (ACO)	1

El algoritmo general del método propuesto es presentado a continuación

---

**Algoritmo 3.** Algoritmo general del método propuesto

---

Capture frame1

caracteristicasTracked = ExtraerCaracteristicas(frame1)

**Mientras** (True)

    Capture frame2

    caracteristicasTracked = Tracker(frame1, frame2, caracteristicasTracked)

**Si** evento **entonces** //El usuario escoge libremente cuando crear las img de control

        ACS = CrearACS(caracteristicasTracked) //puntos 3-D de características

        CrearWCS(frame1, frame2) //Crea el sistema coord. Euclidiano

        TridimensionalWCS = CreateTridimensionalWCS(ACS)

        CreatedWCS = true

**Fin Si**

**Si** (CreatedWCS) **entonces**

        MatrizReproyeccion = ACO(caracteristicasTracked, ACS)

        Repro2D =CoordReproyectadasWCS(MatrizReproyeccion, TridimensionalWCS)

        DibujarReproyeccion(Repro2D, frame1, frame)

**Fin Si**

**Fin Mientras**

---

## 6. Conclusiones

La habilidad de la realidad aumentada para presentar información superpuesta directamente sobre el campo de visión del usuario, abre interesantes oportunidades para la interacción con el ambiente real apoyando las tareas de visualización, acceso a información y toma de decisiones. Sin embargo, las técnicas que permiten dicho proceso necesitan de mucha información a priori para obtener resultados que sean visualmente coherentes, en otras palabras, es necesario conocer la escena o manipularla para que los objetos virtuales aparezcan ubicados en la posición correcta.

Para superar la limitación de la cantidad de información a priori necesaria, se proponen técnicas pasivas que no modifiquen el escenario real sino que obtengan algún tipo de información “útil” a partir de ésta que permita calcular la información de posicionamiento. Dichas técnicas han sido denominadas markerless, ya que evitan la necesidad del uso de marcas, y por lo tanto, requieren intervención mínima del mundo real. Lamentablemente, la ventaja de estas técnicas se convierte también en su debilidad al generar mayores errores en el posicionamiento, puesto que no se tiene una ubicación exacta de donde se superpondrán los objetos virtuales.

En la presente tesis se ha propuesto un método basado en Optimización de Colonia de Hormigas (ACO) para disminuir el error existente entre la información real y la información estimada, tal que permita un cálculo preciso de la pose, y por lo tanto un registro con el menor error posible. Tal como afirma Azuma [Azu94], la eliminación del error en el registro es casi imposible, sin embargo trabajos como el aquí propuestos brinda un gran aporte a lo que se conoce dentro del campo de la realidad aumentada. Dicho aporte viene dado principalmente en la reducción del error en la visualización, lo que permite que esta técnica pueda ser utilizada en campos de aplicación donde se

requiere de gran precisión en la alineación real-virtual y no se utilizan marcas para el registro.

Si bien el método plantea muy buenos resultados comparados con otros reportados en la literatura, el método plantea nuevos retos que se convierten en nuevas investigaciones. Como trabajo futuro se presentan tres principales retos: el primero asociado a la implementación de una técnica de realimentación adaptativa que elimine o disminuya la inestabilidad del método propuesto representada en los *jitters*. En segundo lugar, se propone incorporar técnicas que superen la restricción del método propuesto en cuanto a la renderización en superficies planas. Finalmente, habrá de recordarse que una de las características de la realidad aumentada es su interacción en tiempo real, una de las ventajas de ACO que no fue explotada en el presente trabajo consiste en la paralelización del método, la exigencia del desempeño en tiempo real se abordó en la presente tesis aumentando el número de hormigas y disminuyendo el número de iteraciones necesarias para que el algoritmo converja.

## 7.Referencias

- [Aba04] Ababsa, F-E., Mallem, M. (2004) Robust Camera Pose Estimation Using 2D Fiducials Tracking for Real-Time Augmented Reality. Proceedings of the 2004 ACM SIGGRAPH International Conference on Virtual Reality Continuum and Its Applications in Industry.
- [Azu94] Azuma, R. T., Gary, B. (1994) Improving Static and Dynamic Registration in an Optical See-Through HMD. Proceedings of SIGGRAPH' 94, Computer Graphics, Annual Conference Series, pp. 197-204
- [Azu97] Azuma, R. T. (1997) A Survey of Augmented Reality. Teleoperators and Virtual Environments. Vol. 6, No. 4, pp. 355-385
- [Azz08] Azzari, P., Di Stefano, L., Tombari, F., Mattocchia, S. (2008) Markerless Augmented Reality Using Image Mosaics. Lecture Notes in Computer Science. Vol. 5099, pp. 413-420
- [Baj92] Bajura, M., Fuchs, H. and Ohbuchi, R. (1992). Merging Virtual Objects with the Real World: Seeing Ultrasound Imagery within the Patient. Computer Graphics Vol. 26, No. 2, pp. 203-210
- [Baj95] Bajura, M., Neumann, U. (1995) Dynamic Registration Correction in Augmented Reality Systems. IEEE Virtual Reality Annual International Symposium (VRAIS), pp. 189-196
- [Baj95] Bajura, M., Neumann, U. (1995) Dynamic Registration Correction in Video-Based Augmented Reality Systems. IEEE Computer Graphics. Vol. 15 No. 5, pp. 52-60
- [Bal82] Ballard, D. H., Brown, C. M (1982) Computer Vision. Prentice-Hall. Englewood Cliffs, New Jersey.
- [Bay01] Bayro-Corrochano, E., Rosenhahn, B. (2001). Analysis and Computation of the Intrinsic Camera Parameter. Geometric Computing with Clifford Algebras. Capitulo 16. Springer-Verlag
- [Bay02] Bayro-Corrochano, E., Rosenhahn, B. (2002). A Geometric Approach for the Analysis and Computation of the Intrinsic Camera Parameters. Pattern Recognition, Vol. 35, pp. 169-186
- [Bie93] Bier, E. A., Stone, M. C., Pier, K., Buxton, W., DeRose, T. D. (1993). Toolglass and Magic Lenses: The see-through Interface. Procceedings SIGGRAPH'93. Vol. 27, pp. 73-80
- [Ble08] Bleser, G., Stricker, D. (2008). Advanced Tracking Through Efficient Image Processing and Visual-Inertial Sensor Fusion. IEEE Virtual Reality, pp. 137- 144
- [Blu09] Blum, L., Broll, W., Müller, S. (2009) Augmented Reality Under Water. SIGGRAPH 2009

- [Bre09] Bres, S., Tellez, B. (2009) Localisation and Augmented Reality for Mobile Applications in Cultural Heritage. Proceedings 3<sup>rd</sup> ISPRS International Workshop. Vol. 38
- [Bro96] Brooks, F. (1996). The Computer Scientist as Toolsmith II. Communications of the ACM Vol. 39, No. 3, pp. 61-68
- [Bro96a] Brooks, F. P. Jr (1996) An Introduction to Cybernetics. Chapman & Hall.
- [Caa03] Caarls, J., Jonker, P., Persa, S. (2003). Sensor Fusion for Augmented Reality. Proceedings of European Symposium on Ambient Intelligence, pp. 160-176
- [Cat07] Catchpole, J. (2007). Adaptive Vision Based Scene Registration for Outdoor Augmented Reality. Tesis Doctoral. Universidad de Waikato
- [Cha02] Chai, L., Hoff, WA., Vincent, T. (2002) Three-dimensional Motion and Structure Estimation Using Inertial Sensors and Computer Vision for Augmented Reality. Teleoperators and Virtual Environment. Vol. 11, No. 5, pp. 474-492
- [Che08] Chen He, X., Yung, N. H. C (2008) Corner Detector Based on Global and Local Curvature Properties. Optical Engineering. Vol. 47, No. 5, pp. 1 - 12
- [Che99] Chen CS, Yu CK, Hung YP. (1999) New Calibration-free Approach for Augmented Reality Based on Parameterized Cuboid structure. Proceedings of the 7<sup>th</sup> International Conference on Computer Vision, Corfu, Greece, pp. 30 – 37
- [Com03] Comport, A. I., Marchand, E., Chaumette, F. (2003) A Real-time Tracker for Markerless Augmented Reality. ACM/IEEE International Symposium on Mixed and Augmented Reality, pp- 36-45
- [Dem93] De Ma, S. (1993) Conics-Based Stereo. Motion Estimation and Pose Determination. International Journal of Computer Vision, Vol. 10, No. 1, pp. 7 -25
- [Den90] Deneubourg, J-L., Aron, S., Goss, S., Pasteels, J-M (1990) The self-organizing exploratory pattern of Argentine Ant. Journal of Insect Behaviour, Vol. 3, pp. 159.
- [Dho89] Dhome, M., Richetin, M., Lapresté, J-T., Rives, G. (1989) Determination of the Attitude of 3D Objects from a Single Perspective View. IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 11, No. 12, pp. 1265 – 1278.
- [Dor06] Dorigo, M., Birattari, M., Stutzle, T. (2006) Ant Colony Optimization – Artificial Ants as a Computational Intelligence Technique. IEEE Computational Intelligence Magazine. Noviembre 2006
- [Dor91] Dorigo, M., Maniezzo, V., Colorno, A. (1991). Positive Feedback as a Search Strategy. Departamento de Electrónica, Politecnico di Milano. Italia. Reporte técnico 91-016.
- [Dor92] Dorigo, M. (1992) Optimization Learning and Natural Algorithms. Tesis doctoral. Departamento de Electronica. Politecnico di Milano.
- [Dor96] Dorigo, M., Maniezzo, V., Colorni, A. Ant System: Optimization by a Colony of Cooperating Agents. IEEE Transactions on Systems, Man and Cybernetics – Part B. Vol. 26, No. 1, pp. 29 - 41
- [Dor99] Dorigo, M., Di Carlo, G., Gambardella, L.M. (1999) The Ant Colony Optimization meta-heuristic. New Ideas in Optimization, D. Corne et al. eds. McGraw Hill, pp. 11 -32
- [Dru02] Drummond, T., Cipolla, R. (2002) Real-time Visual Tracking of Complex Structures. IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 27, No. 7, pp. 932 - 946
- [Fau92] Faugueras, O. D., Luong, Q-T., Maybank, S. J. (1992) Camera self-calibration : Theory and Experiments. Proceedings of Second European Conference on Computer Vision, pp. 321-334
- [Fau92a] Maybank, S. J., Faugueras, O. D. (1986) A Theory of Self-calibration of a Moving Camera. International Journal of Computer Vision. Vol. 8, No. 2, pp. 123-151
- [Fei93] Feiner, S.; MacIntyre, B. and Seligmann, D. (1993). Knowledge-Based Augmented Reality. Communications of the ACM, Vol. 36, No. 7, pp. 53-62

- [Fer01] Ferrari, V., Tuytelaars, T., Van Gool, V. (2000) Markerless Augmented Reality with a Real-time Affine Region Tracker. Proceedings IEEE/ACM International Symposium on Augmented Reality, pp. 87 - 96
- [Fou94] Fournier, A. (1994) Illumination Problems in Computer Augmented Reality. Journée INRIA. Analyse/Synthèse D'Images, pp. 1 -21
- [Fre06] Freeman, R., Steed, A. (2006) Interactive Modelling and Tracking for Mixed and Augmented Reality. Proceedings of the ACM Symposium on virtual Reality Software and Technology.
- [Fru03] Frund, J., Gausemier, J., Matyszczok, C., y Radkowski, R. (2003) Applications Areas of AR-Technology within Automobile Advance Development. International Workhop on Potential Industrial Applications of Mixed and Augmented Reality, pp. 1-7
- [Gam96] Gambardella, L. M., Dorigo, M. (1996) Solving Symmetric and Asymmetric TSPs by Ant Colonies. Proceedings IEEE International Conference on Evolutionary Computation (ICEC'96)
- [Gan84] Ganapathy, S. (1984) Decomposition of Transformation Matrices for Robot Vision. Proceedings IEEE International Conference on Robotics and Automation, pp. 130-139
- [Gar08] Gary, B., Kaehler, A (2008) Learning OpenCV. O'Reilly
- [Gos89] Goss, S., Aron, S., Deneubourg, J-L., Pasteels, J-M. (1989) Self-organized shortcuts in the Argentine Ant- *Naturwissenschaften*, Vol. 16, pp. 579-581.
- [Gut08] Gutierrez, M., Vexo, F. y Thalmann, D. (2008) Steeping into Virtual Reality. Springer
- [Hil02] Hildebrand, A., et al. 2002. ARCHEOGUIDE: An Augmented Reality Based System for Personalized Tours in Cultural Heritage Sites. Cultivate Interactive, Vol. 1. 2002
- [Hof96] Hoff, W. A., Kguyen, K., Lion, T. (1996) Computer Vision-Based Registration Techniques for Augmented Reality. Proceedings of 15th Intelligent Robots and Computer Vision. Vol. 2904, pp. 538-548
- [ISpe09] IEEE Spectrum, Engineering Achievements: The Two Lists. Diciembre 2009, pp. 21
- [Jo08] Jo, K., Kakehi, Y., Minamizawa, K., Nii, H., Kawakami, N., Tachi, S. (2008) ARForce: A Marker-based Augmented Reality System for Force Distribution Input. Proceedings of the International Conference on Advances in Computer Entertainment Technology, pp. 160-165
- [Jua05] Juan, M. C., Alcaniz, M., Monserrat, C., Banos, R. M., Guerrero, B. (2005) Using Augmented Reality to Treat Phobias. IEEE Computer Graphics and Applications, Vol. 25, No. 6, pp. 31-37
- [Jua06] Juan, M. C., Joele, D., Baños, R., Botella, C., Alcañiz, M., Van Der Mast, C. (2006) A Markerless Augmented Reality System for the Treatment of Phobia to Small Animals. Proceedings 9th Annual International Workshop on Presence.
- [Kam05] Kamiyama, K., Vlack, K., Mizota, T., Kajimoto, H., Kawakami, N., Tachi, S. (2005) Vision-based Sensor for Real-time Measuring of Surface Traction Fields. IEEE Computer Graphics and Applications, pp. 68-75
- [Kat00] Kato, H., Billinghamurst, M., Poupyrev, I., Imamoto, K., Tachibana, K. (2000) Virtual Object Manipulation on a Table-Top AR Environment. Proceedings of the International Symposium on Augmented Reality, pp.111-119
- [Kat99] Kato, H., Billinghamurst, M (1999) Marker Tracking and HMD Calibration for a Video-based Augmented Reality Conferencing System. Proceedings of the 2nd International Workshop on Augmented Reality
- [Kat99] Kato, H., Billinghamurst, M (1999) Marker Tracking HMD Calibration for a Video-Based Augmented Reality Conferencing System. Proceedings of the 2<sup>nd</sup> IEEE and ACM International Workshop on Augmented Reality.

- [Kim03] Kim, J. B.; Moon, C. H. and Kim, H, J. (2003). A Person Annotation Overlay Using a Wearable Computer Augmented Reality. Proceedings 2nd IEEE and ACM International Symposium on Mixed and Augmented Reality – ISMAR'03, pp. 244-245
- [Kum94] Kumar, R., Hanson, A. R. (1994) Robust Methods for Estimating Pose and Sensitivity Analysis. CVGIP: Image Understanding, Vol. 60, No. 3, pp. 313-342
- [Kut96] Kutulakos, K. N., Vallino, J (1996) Affine Object Representations for Calibration-Free Augmented Reality. IEEE Virtual Reality Annual International Symposium, VRAIS'96, pp. 25-36
- [Kut98] Kutulakos, K, N., Vallino, J. R. (1998) Calibration-free Augmented Reality. IEEE Transactions on Visualization and Computer Graphics, Vol. 4, No. 1, pp. 1-20
- [Lee08] Lee, W., Woo, W. (2008) Exploiting Context-Awareness in Augmented Reality Applications. International Symposium on Ubiquitous Virtual Reality, pp. 51-54
- [Lee08a] Lee, T., Höllerer, T (2008) Hybrid Feature Tracking and User Interaction for Markerless Augmented Reality. Proceedings IEEE Conference on Virtual Reality, pp. 145-152
- [Lee09] Lee, W., Woo, W. (2009) Real-Time Color Correction for Marker-based Augmented Reality Applications. International Workshop on Ubiquitous Virtual Reality, pp. 32-35
- [Lee09a] Lee, T., Höllerer, T. (2009) Multithreaded Hybrid Feature Tracking for Markerless Augmented Reality. IEEE Transactions on Visualization and Computer Graphics, Vol. 15, No. 3, pp. 355-368
- [Len88] Lenz, R. K., Tsai, R. Y (1988) Techniques for Calibration of the Scale Factor and Image Center for High Accuracy 3D Machine Vision Metrology. IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 10, pp. 713-720
- [Lep03] Lepetit, V., Vachetti, L., Thalmann, D., Fua, P. () Fully Automated and Stable Registration for Augmented Reality Applications. Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality, pp. 93-102
- [Low87] Lowe, D. G. (1987) Three-dimensional Object Recognition from Single Two-dimensional Images. Artificial Intelligence, Vol. 31, pp- 355-395
- [Low92] Lowe, D. G. (1992) Robust Model-based Motion Tracking through the Integration of Search and Estimation. International Journal of Computer Vision. Vol. 8, No. 2, pp. 113-122
- [Luc81] Lucas, B. D., Kanade, T. (1981) An Iterative Image Registration Technique with an Application to Stereo Vision. Proceedings of the 17<sup>th</sup> International Joint Conference on Artificial Intelligence. Vol. 2, pp. 674 - 679
- [Mel95] Mellor, J. P. (1995). Enhanced Reality Visualization in a Surgical Environment. Artificial Intelligent Laboratory. A.I reporte técnico No. 1544
- [Mer07] Merrill, D., Maes, P. (2007) Augmented Looking, Pointing and Reaching Gestures to Enhance the Searching and Browsing of Physical Objects. Proceedings 5<sup>th</sup> International Conference on Pervasive Computing
- [Mil94] Milgram, P., Kishino, F. (1994) A Taxonomy of Mixed Reality Visual Displays. IEICE Transactions on Information Systems, Vol E77-D, No. 12
- [Mis09] Mistry, P., Maes, P (2009) Augmenting Sticky Notes as an I/O Interface. Proceedings 13<sup>th</sup> International Conference on Human-Computer Interaction.
- [Moo09] Mooser, J., You, S., Neumann, U., Wang, Q. (2009) Applying Robust Structure from Motion to Markerless Augmented Reality. IEEE Workshop on Applications of computer Vision.
- [Naj04] Najafi, H., Navab, N., Klinker, G. (2004) Automated Initialization for Markerless Tracking: A Sensor Fusion Approach. Proceedings of the 3<sup>rd</sup> International IEEE/ACM Symposium on Mixed and Augmented Reality, pp. 79 – 88
- [Nak05] Nakazato, Y., Kanbara, M., Yokoya, N (2005) Wearable Augmented Reality System Using Invisible Visual Markers and an IR Camera. Proceedings of the 9th IEEE International Symposium on Wearable Computers, pp. 198-199

- [Neu99] Neumann, U., You, S. (1999) Natural Feature Tracking for Augmented Reality. IEEE Transaction on Multimedia, Vol. 1, No. 1, pp. 53 - 64
- [Obh92] Ohbuchi, R.; Chen, D. and Fuchs, H. (1992). Incremental Volume Reconstruction and Rendering for 3D Ultrasound Imaging. SPIE Proceedings, Vol. 1808, pp. 312-323
- [Owe03] Owen, C. B. et al. Augmented Imagery for Digital Video Applications. CRC Handbook of video Databases, CRC Press LLC.
- [Pan06] Pang, Y., Yuan, M. L., Nee, A. Y. C., Ong, S. K., Youcef-Toumi, K (2006) a Markerless Registration Method for Augmented Reality based on Affine Properties. Proceedings of the 7<sup>th</sup> Australasian User Interface Conference, Vol. 50
- [Pas99] Pasman, W., Van Der Schaaf, A., Lagendijk, R., Jansen, F. W. (1999) Accurate Overlaying for Mobile Augmented Reality. Computer & Graphics. Vol. 23, No. 6, pp. 875-881
- [Pri02] Prince, S. J. D., Xu, K., Cheok, A. D. (2002) Augmented Reality Camera Tracking with Homographies. IEEE Computer Graphics and Applications, Vol. 22, No. 6, pp. 39-45
- [Rei07] Reitmayr, G., Eade, E., Drummond, T. W. (2007) Semi-automatic Annotations in Unknown Environments. Proceedings of the 6<sup>th</sup> IEEE and ACM International Symposium on Mixed and Augmented Reality,
- [Reg03] Regenbrecht, H. et al. (2003) An Augmented Virtuality Approach to 3D Videoconferencing. Proceedings of the Second IEEE and ACM International Symposium on Mixed and Augmented Reality, pp. 290-291
- [Ros05] Rosten, E.; Reitmayr, G. and Drummond, T. (2005). Real Time Video Annotations for Augmented Reality. International Symposium on Visual Computing
- [Ros95] Rose, E.; Breen, D.; Ahlers, K.; Crampton, C.; Tuceryan, M.; Whitaker, R. and Greer, D. (1995). Annotating Real-World Objects Using Augmented Reality. Proceedings of Computer Graphics International'95, pp. 357-370
- [Sat01] Satoh, K.; Anabuki, M.; Yamamoto, H.; Tamura, H. (2001). A Hybrid Registration Method for Outdoor Augmented Reality. Proceedings of the IEEE and ACM International Symposium on Augmented Reality, pp. 67-76
- [Sat07] Sato, K., Minamizawa, K., Kawakami, N., Tachi, S. (2005) Haptic Telexistence. 34th International Conference on Computer Graphics and Interactive Techniques – ACM SIGGRAPH 2007
- [Seo00] Seo, Y., Hong, K. S. (2000) Calibration-free Augmented Reality in Perspective. IEEE Transactions on Visualization and Computer Graphics, Vol. 6, No. 4, pp. 346-359
- [Shi94] Shi, J., Tomasi, C. (1994) Good Features to Track. IEEE Conference on computer Vision and Pattern Recognition, pp. 593 -600
- [Sim00] Simon, G., Fitzgibbon, A. W., Zisserman, A. (2000) Markerless Tracking Using Planar Structures in the Scene. IEEE/ACM International Symposium on augmented Reality, pp. 120 – 128
- [Skr04] Skrypnik, I., Lowe, D. G. (2004) Scene Modelling, Recognition and Tracking with Invariant Image Feature. Proceedings of the 3<sup>rd</sup> IEEE/ACM International Symposium on Mixed and Augmented Reality, pp. 110 - 119
- [Sta94] State, A.; Chen, D. T.; Tector, C.; Brandt, A.; Chen, H.; Ohbuchi, R.; Bajura, M. and Fuchs, H. (1994). Case Study: Observing a Volume-Rendered Fetus within a Pregnant Patient. Proceedings of IEEE Visualization'94, pp. 364-368
- [Sta96] State, A., Hirota, G., Chen, D. T., Garret, W. F., Livingston, M. A., (1996) Superior Augmented Reality Registration by Integrating Landmark Tracking and Magnetic Tracking. Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques. pp. 429 -438

- [Stri98] Stricker, D.; Klinker, G.; Reiners, D. A. (1998). A Fast and Robust line-based Optical Tracker for Augmented Reality Applications. Proceedings on the International Workshop on Augmented Reality.
- [Stü00] Stütze, T., Hoos, H. H. (2000) MAX-MIN Ant System. Future Generation Computer Systems, Vol. 16, No. 8, pp. 889-914
- [Tat07] Tateno, K., Kitahara, I., Ohta, Y. (2007) A Nested Marker for Augmented Reality. IEEE Virtual Reality Conference – Charlotte, pp. 259-262
- [Tho00] Thomas, B., Close, B., Donoghue, J., Squires, J., De Bondi, P., Morris, M., Piekarski, W. (2000) ARQuake: an Outdoor/indoor Augmented Reality First Person Applications. IV International Symposium on Wearable Computers, pp. 139-146
- [Tsa85] Tsai, R. Y. (1985) A Versatile Camera Calibration Technique for High Accuracy 3D Machine Vision Metrology using off-the-shelf TV Cameras and Lenses. Reporte RC 11413, IBM T. J. Watson Research Center, Yorktown Heights
- [Tuc95] Tuceryan, M., Greer, D. S., Whitaker, R. T., Breen D. E., Crampton, C., Rose, E., Ahlers, K. H. (1995) Calibration Requirements and Procedures for a monitor-Based Augmented Reality System. IEEE Transactions on Visualization and Computer Graphics, Vol. 1, No. 3, pp. 255-273
- [Tuc95] Tuceryan, M. et al. (1995) Calibration Requirements and Procedures for a Monitor-Based Augmented Reality System. IEEE Transaction on Visualization and Computer Graphics. Vol.1. No. 3, pp. 255-297
- [Uml02] Umlauf, E. J., Piringer, H., Reitmayr, G., Schmalstieg, D. (2002) ARLib: the augmented library. Proceedings of the International Workshop on ARToolkit.
- [Vla01] Vlahakis, V., et al. 2001. First Results of an Augmented Reality, Mobile Computing System in Cultural Heritages Sites. Proceedings of the Conference on Virtual Reality, Archeology and Cultural Heritage, pp. 131-140. 2001
- [Wei91] Weiser, M (1991) The Computer for the 21<sup>st</sup> Century. Scientific American. Scientific American, Vol. 265, No. 3, pp. 66-75
- [Wei99] Weiser, M. (1999) The Computer for the 21st Century. Mobile Computing and Communications Review. Vol. 3, No. 3, pp. 3-11
- [Wen92] Weng, J., Cohen, P., Herniou, M (1992) Camera Calibration with Distortion Models and Accuracy Evaluation. IEEE Transaction on Pattern Analysis and Machine Intelligence. Vol 14, No. 10, pp. 965-980
- [Whi04] White, M. et al. ARCO – An Architecture for Digitalization, Management and Presentation of Virtual Exhibitions. Proceedings of the Computer Graphics International (CGI'04), pp. 622-625. 2004
- [You99] You, S.; Neumann, U.; Azuma, R. (1999) Hybrid Inertial and Vision Tracking for Augmented Reality Registration. Proceedings of the IEEE Virtual Reality, pp. 260-267
- [Yu07] Yu, Y. K., Wong, K. W., Yuen-Chang, M. M. (2007) Pose Estimation for Augmented Reality Applications Using Genetic Algorithm.
- [Yua04] Yuan, M. L., Ong, S. K., Nee, A. Y. C. (2004) Registration Using Projective Reconstruction for Augmented Reality Systems. Proceedings of the 2004 ACM SIGGRAPH International Conference on Virtual Reality Continuum and its Applications in Industry
- [Yua05] Yuan, M.L., Ong, S. K., Nee, A. Y. C. (2005) A Generalized Registration Method for Augmented Reality Systems. Computers and Graphics, No. 29, pp. 980-997
- [Yua06] Yuan, C. (2006) Markerless Pose Tracking for Augmented Reality. Lecture Notes in Computer Science, Vol. 4291, pp. 721-730
- [Zau03] Zauner, J., Haller, M., Brandl, A. (2003) Authoring of a Mixed Reality Assembly Instructor for Hierarchical Structures. Proceedings of the Second IEEE and ACM International Symposium on Mixed and Augmented Reality, pp. 237-246