

Reconocimiento de objetos en escenas complejas para aplicaciones de realidad aumentada

Mario Gerardo Canul Ku¹, mariocanul@cimat.mx, Guanajuato, Guanajuato , Dr. Diego Jiménez Badillo², diego_jimenez@inah.gob.mx , Dr. Salvador Ruiz Correa ³, src@cmls.pw , San Luis Potosí S.L.P., Dr. Jean-Bernad Hayet ¹,jbhayet@cimat.mx, Guanajuato, Guanajuato.

¹Centro de Investigación en Matemáticas, A.C. (CIMAT), ²Museo del Templo Mayor, ³ Instituto Potosino de Investigación Científica y Tecnológica (IPICYT).

“Primer Congreso Internacional Patrimonio Cultural y Nuevas Tecnologías: Una visión Contemporánea”

Resumen

En este trabajo se presenta una metodología para el reconocimiento de objetos, en particular para el desarrollo de aplicaciones de realidad aumentada, cuya aportación se centra en la creación de un sistema novedoso que reconoce piezas arqueológicas, sobreponiendo información y objetos virtuales de tal forma que para el ojo humano en apariencia la escena real y virtual son una misma, todo ello en tiempo real. El reconocimiento hace uso de información local inherente de cada objeto, lo cual ayuda a que cuando un objeto es ocluido parcialmente, este puede ser reconocido sin comprometer el funcionamiento del sistema. Esto nos permitió realizar experimentos en *in-situ* y *ex-situ* de un conjunto de piezas exhibidas en la sala cinco del Museo del Templo Mayor, simulando la situación cuando una persona ocluye parcialmente a una pieza. La información sobrepuesta es acerca de la pieza reconocida e incluye imágenes, vídeos, texto, elementos gráficos de señalización, mapas satélites y panoramas a nivel de calle del sitio donde se localizó dicha pieza.

Introducción

En la actualidad las tecnologías de la información forman parte de las herramientas empleadas para preservar, registrar, restaurar y difundir el patrimonio histórico del país. Especialmente nos interesa

aquellas que nos permiten el desarrollo de técnicas o metodologías para el reconocimiento y modelado de objetos, su entorno tridimensional donde fueron descubiertos, etc. Mediante ellas se pueden desarrollar diferentes tipos de aplicaciones que son de interés en áreas como robótica móvil, reconstrucción 3D de escenarios, Realidad Aumentada, recorridos virtuales de sitios de difícil acceso, etcétera.

La Realidad Aumentada (*RA*) es un área de investigación que ha venido en auge en los últimos treinta años, ya que brinda la posibilidad al usuario de poder observar objetos tridimensionales sobrepuestos en el mundo real a través de un dispositivo de captura de imágenes y despliegue de las mismas en una pantalla, tal y como un teléfono inteligente, *tablet* o computadora portátil. Esto nos permite diseñar diferentes formas de interacción del usuario con estos dispositivos y con ello explotar la experiencia de las personas que visitan los museos, esto se logra al proporcionarle de manera interactiva información de interés a cerca de alguna pieza exhibida, como por ejemplo, el lugar exacto y la forma en que fue localizada la pieza, ya sea a través de imágenes o la reconstrucción 3D que se haya realizado durante el levantamiento, su restauración y posterior exhibición en la sala, véase la Figura 1.

Los museos son lugares en donde cada una de sus salas se diseñan y acondicionan bajo ciertos estándares y reglamentos que permitan la conservación de las piezas exhibidas y exposición de las mismas, por lo que la actualización de la información de las cédulas informativas puede ser lenta, incluso introducir algún tipo de marcador para reconocer alguna pieza puede resultar en un proceso difícil o impráctico.

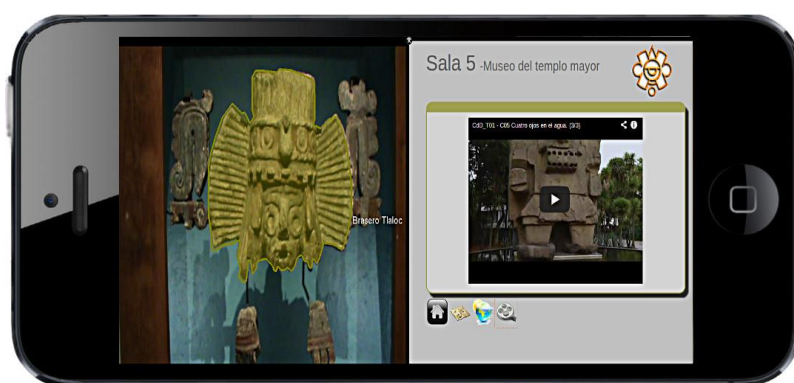


Figura 1. Información interactiva a través de una aplicación de Realidad Aumentada, permite una mayor experiencia a los visitantes de un museo.

La mayoría de los museos son frecuentados por muchos visitantes y es muy común que un visitante ocluya de manera parcial o total a la pieza. Por lo cual vimos la necesidad de desarrollar una metodología que permita integrar un sistema de *RA* a una sala de exhibición sin que este obligue alguna adecuación de la sala, es decir sin utilizar marcadores, sino información propia de la pieza y a su vez poder brindar información actualizada y que pueda visualizarse sobre la pieza aún que está sea ocluida de manera parcial por algún otro visitante. La información es visualizada por medio de una página *web*, ya que mediante de la tecnología *HTML-5* nos permite introducir imágenes, texto y contenido multimedia ya existente de algunas de las piezas de la sala cinco del Museo del Templo Mayor.

2.1 Contribuciones

Nuestras principales contribuciones en el trabajo que hemos desarrollado son:

- El desarrollo de una biblioteca que integra código especializado para el procesamiento de imágenes y generación de gráficos por computadora de manera estándar para diferentes sistemas operativos.
- El diseño de una metodología cuyo producto fue la creación de un sistema de Realidad Aumentada especializado en el reconocimiento de piezas arqueológicas.
- La realización de contenido virtual que combina tecnologías multimedia, servicios *web*, base de datos móvil, tal que el desarrollo sea modular, flexible y escalable a los requerimientos de diferentes tipos de museo.
- Inclusión de un sistema de RA con características novedosas y diferentes a los actualmente desarrollados en lo que respecta a arqueología.

Trabajo Relacionado

El reconocimiento de objetos es un área muy activa en las últimas décadas y han surgido diferentes metodologías que pueden ser categorizadas como:

- Ajuste de modelos geométricos.
- Modelos basados en apariencia.
- Desplazamiento de ventanas.
- Modelos basados en características.
- Bolsa de palabras.

Cada una de las metodologías han sido diseñadas para reconocer objetos específicos como son calles, edificios, personas, etc. Las primeras surgen en las décadas de 1950 y 1970, las cuales se caracterizaron en el uso de patrones de referencia del objeto a reconocer y técnicas de correlación para asociar dichos patrones presentes en las imágenes de prueba. Las siguientes metodologías basadas en apariencia de los objetos surgieron en la década de los noventa, cuya principal diferencia es la obtención de características globales estadísticamente invariantes y la extracción de las variables más relevantes por medio del análisis de sus componentes principales, el trabajo más representativo fue el desarrollado por *Viola y Jones (2001)*, cuya aportación es la introducción del concepto de imagen integral para reconocer rostros en tiempo real combinando el uso de ventanas de desplazamiento y modelos basados en apariencia.

Las metodologías más recientes surgen a partir de los años 2000, las cuales hacen uso de características locales invariantes a transformaciones de escala, perspectiva y orientación codificadas en un descriptor, el cual es un arreglo numérico que codifica información local de un punto invariante en una imagen, es decir un punto que siempre pueda ser localizado y que sea lo suficientemente diferente a otros puntos.

En ese lapso de tiempo surgen los descriptores *SIFT* (*Scale Invariant Feature Transform*, David Lowe , 1999) y *SURF* (*Speeded Up Robust Features*, Bay y Tuytelaars, 2006), los cuales han sido ampliamente usados en diferentes aplicaciones.

El reconocimiento de objetos es una parte importante en la realidad aumentada y forma parte de nuestro interés para reconocer piezas arqueológicas. El primer sistema de realidad aumentada fue creado por Ivan Sutherland en 1968. Sin embargo fue hasta el año de 1992 que el termino 'Realidad aumentada' fue acuñado gracias al trabajo de los ingenieros Caudell y Mizell. Dicho trabajo consistió en el desarrollo de un sistema para asistir a los operadores en labores de ensamblaje de piezas de aviones y con ello minimizar errores durante el proceso.

En el año de 1998 la compañía de *Sony* en conjunto con el ingeniero Jun Rekimoto introducen un nuevo método que emplea marcadores para sobreponer objetos virtuales en ello (*Rekimoto, 1998*). Esta técnica fue un parte aguas en los actuales sistemas de realidad aumentada y en el año 2006 *Sony* crea un vídeo juego *Eye of Judgement* que hace uso de los marcadores para la consola playstation, véase la Figura 2.

El uso de marcadores fue ampliamente usado y fue hasta el año de 1999 cuando Kato y Billinghurst crearon *Artoolkit*, siendo la primera biblioteca gratuita y especializada para el desarrollo de aplicaciones de realidad aumentada. Al ser gratuita se volvió muy popular entre los programadores y es posible encontrar diferentes aplicaciones en tiempo real tanto en dispositivos móviles o en computadores, (*Kato, Billinghurst, 1999*).

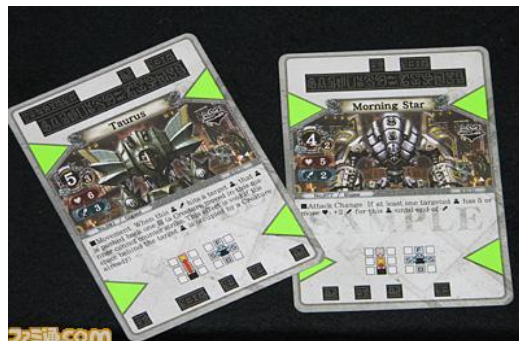
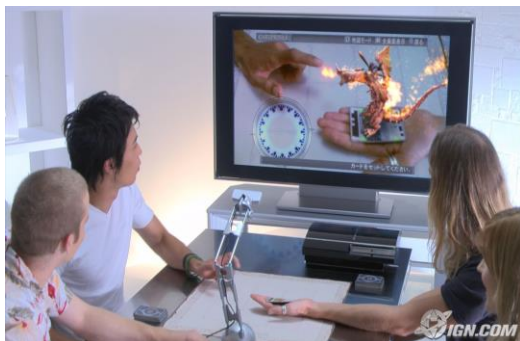


Figura 2. Ilustración del vídeo juego *Eye of Judgement*, en la primera imagen se muestra el juego en funcionamiento y en la segunda los marcadores empleados.

Imágenes: <http://www.gamingsurvival.com/2010/07/02/eye-of-judgement-ps3-to-close-online-servers-in-september/>

En años recientes Castle propone su sistema *PTAMM (Parallel Tracking and Multiple Mapping)*, el cual consiste en construir múltiples mapas con base en características naturales del entorno y hacer seguimiento de las mismas en paralelo, lo cual permite estimar la posición de un objeto en tiempo real y de manera eficiente, logrando con ello una nueva tendencia en los sistemas de realidad aumentada (Castle, 2009).

Recientemente, en abril de 2012 fue lanzado el proyecto de *Google Glass* y ofrece nuevas opciones para el desarrollo de futuros sistemas de realidad aumentada, ya que proporciona integración directa con el sistema operativo Android para el desarrollo de aplicaciones, una cámara de vídeo, acceso a internet inalámbrico, giroscopio, sensores de luz, pantalla para el despliegue de notificaciones, etc, véase la Figura 3.

En el área de arqueología y con base en la literatura revisada sólo se conoce un sistema de AR aplicado a esta disciplina, conocido como *Archeoguide (Vlahakis, 2002)* y propiamente no reconoce objetos, sino que hace uso de la posición del dispositivo a través del *GPS* y una conexión a una red local para desplegar en dicha posición información virtual asociada al sitio. Por esa razón se decidió proponer un nuevo sistema para reconocer piezas arqueológicas usan características inherentes a ellas, tal que pueda ser usado en tiempo real.



Figura 3. Imágenes acerca del diseño del Google Glass, en la primera se observa el dispositivo

ensamblado y en la segunda sin ensamblar.

Imágenes: <https://plus.google.com/u/0/+GoogleGlass/photos>

Reconocimiento de Objetos

El reconocimiento de objetos esta compuesto de diferentes etapas a partir de imágenes de referencia de los objetos a reconocer y de la escena en donde se encuentran dichos objetos, en nuestro caso las piezas de la sala 5 del Museo del Templo Mayor que corresponden al Dios Tláloc.

La primera etapa consiste en obtener los descriptores *SIFT* de las imágenes de referencia de cada uno de las piezas a reconocer y posteriormente almacenarlos sistemáticamente en una base de datos. La siguiente etapa es adquirir los descriptores de las imágenes de la escena a evaluar y relacionarlos por pares con los descriptores almacenados en la base de datos, para ello se hace uso de una técnica de clasificación denominada clasificador de vecino más cercano (*KNN*, *k-Nearest Neighbors*) y con base en el establecimiento de correspondencias identificar la región en donde se encuentran las piezas de nuestra base de datos. Posteriormente se procede a estimar una homografía por cada pieza, la cual nos permite colocar bajo una misma perspectiva o proyección las imágenes de referencia y la imagen de la escena. Con ello es posible estimar la posición de cada objeto en la escena y en dicha posición desplegar un objeto virtual sobrepuesto. Finalmente se procede a dibujar un contorno sobrepuesto a la pieza que funge como señalización al usuario y así realizar la interacción con el mismo al momento que la persona seleccione el contorno, cuando eso ocurre se despliega una página *web* con contenido multimedia acerca de la pieza reconocida, véase la Figura 4.

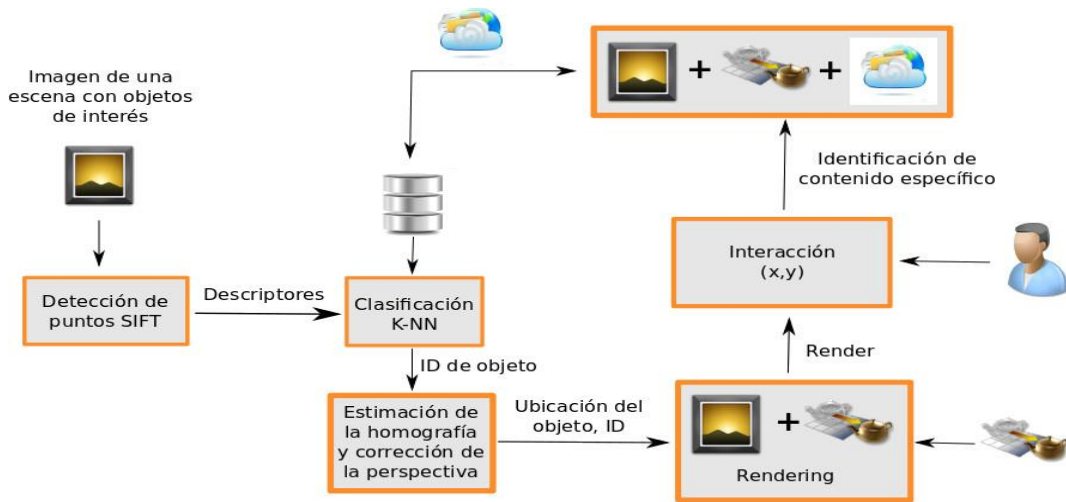


Figura 4. Ilustración de las etapas que componen el reconocimiento de objetos.

5.1 SIFT

Los descriptores *SIFT* fueron propuestos por David Lowe y su adquisición se divide en dos partes: La localización de puntos invariantes a diferentes escalas y la construcción del descriptor asociado a cada uno de ellos (David Lowe, 1999).

La localización de los puntos invariantes se obtiene por medio de convolución de una imagen con la aproximación del Laplaciano. Dicha aproximación se realiza por medio de la diferencia de funciones gaussianas de diferentes escalas y se define como:

$$D(x, y, \sigma) = \left(\frac{G(x, y, k\sigma) - G(x, y, \sigma)}{k\sigma - \sigma} \right) * I(x, y)$$

Donde $I(x, y)$ corresponde a un píxel de la imagen, la letra sigma al valor de la escala y el símbolo $*$ el operador de convolución. Mediante dicha aproximación se procede a derivar e igualar a zero para identificar sus máximos y mínimos, ya que en ellos se localizan los puntos invariantes que se requieren.

La construcción del descriptor se realiza al obtener orientaciones de gradiente del vecindario donde se encuentra un punto invariante identificado en la etapa anterior. El vecindario consiste a una región

cuadrada centrado al punto de dimensión de 16x16 píxeles y se particiona en cuatro subregiones de 4x4 píxeles. En cada subregión se calculan gradientes en ocho orientaciones o ángulos, y se procede a hacer un conteo para generar un histograma de los diferentes valores que se obtiene de los gradientes, logrando con ello la obtención de 128 valores diferentes, los cuales corresponden al descriptor, es decir, el descriptor se compone de un arreglo de 128 valores, véase la Figura 5.

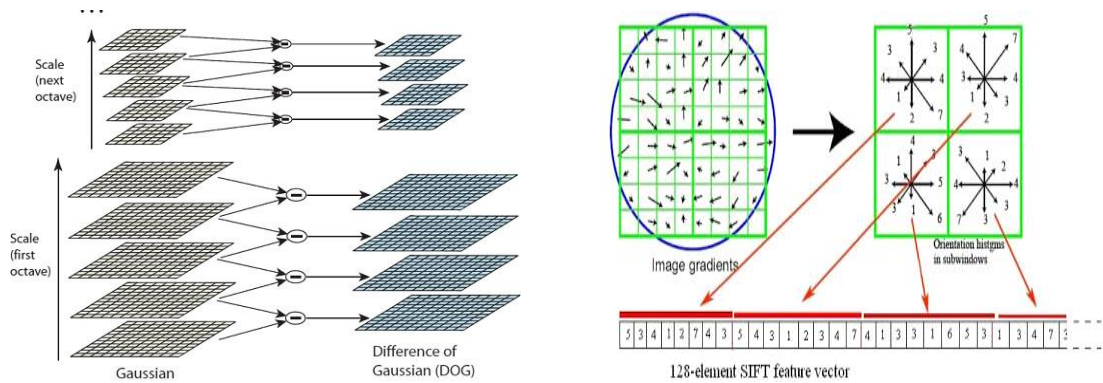


Figura 5. Imágenes acerca de la obtención del descriptor. En la primer imagen se muestra la diferencia de las imágenes convolucionadas con funciones gaussianas a diferentes escalas y en la segunda se muestra la construcción del descriptor por medio de orientaciones de gradiente.

Imágenes: <https://picasaweb.google.com/lh/photo/vyaYFzPsGz6RzldJnvEaDQ>

5.2 Clasificación

El problema de clasificación se puede describir como un problema de búsqueda dentro de un conjunto de entrenamiento, es decir, si suponemos que cada descriptor es único y existe una copia en el conjunto de entrenamiento, el objetivo es localizar la copia, devolver tanto el descriptor como el identificador del objeto al cual pertenece y ello establecer una correspondencia entre el descriptor de consulta y el obtenido por medio del proceso de búsqueda.

El conjunto de entrenamiento se define como conjunto finito de objetos de referencia, en donde cada objeto es representado por un conjunto discreto de descriptores SIFT y por una etiqueta que los identifica de los demás. Todos los descriptores y etiquetas son almacenados en una tabla cuyas filas

representan al descriptor y identificador del objeto.

Entonces para resolver el problema de búsqueda es empleado el clasificador *KNN*. La clasificación consiste en encontrar los *K* vecinos más cercanos al descriptor de consulta y por medio de votación indicar a que pieza pertenece. La distancia euclidiana es utilizada para medir la cercanía entre los descriptores. Esta sencilla forma de clasificación tiene sus inconvenientes, ya que si la dimensión del conjunto de entrenamiento es grande la consulta puede ser muy tardada, puesto que para cada descriptor de consulta se tendría que calcular la distancia de todos los descriptores hacia el descriptor de consulta, por ende la eficiencia tiende a reducirse y para aumentarla se procede a realizar un particionamiento del espacio *n*-dimensional en donde pertenecen los descriptores por medio de árboles KD aleatorios.

Un árbol KD es una estructura de datos empleada para almacenar en cada uno de sus nodos el valor de un umbral de referencia, un arreglo numérico o descriptor, el nodo padre al cual pertenece y dos nodos hijos (el número de hijos puede variar al igual que el número de umbrales) , teniendo en total el mismo número de nodos como de descriptores. El umbral es empleado para indicar si un valor se encuentra hacia el lado de uno de sus nodos hijo. El valor del umbral se calcula a partir de la media o mediana de cada una de sus componentes del descriptor y cada nivel del árbol corresponde a una componente diferente a su nivel predecesor.

Ejemplo: Supongamos que tenemos un conjunto de descriptores de dos dimensiones definido como $P = \{ (2, 3), (5, 4), (9, 6), (4, 7), (8, 1), (7, 2) \}$, El primer nivel del árbol corresponde a la componente 'x', el siguiente a la componente 'y', luego a la 'x' y así sucesivamente hasta que todos los descriptores estén almacenados en un nodo del árbol. Para crear el primer nivel correspondiente a nodo raíz se procede a calcular la mediana de la componente 'x', el valor de la mediana funge como umbral y con base a nuestro conjunto de entrenamiento corresponde al valor de 7. Mediante este umbral se particiona el conjunto en dos subconjuntos, el primero cuyas componentes en 'x' son menores a siete y las mayores o

iguales a siete. El siguiente nivel se construye igual forma solo que empleando la componente 'y' hasta que cada partición tenga un único descriptor, véase la Figura 6.

K-D tree

— pointList = [(2,3), (5,4), (9,6), (4,7), (8,1), (7,2)]

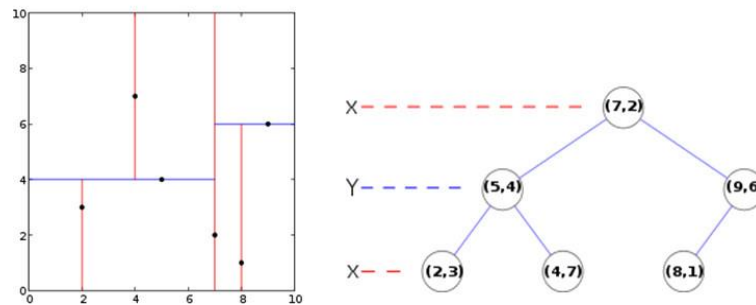


Figura 6. Ilustración en donde se muestra la generación de un árbol KD para un conjunto de descriptores de dos dimensiones.

Imagen: <http://mind.kaist.ac.kr/Francis/project2.html>

El trabajo propuesto por Muja y Lowe consiste en particionar el espacio de los descriptores por medio de árboles KD aleatorios y en la realización de clasificador *KNN*, (Muja, Lowe, 2009). El particionamiento se realiza por medio cuatro árboles KD que constan de cinco nodos donde cada uno se obtienen por medio de las cinco varianzas máximas de cada una de las componentes del conjunto de descriptores, posteriormente se calculan las medias de dichas componentes y son fijadas como el umbral de partición. Los subsiguientes niveles se obtienen de igual manera y se prosigue hasta haber particionado todo el conjunto de descriptores. Este particionamiento hace que la complejidad de la búsqueda sea del orden de $O(\log n)$, con lo cual la eficiencia se incrementa.

La búsqueda se realiza al visitar cada uno de los niveles de los árboles únicamente descendiendo hacia las ramas acotadas por el valor de los umbrales de los nodos hijos hasta encontrar el descriptor que más se parece al descriptor de consulta, esto se logra al tener una distancia pequeña o igual a cero.

Por lo tanto, dado un conjunto de descriptores *SIFT* correspondientes a las piezas a reconocer, el clasificador *KNN* devolverá las correspondencias de las piezas reconocidas y con ellas estimar sus

homografías correspondientes.

La homografía es una transformación lineal empleada para transformar un plano en perspectiva a otra. En nuestro caso suponemos que los puntos invariantes de cada descriptor se encuentran sobre un plano y se quiere encontrar la homografía que asocia puntos en la imagen de la escena con puntos de la imagen de nuestro objeto reconocido. La estimación de la homografía se realiza por medio del algoritmo de transformación directa y para más detalles puede verse en *Hartley y Zisserman (2004)*.

Dado que se ha obtenido la homografía a partir de correspondencias de puntos entre planos, es posible aplicar dicha transformación al contorno que rodea la pieza de referencia, tal que aparezca con la misma perspectiva en la región reconocida de la imagen de prueba, véase la Figura 7. Haciendo que con dicho contorno sirva de señalización únicamente asignándole un color diferente a cada una de las piezas que conforman nuestro conjunto de entrenamiento.

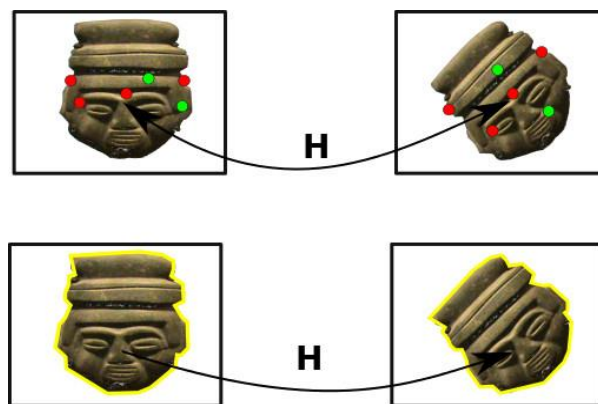


Figura 7. Ilustración en donde se muestra el concepto de homografía entre dos imágenes de diferente perspectiva, así como su uso para proyectar un contorno sobre la pieza reconocida.

5.3 Interacción

La interacción es una característica importante en el sistema propuesto, ya que se puede proporcionar información de la pieza a través de vídeos, imágenes y no únicamente con un breve texto que

normalmente las cédulas informativas proporcionan. Esto permite que el visitante disfrute de una mayor experiencia en su visita al museo, además de enriquecer su bagaje cultural acerca de las piezas exhibidas y darle una idea de la forma de vida que tenían las culturas que construyeron dichas piezas. En nuestro caso fue integrar tecnologías informáticas y ciencias exactas para revitalizar la riqueza cultural de las culturas mesoamericanas que habitaron en México.

La interacción con el usuario se realiza al proyectar un contorno sobrepuesto en la pieza visualizada en la pantalla en tiempo real, generando con ello que el usuario fije su atención y seleccione el contorno, produciendo con ello una respuesta del sistema ante dicha acción. Este evento consiste en desplegar de manera dinámica una página *web* con contenido multimedia como son mapas a nivel de calle en donde se localizaron algunas de las piezas exhibidas en la sala 5 del museo del Templo Mayor.

El uso de páginas web nos permite centralizar y sistematizar la actualización de la información de manera sencilla, véase la Figura 8. Esto permite que gente experta se encargue de elaborar los contenidos de cada pieza, sin que tenga que hacer cambios al sistema a nivel de programación y con ello proporcionarle forma sencilla de sincronizar automáticamente los cambios en los contenidos con los diferentes usuarios del sistema. Es decir el sistema puede ser usado por diferentes usuarios de manera asíncrona, logrando que tanto visitantes como expertos puedan interactuar de manera

transparent



e.
Figura 8. Imagen de un mural del dios Tlaloc en donde se muestra el contorno que es empleado

como señalización para fijar la atención del usuario.

5.4 Evaluación

La evaluación del sistema es importante ya que se desea tener un porcentaje alto de reconocimiento o clasificación de las piezas ante condiciones propias de un museo, como son iluminación, cambios de perspectiva, escala y oclusiones parciales. Para ello suponemos que podemos acotar por medio de un cuadrilátero la posición real de la pieza y la posición estimada de la misma, mientras sea mayor área de intersección mayor será el porcentaje de clasificación, véase la Figura 9. Entonces con base en una secuencia de imágenes cuya posición real de una pieza sea conocida, se puede evaluar imagen por imagen el área de intersección y con ello calcular falsos positivos o falso negativos.

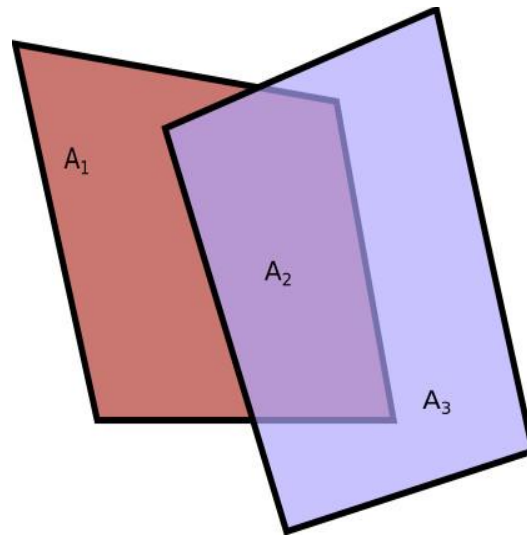


Figura 8. Ilustración en donde se muestra el área de intersección entre dos cuadriláteros.

La primera medida básicamente es el área de intersección entre los cuadriláteros y se define como la división del área de intersección entre la suma de las áreas de los cuadriláteros menos su intersección, el resultado es un valor en el intervalo cero y uno. Mientras el valor se encuentre más cercano a cero indica que no se reconoció la pieza y cercano a uno que fue reconocida correctamente. Se fija un umbral de 0.8 para indicar que realmente se reconoce el objeto ya que se desea un alto porcentaje de

clasificación, si no se rebasa dicho umbral el objeto no es reconocido.

Dado que se conoce la posición real del objeto dentro de la secuencia de imágenes, se puede contabilizar el número de falsos positivos, de falsos negativos, verdaderos positivos y de verdaderos negativos y con ello fijar una segunda medida de evaluación, la cual se conoce como Medida-F y se define como :

$$F_{\beta} = (1 + \beta^2) \cdot \frac{\text{precision} \cdot \text{recall}}{(\beta^2 \cdot \text{precision}) + \text{recall}}$$

Donde *precision* y *recall* son dos medidas que se obtienen con base en los falsos y verdaderos positivos. Al igual la Medida-F se encuentra acotada en el intervalo cero y uno.

Resultados

En esta sección se muestran dos aplicaciones desarrolladas como parte de nuestro trabajo de investigación. La metodología propuesta en este documento, nos permite adaptarla sencillamente a diferentes aplicaciones. Cada aplicación resuelve un problema en específico y con ello se abarca un amplio panorama de sus posibles usos.

La primera aplicación se encarga de colorear una pieza que ha sufrido deterioro en su pigmentación, debido al paso del tiempo, a las condiciones ambientales a las que fue expuesta y al daño a causa de una acción humana. Esta aplicación es muy útil tanto para visitantes como a los arqueólogos, ya que permite visualizar como era la pigmentación de la pieza antes de su deterioro.

La segunda aplicación corresponde al sistema de realidad aumentada desarrollado como parte de nuestro objetivo. Este se encarga de detectar varias piezas arqueológicas que tienen mucha textura y bordes. Las piezas a detectar son exhibidas en la sala 5 del Museo del Templo Mayor y forman parte de la colección de piezas alusivas al dios Tláloc. Cada pieza que es detectada, el sistema se encarga de

sobreponer una región de interés coloreada con un color específico al contorno del objeto. La región de interés coloreada sirve para señalar donde el usuario puede interactuar con el sistema. Cuando éste interactúa con el sistema, se le proporciona un contenido informativo acerca de la pieza seleccionada.

6.1 Coloreado de piezas arqueológicas

Los estudios de pigmentación de piezas arqueológicas, permiten conocer los colores que tuvo una pieza cuando ésta fue fabricada. Debido al paso del tiempo estos pigmentos se deterioran y son pocos visibles.

Un ejemplo de ello es el monolito de Coyolxauhqui expuesto en el Museo del Templo Mayor. Debido al deterioro los pigmentos originales han desaparecido casi por completo y son fáciles de visualizar.

Con base a estudios de pigmentación se pudo determinar como lucía el monolito al momento de su creación. Es por ello que se desarrolló una aplicación de *RA* que ejemplifiqué el coloreado de una pieza arqueológica.



Figura 9. Ejemplo del coloreo de una pieza arqueológica, en la imagen izquierda se muestra un cartel de Coyolxauhqui, el cual corresponde a como esta pieza luce actualmente y en la figura derecha se muestra como estaba coloreada antes de su deterioro.

Normalmente un museo es un lugar concurrido por muchas personas interesadas en conocer las colecciones de objetos o piezas que se exhiben. En las salas de exhibición es común encontrar personas que ocultan una o varias piezas. Entonces, si otro visitante desea ver como era la

pigmentación de la pieza antes de su deterioro, este únicamente logra ver una parte de la pieza ocluida. En la Figura 9 se muestra una escena típica que podemos encontrar en un museo.

En la Figura 9a se realiza una recreación de una sala en la cual se exhibe una pieza arqueológica. La pieza corresponde a un cartel que contiene la imagen del monolito Coyolxauhqui descubierto por accidente en 1978 en lo que actualmente es el Museo del Templo Mayor. En la imagen se puede observar que observa que el cartel de Coyolxauhqui es parcialmente ocluido por una persona. Sin embargo, si un visitante utiliza la aplicación desarrollada para visualizar la pigmentación que tenía el monolito. La aplicación logra detectar la imagen del cartel y ésta es coloreada con los colores reales del monolito, ver Figura 9b.

En la Figura 10a se realiza una recreación de una sala en la cual se exhibe la piedra del Sol. La pieza recreada corresponde a un cartel que contiene la imagen de la piedra del Sol Azteca descubierto en 1479 en lo que actualmente es el Zócalo de la Ciudad de México. En la imagen se puede observar que observa que el cartel de la piedra del Sol es parcialmente ocluido por una persona. Sin embargo, cuando un visitante utiliza la aplicación. La aplicación logra detectar la imagen del cartel y ésta es coloreada como lucía antes de su deterioro, ver Figura 10b.



Figura 9. Ejemplo del coloreo de la piedra del Sol. En la figura izquierda se muestra un cartel con la apariencia actual de la piedra del Sol y en la figura derecha se muestra como lucía antes de su deterioro.

6.2 Información asistencial a visitantes de museos

Uno de nuestros objetivos fue crear un sistema de RA aplicado a la arqueología. Éste se desarrolló con la finalidad de detectar e identificar piezas arqueológicas de la sala 5 del Museo del Templo Mayor. Éstas piezas están asociadas a las deidades acuático-terrestres y de la fertilidad. Cuando un usuario interactúa con el sistema, este puede obtener información complementaria a la cédula informativa de una pieza exhibida. La información que se presenta le permite conocer la cosmovisión que los aztecas tenían antes de la conquista española.

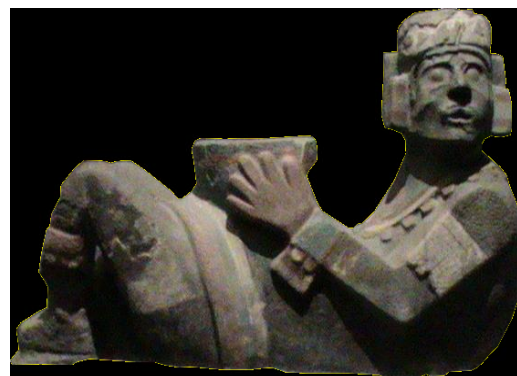
Los contenidos acerca de cada pieza se realizaron con base a los documentos que nos proporcionaron. El sistema hace uso de tecnologías *web* para enriquecer el contenido mediante mapas satelitales, visión de calle, información del estado del clima y vídeos. Para facilitar la detección de cada pieza, a cada una se le asigna un color para señalar la diferencia entre cada una de ellas. En total utilizamos diez piezas representativas de la sala, de las cuales se muestran en la Figura 10 cinco de ellas.



a)

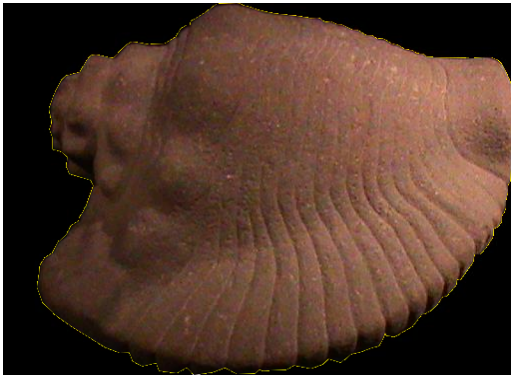


b)



c)

d)



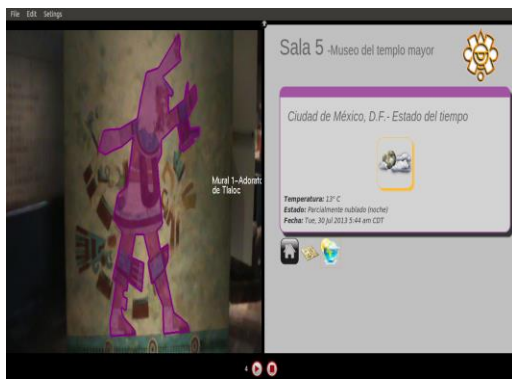
e)

Figura 10. Piezas arqueológicas exhibidas en la sala 5 del Museo del Templo Mayor , las cuales forman parte de las piezas a reconocer.

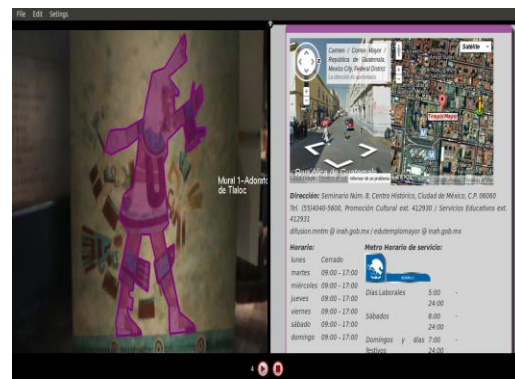
Para probar la funcionalidad del sistema, generamos un vídeo del recorrido realizado en la sala del museo. Con dicho vídeo se realizó la interacción con el sistema, cuyos resultados se muestran en las imágenes de la Figura 11.

En la imagen 11a se muestra uno de los dos murales encontrados en el adoratorio de Tláloc. Pudimos constatar que la cédula informativa de ésta pieza presenta información muy breve acerca de ella. Nuestro sistema se encarga de detectarla en el vídeo y la señala con una región de interés en color púrpura. Al interactuar con dicha región el sistema proporciona información acerca del mural, proporciona un mapa satelital de la dirección y horarios del Museo Templo Mayor. Además de la información de la pieza, es posible proporcionar otro tipo información, por ejemplo, el estado actual del clima en la Ciudad de México, ver Figura 11b.

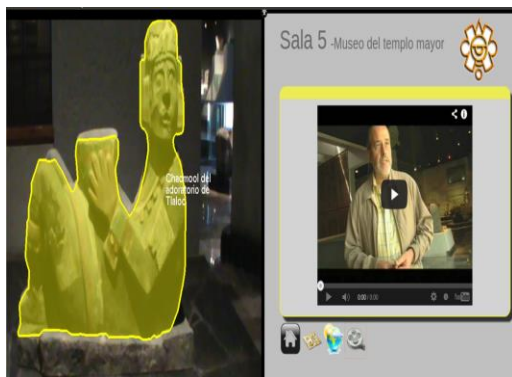
Una pieza importante de la sala es el brasero de Tláloc debido a su buen estado de conservación. En la Figura 11f se muestra que el contenido informativo puede incluir un vídeo acerca de la pieza. El contenido informativo del brasero incluye imágenes y texto para causar un mayor interés de conocer más acerca de la pieza. La región de interés aparece coloreada en verde-amarillo, esta se puede ver en la Figura 11e.



a)



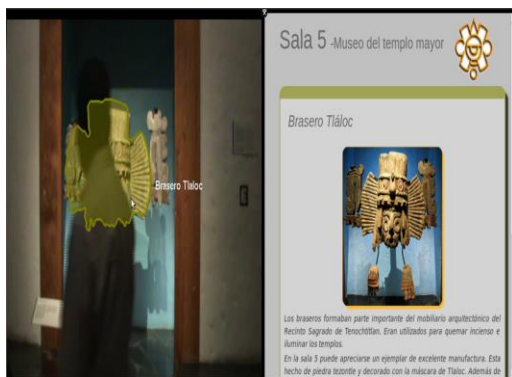
b)



c)



d)



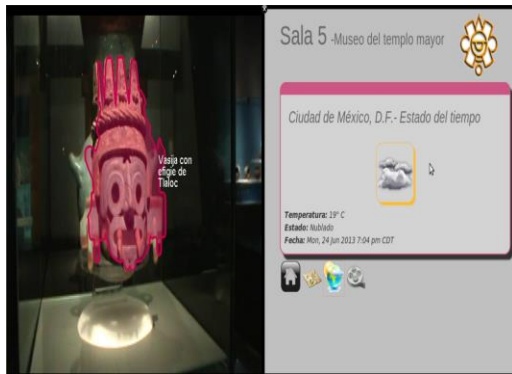
e)



f)

Figura 11 Piezas arqueológicas exhibidas en la sala 5 del Museo del Templo Mayor , las cuales muestran el funcionamiento del sistema.

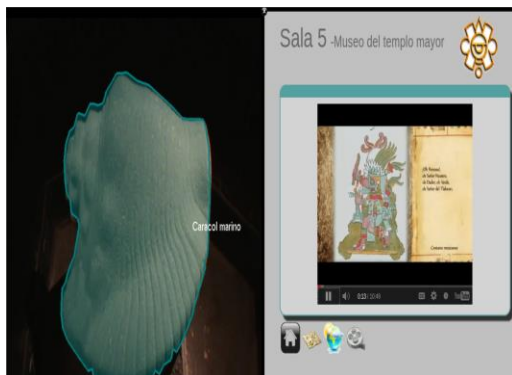
El chac mool es un tipo de esculturas que aparecen al principio del Período Posclásico en diversos sitios de México. En la Figura 11c se observa coloreado en color amarillo cuando este es seleccionado. El contenido que se despliega incluye un vídeo y un mapa satelital donde se indica donde fue descubierto.



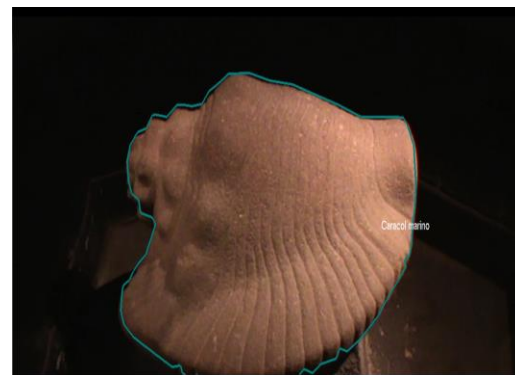
a)



b)



c)



d)

Figura 12 Piezas arqueológicas exhibidas en la sala 5 del Museo del Templo Mayor , las cuales forman parte de las piezas a reconocidas.

Otra pieza importante de la sala es el caracol marino. En la Figura 12c se observa en color azul la región de interés que se ajusta al contorno de la pieza. Esta región permite al usuario percibir una zona donde interactuar, cuando esto ocurre se despliega el contenido de la pieza. Este contenido consiste de un vídeo que habla sobre la deidad Tláloc y el significado de cada uno de sus rasgos característicos, ver Figura 12b.

Por último en la 12a mostramos en color magenta la vasija con efígie de Tláloc. Su contenido consiste en texto, imágenes y el estado del clima de la Ciudad de México.

Aunque este sistema está hecho ad-hoc para la detección e identificación de piezas arqueológicas, tiene la flexibilidad de ajustarlo a otros tipos de objetos de exhibición, como pueden ser pinturas o cierto tipo de esculturas. Es flexible, novedoso, escalable y con base a nuestra inspección bibliográfica, sólo se

conoce un sistema de *RA* aplicado a la arqueología, se conoce como *Archeoguide* (Vlahakis, 2002). Éste difiere completamente a nuestro sistema desarrollado, ya que no detecta objetos, sino que identifica la posición del dispositivo que ejecuta el sistema con base a un GPS y su conexión a una red local.

Conclusiones

El trabajo presentado en este documento trató sobre el reconocimiento de objetos en escenas complejas, principalmente para aplicaciones de realidad aumentada enfocadas a la arqueología. El reconocimiento se realizó al representar a un objeto por medio de características locales. Al estimar la posición del objeto en la escena, nos permitió sobreponer información virtual específica, logrando con ello un sistema de realidad aumentada.

La revisión de la literatura nos permitió elaborar una conjunto de métodos y algoritmos para llevar a cabo el reconocimiento de objetos, en escenas del museo y en escenas controladas, estableciendo con ello una metodología para llevar a cabo la estimación de la posición del objeto.

Uno de los objetivos del presente trabajo involucró el desarrollo de un sistema de software interactivo, el cual fue empleado en la realización de los experimentos. En ellos se implementaron los algoritmos basados en los conceptos presentados en la sección de reconocimiento de objetos del presente documento, anexando otras rutinas para el diseño y manejo de eventos de la interfaz gráfica de usuario. Este sistema de software además de permitir interactividad con el usuario, logró integrar adecuadamente las bibliotecas *OpenCV*, *OGRE* y *OpenGL*, obteniendo con esto una aplicación de software que puede servir como base para el desarrollo de otros sistemas de *RA*.

Referencias

- Castle, R. O. (2009). Simultaneous Recognition, Localization and Mapping for Wearable Visual

Robots. Tesis de doctorado, University of Oxford.

- Caudell, T. y Mizell, D. (1992). Augmented reality: an application of heads-up display technology to manual manufacturing processes. En System Sciences, 1992. Proceedings of the Twenty-Fifth Hawaii International Conference on, volumen ii, paginas 659–669 vol.2.
- Hartley, R. I. y Zisserman, A. (2004). Multiple View Geometry in Computer Vision. Cambridge University Press, ISBN:0521549518.
- Kato, H. y Billinghurst, M. (1999). Marker tracking and hmd calibration for a video-based augmented reality conferencing system. En Augmented Reality, 1999. (IWAR '99) Proceedings. 2nd IEEE and ACM International Workshop, pages 85–94.
- Lowe, D. G. (1999). Object recognition from local scale-invariant features. En Proceedings of the International Conference on Computer Vision-Volume 2 - Volume 2, ICCV '99, paginas 1150–1157, Washington, DC, USA. IEEE Computer Society.
- Muja, M. y Lowe, D. G. (2009). Fast approximate nearest neighbors with automatic algorithm configuration. International Conference on Computer Vision Theory and Application VISSAPP'09), paginas 331–340. INSTICC Press.
- Rekimoto, J. (1998). Matrix: A realtime object identification and registration method for augmented reality. Proceedings of the Third Asian Pacific Computer and Human Interaction, APCHI '98, page 63, Washington, DC, USA, IEEE Computer Society.
- Sutherland, I. E. (1968). A head-mounted three dimensional display. Proceedings of the December 9-11, 1968, fall joint computer conference, part I, AFIPS '68 (Fall, part I), pages 757–764, New York, NY, USA. ACM.
- Viola, P. y Jones, M. (2001). Rapid object detection using a boosted cascade of simple features.

- Vlahakis, V., Ioannidis, N., Karigiannis, J., Tsotros, M., Gounaris, M., Stricker, D., Gleue, T., Dhne, P., y Almeida, L. (2002). Archeoguide: An augmented reality guide for archaeological sites. *IEEE Computer Graphics and Applications*, 22(5):52–60.