# Síntesis de imágenes a partir de imágenes reales de una escena mediante un algoritmo parcialmente autónomo*

## Image Synthesis from Real Scene Images by Means of a Partially Autonomous Algorithm**

## Síntese de imagens a partir de imagens não retificadas de uma cena***

*Arturo Fajardo-Jaimes*****
*Pedro Raúl Vizcaya-Guarín******

* Fecha de recepción: 11 de mayo de 2009. Fecha de aceptación para publicación: 23 de julio de 2009. Este artículo se deriva del proyecto de investigación *Síntesis de imágenes a partir de vistas de una escena*, desarrollado por el Departamento de Electrónica de la Pontificia Universidad Javeriana.

** Submitted on May 11, 2009. Accepted on July 23, 2009. This article results from the research project on *Image Synthesis of Views from a Scene*, developed by the Electronics Department at the Pontificia Universidad Javeriana.

*** Data de recepção: 11 de maio de 2009. Data de aceitação para publicação: 23 de julho de 2009. Este artigo deriva do projeto de pesquisa em síntese de imagens a partir de vistas de una cena, desenvolvido pelo Departamento de Eletrônica da Pontifícia Universidade Javeriana.

**** Ingeniero electrónico. Magíster en Ingeniería Electrónica, Pontificia Universidad Javeriana, Bogotá, Colombia. Profesor asistente de la Pontificia Universidad Javeriana. Correo electrónico: fajardoa@javeriana.edu.co.

***** Ingeniero electrónico, Pontificia Universidad Javeriana, Bogotá, Colombia. Máster y doctor en Ingeniería Eléctrica, Rensselaer Polytechnic Institute, New York, Estados Unidos. Profesor titular de la Pontificia Universidad Javeriana. Correo electrónico: pvizcaya@javeriana.edu.co.

## Resumen

En este artículo se presenta un algoritmo de síntesis de imágenes que brinda una solución al problema de sintetizar vistas de una escena real tomadas por una cámara virtual, ubicada entre dos cámaras reales. En particular, se presenta cómo bajo ciertas condiciones de las escenas un par de vistas base es suficiente para determinar todo el conjunto de vistas posibles sobre la línea que une sus centros ópticos, conocida como la línea base, sin necesidad de reconstruir explícitamente un modelo en 3D. Los resultados experimentales muestran cómo el algoritmo funciona adecuadamente en escenas simples (compuestas por un objeto de geometría sencilla, opacos y sin oclusiones).



## Palabras clave

Algoritmos, procesamiento de imágenes, cámaras fotográficas digitales.



## Abstract

This paper presents an image synthesis algorithm for rendering views of a real scene taken with a virtual camera which is located between two real cameras. Specifically, this paper presents how, under certain conditions, a pair of views is enough to obtain a full set of possible views by following the line that joins their optical centers (baseline) without reconstructing explicit 3D models. Test results show the viability of the proposed algorithm for simple scenes (an object with a simple geometry, which is opaque, and does not present any occlusions).



## Key words

Algorithms, image processing, digital cameras.



## Resumo

Neste artigo apresenta-se um algoritmo de síntese de imagens que oferece uma solução ao problema de sintetizar vistas de uma cena real tomadas por uma câmera virtual, localizada entre das câmeras reais. Em particular, apresenta-se como sob certas condições das cenas um par de vistas base é suficiente para determinar todo o conjunto de vistas possíveis sobre a linha que une seus centros óticos, conhecida como a línea base, sem necessidade de reconstruir explicitamente um modelo em 3D. Os resultados experimentais mostram como o algoritmo funciona adequadamente em cenas simples (compostas por um objeto de geometria simples, opacos e sem oclusões).



## Palavras chave

Algoritmos, processamento de imagens, câmaras fotográficas digitais.

## Introducción

La solución al problema de sintetizar imágenes a partir de vistas conocidas de una escena abre vastas posibilidades en cuanto al análisis de escenas complejas presentes en la vida cotidiana (Inamoto y Saito, 2007). En el caso particular de escenas de seguridad donde se tengan varias cámaras fijas, se podrían encontrar vistas no observadas por alguna de las cámaras, lo cual permitiría la posterior identificación de una persona. Para abordar el problema de la predicción de nuevas vistas a partir de vistas base se puede obtener el modelo tridimensional de una escena para después reproyectarlo y así sintetizar nuevas vistas (Kubota *et al.*, 2006; Zheng y Wu, 2001). Las principales desventajas de este tipo de solución son la complejidad y la acumulación de error en que se incurre al reconstruir el modelo en 3D.

En el SIGGRAPH 92, Beier y Nelly (1992) introdujeron una técnica de procesamiento de imágenes llamada conformación (*morphing*) para la metamorfosis de una imagen a otra. Esta técnica permitió generar nuevas vistas a través de la interpolación lineal de puntos correspondientes en ambas imágenes (Chen y Williams, 1993). Dicha investigación se concentró en encontrar qué interpolación producía vistas físicamente válidas de la escena, es decir, aquella que simula la vista producida por una cámara real en otra posición, donde encontraron que sólo algunas vistas interpoladas resultaban serlo. Seitz y Dyeer (1995 y 1996) demostraron que sólo bajo ciertas condiciones de la geometría epipolar la interpolación lineal produce vistas físicamente válidas.

En este artículo se presenta el desarrollo y la validación de un algoritmo implementado en Matlab®, que permite sintetizar nuevas vistas físicamente válidas de una escena a partir de vistas conocidas de esta, basándose en el método propuesto por Seitz y Dyer. Se empieza por exponer brevemente el modelo de la cámara utilizado y la geometría epipolar, para después describir el algoritmo. Finalmente, se ilustran algunos resultados relevantes y las principales conclusiones a las que se llegaron al realizar la investigación, junto con algunas sugerencias para mejorar el desempeño del algoritmo en trabajos posteriores. El algoritmo que se presenta se diferencia de los algoritmos existentes por ser

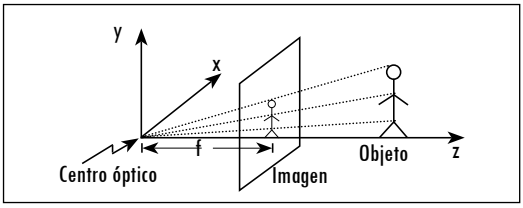
parcialmente autónomo, ya que una vez calibradas las cámaras, sintetiza las vistas sin interactuar con el usuario.

## 1. Desarrollo y métodos

### *1.1 Modelo de la cámara*

La proyección en perspectiva (Nalwa, 1993) es la proyección de puntos tridimensionales del espacio sobre una superficie bidimensional por medio de líneas rectas que pasan a través de un solo punto, llamado el centro óptico. En la Figura 1 se observa la formación de la imagen.

Figura 1. Proyección en perspectiva




Fuente: Nalwa, 1993.

La distancia entre el centro óptico de la cámara y el plano imagen es conocida como *distancia focal* (f). Del modelo de proyección perspectiva se obtiene que:

$$x = f * \frac{X}{Z} \text{ y } y = f * \frac{Y}{Z} \tag{1}$$

Donde $x$ y $y$ son las coordenadas del punto en la imagen y $X$, $Y$ y $Z$ son las coordenadas del punto en 3D. Estas ecuaciones son no lineales, lo cual impide su formulación de forma matricial. Sin embargo, como se expone en (Nalwa, 1993; González, 2000), al reescribir el punto 3D en coordenadas homogéneas como $\{kX\ kY\ kZ\ k\}^T$, donde $k$ es una constante arbitraria, la proyección de un punto tridimensional se puede escribir como sigue:

$$Pw_h = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \frac{1}{f} & 0 \end{bmatrix} \begin{bmatrix} kX \\ kY \\ kZ \\ k \end{bmatrix} = \begin{bmatrix} kX \\ kY \\ kZ \\ \frac{kZ}{f} \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = m_h \tag{2}$$

Donde $w_h$ representa un punto en tercera dimensión en coordenadas homogéneas, y $m_h$, su proyección en el plano imagen en coordenadas homogéneas.

La matriz $P$ se denomina *matriz de transformación perspectiva* (MPP). La tercera componente del vector resultante carece de significado, por lo cual en muchas aplicaciones la tercera fila de dicha matriz es eliminada. La ecuación es válida si el sistema de coordenadas del mundo real y el del plano imagen son coincidentes; sin embargo, en la mayoría de aplicaciones se presenta una situación donde los dos sistemas no son coincidentes, por ello es necesario hacer coincidir los sistemas de coordenadas. La MPP bajo estas condiciones está descrita por (González, 2000; Fusiello *et al.*, 1999).

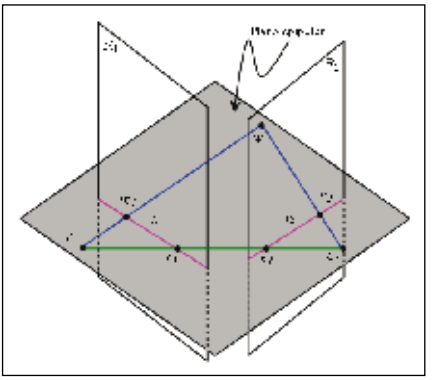
$$\mathrm{P} = A[R|t] \tag{3}$$

La matriz $A_{3x3}$ contiene los parámetros intrínsecos de la cámara, mientras que los parámetros extrínsecos de la cámara (posición y orientación) se encuentran codificados en la matriz de rotación $R_{3x3}$ y en el vector de translación $t_{3x1}$.

### *1.2 Geometría epipolar*

La geometría epipolar es la construcción básica que relaciona dos imágenes de una misma escena. En la Figura 2, sea $c_1$ el centro óptico de la primera cámara (izquierda) y sea $c_2$ el centro óptico de la segunda cámara (derecha), la línea que forman $c_1$ y $c_2$ se proyecta en los planos $R_1$ y $R_2$ en dos puntos llamados epipolos ($e_1$ y $e_2$). Las líneas pertenecientes a los planos $R_1$ y $R_2$, que pasan por los epipolos, se llaman líneas epipolares ($l_1$, $l_2$). El plano definido por el punto 3D y los centros ópticos se llama *plano epipolar*; este plano contiene también los epipolos y la proyección del punto 3D en el plano imagen de ambas cámaras, $m_1$ y $m_2$, los cuales constituyen un par de puntos correspondientes.

**Figura 2. Geometría epipolar**

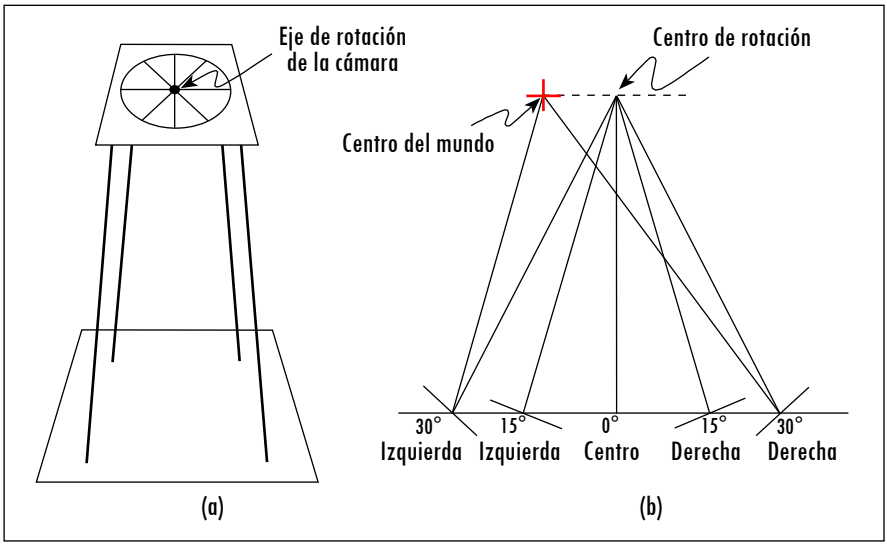


Fuente: presentación propia de los autores.

### *1.3 Montaje físico*

Para obtener vistas se diseñó un montaje físico (Córdoba *et al.*, 2002), que consistió en una base metálica ilustrada en la Figura 3a, la cual permitió obtener vistas de la escena en varias posiciones con una sola cámara, como se muestra en la Figura 3b. El uso de una sola cámara garantizó que los parámetros intrínsecos involucrados en la adquisición de cada vista fueran los mismos.

**Figura 3. Montaje físico: (a) estructura de la base; (b) posibles posiciones de la base**



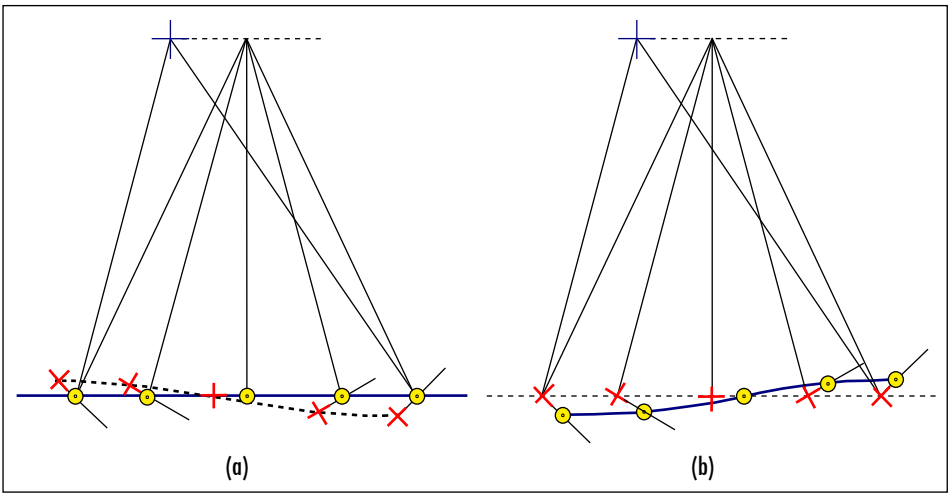

Fuente: presentación propia de los autores.

Como el centro óptico de la cámara no coincide con su eje de rotación, al rodarla sobre su eje y desplazar la base horizontalmente, su centro óptico forma una trayectoria curva ilustrada en la Figura 4a. Como el movimiento de los centros ópticos debe describir una línea recta para que coincida con la trayectoria teórica planteada en el método de síntesis de imágenes, el montaje físico fue ajustado para tratar que la trayectoria del centro mecánico de rotación de la cámara permitiera que la trayectoria descrita por el centro óptico de la cámara fuera en línea recta, como se ilustra en la Figura 4b.

### *1.4 Síntesis de imágenes*

En el algoritmo planteado en este artículo se utilizan técnicas que combinan interpolaciones bidimensionales de forma y color para crear efectos de transición entre imágenes. Estas se conocen como técnicas de conformación (*morphing*) y

Figura 4. Trayectorias descritas por el centro óptico con respecto al eje mecánico de rotación

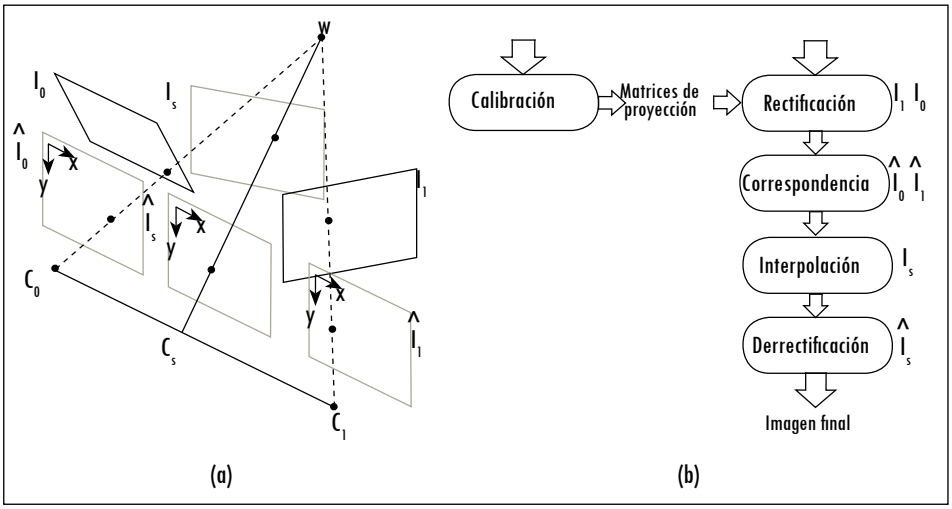



Fuente: presentación propia de los autores.

se caracterizan por producir resultados convincentes con un bajo costo computacional. Sin embargo, al aplicar estas técnicas directamente a las imágenes, no necesariamente se producen vistas físicamente válidas. La única forma de asegurar este resultado es contar con imágenes en que las líneas epipolares sean paralelas (Seitz y Dyer, 1995 y 1996; Córdoba *et al.*, 2002).

Para obtener a partir de vistas no rectificadas imágenes físicamente válidas, es necesario seguir un proceso cuyo diagrama en bloque se muestra en la Figura 5b. En primer lugar, a partir de la calibración de la cámara, las vistas se rectifican para hacer coincidir sus líneas epipolares; en este punto, el problema de sintetizar nuevas vistas se restringe a interpolar vistas paralelas. Una vez se tiene la vista interpolada en el plano común, se debe reproyectar la imagen del plano común al plano imagen deseado. Este último se puede definir de muchas maneras, la más natural es definir su orientación interpolando las orientaciones de los planos imagen de las vistas reales. De forma geométrica se ilustra este proceso en la Figura 5a.

Figura 5. Síntesis de nuevas vistas a partir de vistas no paralelas. Representación geométrica (a) y diagrama en bloques (b) del proceso de síntesis de imágenes físicamente válidas a partir de imágenes no rectificadas




Fuente: (a) Seitz y Dyer, 1995; (b) presentación propia de los autores.

## 1.4.1 Calibración

La calibración consiste en determinar los parámetros de la trasformación entre puntos 3D de la escena y puntos 2D de la imagen, es decir, encontrar la MPP. Calibrar la cámara implica determinar los doce elementos de la matriz *P*, para lo cual es necesario plantear un sistema de ecuaciones con un mínimo de seis puntos. Para *n* puntos se tiene el siguiente sistema de ecuaciones:

$$WX = C \tag{4}$$

Puesto que el sistema de ecuaciones resultante es homogéneo, existen infinitas soluciones que hacen necesario fijar una de las incógnitas (González, 2000); en este caso se fijó $p44=1$, donde $W_{2nx11}$ es una matriz obtenida a partir de los puntos de calibración, $X_{11x1}$ es la matriz de incógnitas y $C_{2nx1}$ es la matriz de cuyos elementos son los pares de coordenadas de los puntos proyectados. La solución que minimiza el error de mínimos cuadrados es:

$$\min_{x}(E^TE) = \min_{x}((WX-C)^T(WX-C)) \tag{5}$$

Puesto que el mínimo se logra cuando la derivada se hace cero, se tiene que:

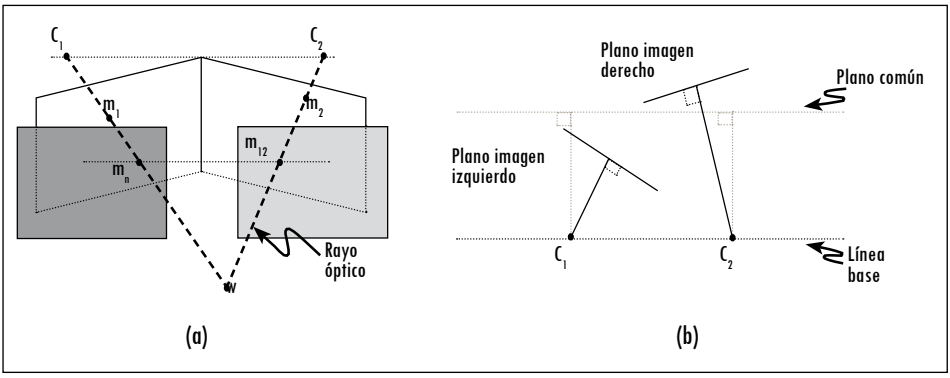
$$X = (W^T W)^{-1} W^T C \tag{6}$$

Al obtener el vector $X$, se obtiene la matriz de trasformación perspectiva $P$ y, por lo tanto, los parámetros de la cámara.

## 1.4.2 Rectificación

El proceso de rectificación consiste en encontrar nuevas proyecciones en las cuales las líneas epipolares sean colineales y paralelas a uno de los ejes de la imagen. Esto se ilustra en la Figura 6a, donde los píxeles $m_1$ y $m_2$ corresponden a la proyección del mismo punto $w$ en los planos imagen, y $m_{r1}$ y $m_{r2}$, a su proyección en los planos rectificados. Para rectificar, el algoritmo rota las cámaras sobre su centro óptico hasta que los planos imagen sean coplanares (Figura 6b). Esto se logra promediando los parámetros intrínsecos de las cámaras y redefiniendo la orientación del plano imagen de tal forma que sea paralelo a la línea base. Una vez calculadas las transformaciones para llevar cada una de las imágenes originales al plano rectificado, estas se aplican a las imágenes originales para producir las imágenes rectificadas (Fusiello *et al.*, 1999).

**Figura 6. Rectificación de un par de imágenes**




Fuente: presentación propia de los autores.

## 1.4.3 Correspondencia

Esta etapa consiste en encontrar los puntos que corresponden a la proyección de un mismo punto tridimensional en ambas imágenes. El algoritmo implementado extrae de la imagen bordes que delimiten cambios entre superficies o

regiones de las superficies, presentes en los objetos, para luego aplicar el proceso de correspondencia a ellas. A fin de obtener los bordes de las imágenes se utilizó el extractor de bordes de Canny, el cual genera bordes definidos que delimitan las superficies y captura mejor los detalles presentes, gracias a que tiene como parámetros de minimización el error de detección, el error de localización y el error de respuesta múltiple, simultáneamente, a diferencia de otros operadores (González, 2000).

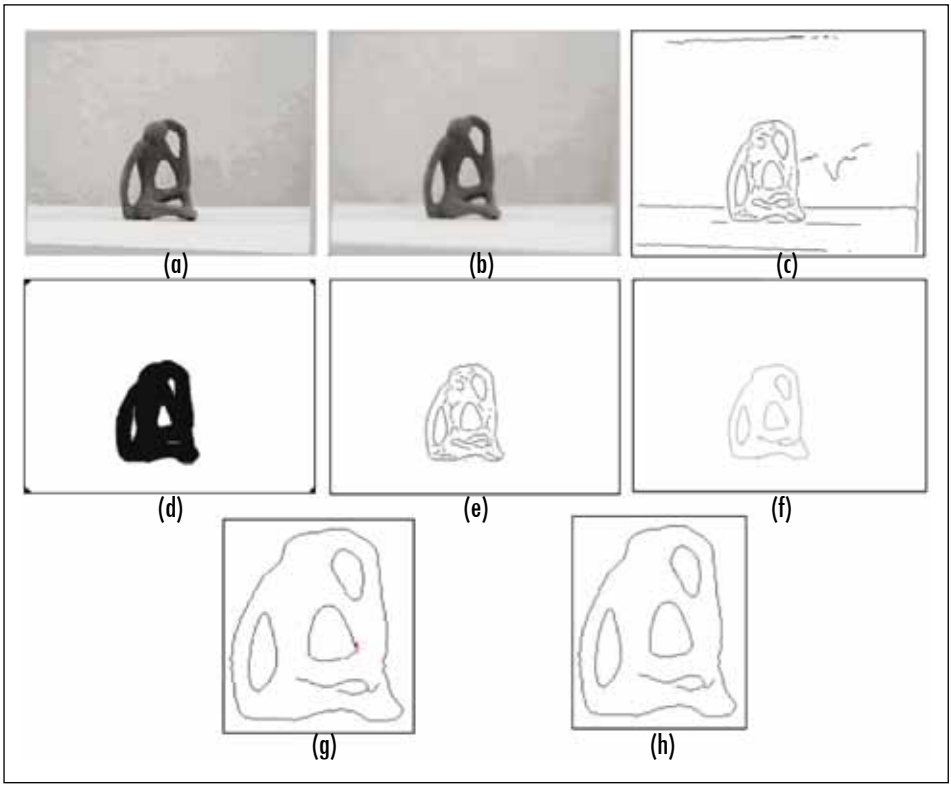
Para mejorar el rendimiento del extractor se implementó una etapa de preprocesamiento que suaviza las texturas de las superficies, lo cual evita detectar algunos bordes que no pertenezcan a límites entre superficies de alto contraste. Dicha etapa se puso en funcionamiento disminuyendo la resolución de la imagen y aplicando a estas imágenes los filtros *wiener* y *medfilt2* del *toolbox* de imágenes de Matlab®, para luego aplicar el extractor de bordes a la imagen preprocesada. De los bordes obtenidos se descartaron aquellos que no pertenecen al objeto, es decir, producidos por texturas del fondo de la imagen. Para ello es necesario diferenciar entre el fondo de la imagen y el objeto, lo que se logra bajo el supuesto de que la intensidad predominante determina el fondo y genera una máscara que elimina los bordes encontrados en el fondo de la imagen.

Posteriormente, la imagen de bordes se ajusta al tamaño de la imagen original, de manera que la posición de los bordes en la imagen original sea conocida. Del procedimiento anterior aún existen algunos pequeños bordes que no representan características importantes en la imagen; para eliminarlos se realiza otro procedimiento, en el que se compara el tamaño en píxeles de cada uno de los bordes en la imagen con un tamaño mínimo permitido.

Finalmente se le hace un posprocesamiento a la imagen de bordes, que consiste en completar contornos, a fin de obtener una imagen de bordes que representa los cambios entre superficies o regiones de alto contraste. Para este procedimiento se hallan los puntos terminales de los bordes y se conectan a cada uno un máximo de dos puntos terminales, los más cercanos dentro de una vecindad de tamaño definido por el usuario. Todo este proceso se ilustra en la Figura 7.

A partir de los bordes extraídos de las imágenes se genera un espacio de búsqueda de posibles correspondencias. Debido al proceso de rectificación realizado previamente en las imágenes, los puntos de correspondencia se encuentran sobre la misma línea horizontal en ambas imágenes. Este espacio de búsqueda se ilustra en la Figura 8, donde la línea vertical muestra las posiciones de los bordes en la línea de búsqueda derecha, y la horizontal, las posiciones de los bordes en la línea de búsqueda izquierda.

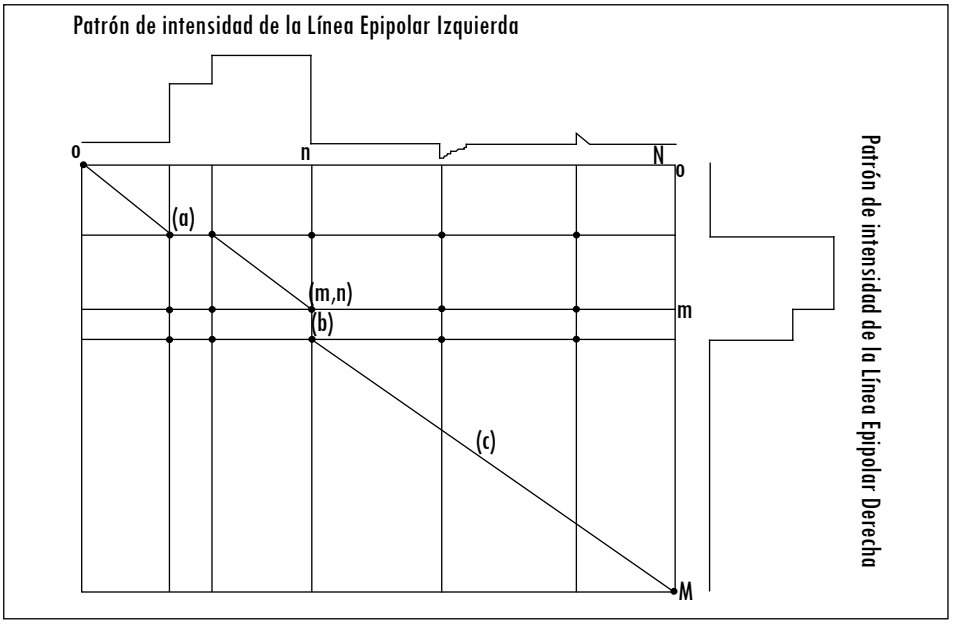
Figura 7. Ejemplo del proceso de extracción de bordes: (a) imagen de intensidad, (b) imagen filtrada, (c) imagen de bordes por Canny, (d) máscara de fondo, (e) imagen de bordes después de aplicar la máscara, (f) terminales después de eliminar bordes pequeños, (g) camino para conectar terminales y (h) imagen de bordes definitiva




Fuente: presentación propia de los autores.

En este plano de búsqueda se definen como nodos las intersecciones de los bordes, y son numerados de izquierda a derecha en cada línea de búsqueda de 0 a $M$ en la línea derecha y de 0 a $N$ en la línea izquierda. Por conveniencia, el inicio y el final de cada línea de búsqueda se consideran un borde. En esta configuración la búsqueda de correspondencias se reduce a encontrar un camino óptimo desde el nodo (00) al (NM). Para encontrar este camino se usó un algoritmo basado en el método de búsqueda por línea (Otha y Kanade, 1985). Una vez se ha encontrado el camino de correspondencia, se dice que todo nodo $\{m,n\}$ que pertenece al camino óptimo conforma un par de bordes $m$ y $n$ correspondientes en la línea epipolar analizada.

Figura 8. Plano dimensional para búsqueda por línea




Fuente: Seitz y Dyer, 1995.

Los cálculos de los costos en este algoritmo de búsqueda se basan en el costo de un camino primitivo en el plano bidimensional de búsqueda. El costo de un camino primitivo se define como la similitud entre intervalos delimitados por bordes en ambas imágenes. Si $a_1 \ldots a_k$ y $b_1 \ldots b_l$ son los valores de la intensidad de los píxeles contenidos en los dos intervalos, entonces la media y la varianza de todos los píxeles en los dos intervalos se calcula como:

$$m = \frac{1}{2}\left(\frac{1}{k}\sum_{i=1}^{k} a_i + \frac{1}{l}\sum_{j=1}^{l} b_j\right) \quad \sigma^2 = \frac{1}{2}\left(\frac{1}{k}\sum_{i=1}^{k}(a_i - m)^2 + \frac{1}{l}\sum_{j=1}^{l}(b_j - m)^2\right) \tag{7}$$

En esta definición los dos intervalos contribuyen por igual al valor de la media y la varianza aunque sus longitudes sean diferentes. El costo de un camino primitivo que determina correspondencia entre estos dos intervalos se define como:

$$C_p = \sigma^2\sqrt{k^2 + l^2} \tag{8}$$

De manera intuitiva, el significado de esta definición puede expresarse de la siguiente manera. Se asumen que los píxeles en los dos intervalos vienen de una superficie homogénea en el espacio tridimensional y que, por lo tanto, deben

tener intensidades similares, es decir, su varianza debe ser pequeña. Para que el algoritmo de búsqueda por línea encuentre correspondencias válidas, las escenas deben ser monótonas (Otha y Kanade, 1985; Seitz y Dyer, 1995), es decir, todos los puntos correspondientes aparecen en el mismo orden a lo largo de las líneas epipolares conjugadas de las imágenes.

### 1.4.4 Interpolación

Debido al proceso de rectificación hecho en las imágenes, el problema de dos vistas de una escena se reduce a las condiciones mostradas en la Figura 9a. Por conveniencia se supone que la cámara es movida desde el origen del mundo hacia la posición ($c_x$, $c_y$, 0) y la distancia focal cambia de $f_0$ a $f_1$. Por lo que las MPP para las imágenes izquierda ($P_0$) y derecha ($P_1$) son de la forma:

$$P_0 = \begin{bmatrix} f_0 & 0 & 0 & 0 \\ 0 & f_0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \qquad P_1 = \begin{bmatrix} f_1 & 0 & 0 & -f_1 \cdot c_x \\ 0 & f_1 & 0 & -f_1 \cdot c_y \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{9}$$

Sean $m_0 \in I_0$ y $m_1 \in I_1$ proyecciones del mismo punto tridimensional $w = \{X\ Y\ Z\ 1\}^T$. Al desarrollar una interpolación lineal de los píxeles $m_0$, $m_1$, tenemos que:

$$(1-s)\cdot m_0 + s\cdot m_1 = (1-s)\cdot \frac{1}{Z} P_0 \cdot w + s\cdot \frac{1}{Z} P_1 \cdot w = \frac{1}{Z} P_s \cdot w \tag{10}$$

Donde:

$$P_s = (1-s)\, P_0 + sP_1 \tag{11}$$

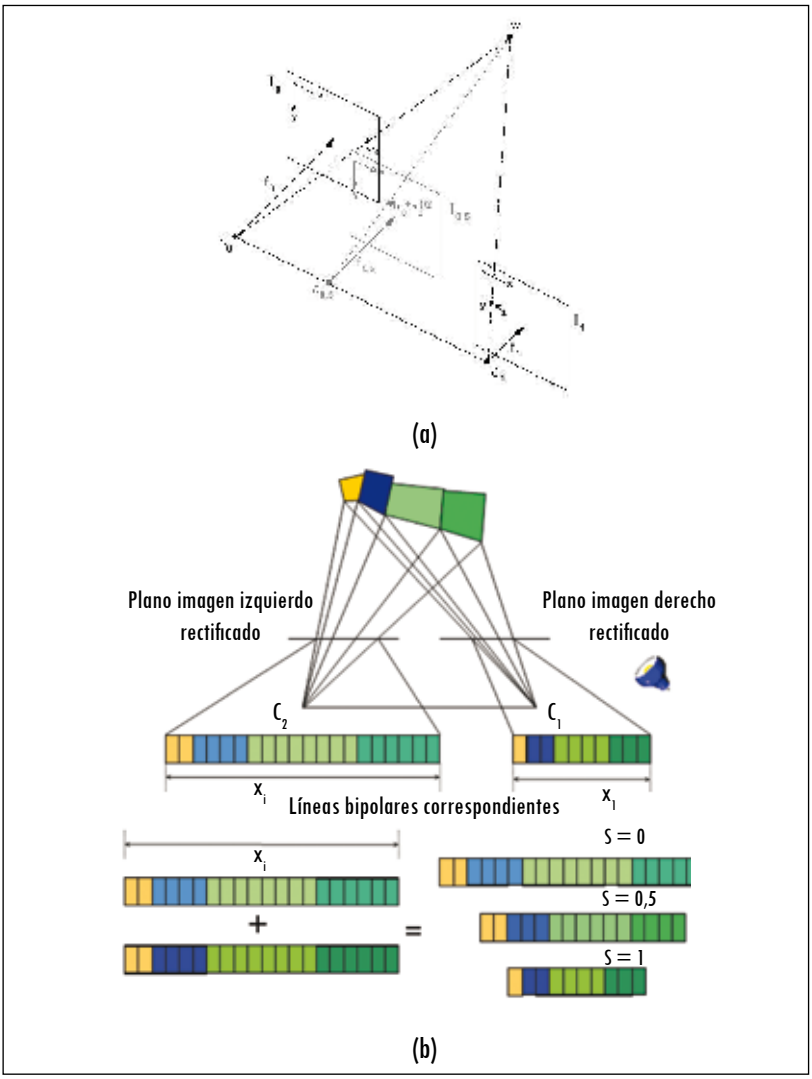
Por lo que la interpolación de las imágenes produce una nueva vista físicamente válida con matriz de proyección $P_s$, producto de la interpolación lineal de $P_0$ y $P_1$. Esta nueva vista tiene una distancia focal $f_s$ y centro óptico $c_s$ determinados por:

$$f_s = (1-s)\cdot f_0 + s\cdot f_1 \quad ; \quad c_s = \left( s\frac{f_1}{f_s}\cdot c_x, s\frac{f_1}{f_s}\cdot c_y, 0 \right) \tag{12}$$

Como se observa de (12), la imagen interpolada está definida por un parámetro de interpolación $s$, el cual determina la ubicación del centro óptico y la distancia focal de la cámara virtual generada por la interpolación. Para interpolar un segmento correspondiente, en principio, se interpolan linealmente las posiciones de los bordes. Luego se usa una estrategia para encontrar el color del segmento interpolado, igualando las longitudes de los dos segmentos originales

a la longitud del mayor. Ya con los segmentos de igual longitud se realiza una interpolación lineal del color, píxel a píxel. Una vez obtenido el color del segmento, es necesario ajustar su longitud a la longitud calculada a partir de la interpolación de los bordes que definen sus extremos. Con el procedimiento para cada segmento correspondiente se genera la línea epipolar interpolada (Figura 9b).

Figura 9. (a) Vistas paralelas de una escena, (b) interpolación de un par de segmentos correspondientes

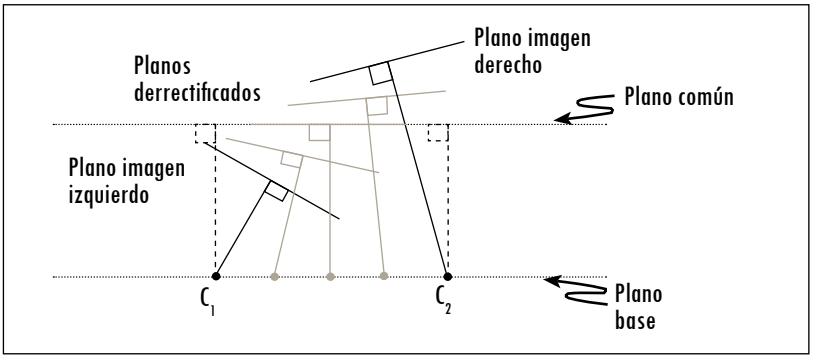



Fuente: (a) Seitz y Dyer, 1996; (b) presentación de los autores.

### 1.4.5 Derrectificación

Una vez se han interpolado las imágenes rectificadas, se desea llevar las imágenes obtenidas a un plano que simule un movimiento natural de la cámara virtual, a la vez simuladas por el algoritmo a lo largo de una línea recta. Esto se logra aplicando a cada una de las imágenes interpoladas una transformación lineal para llevar la imagen del plano rectificado a la posición del plano imagen deseado. El cálculo de dicha transformación se presenta en (Córdoba *et al.*, 2002). La trayectoria descrita por la cámara virtual se muestra en la Figura 10, y esta se consigue a través de la interpolación de los parámetros intrínsecos de las MPP derecha e izquierda y redefiniendo la orientación del plano imagen para cada posición de la cámara virtual.

Figura 10. Trayectoria de derrectificación




Fuente: presentación propia de los autores.

La orientación del plano imagen se obtiene de la interpolación de las orientaciones de los planos imágenes originales respecto al plano rectificado. Para este proceso se obtienen los ángulos que determinan las rotaciones aplicadas en el proceso de rectificación a cada uno de los planos imagen, para ser llevados al plano rectificado. Una vez conocidos estos ángulos, se determina la rotación del plano imagen interpolado, para llevarlo al plano imagen deseado. El ángulo de rotación ($\theta_s$) sobre el eje *Y* de la imagen sintetizada se obtiene interpolando los ángulos en que se rotaron los planos imágenes originales sobre el eje *Y*, para ser rectificados ($\theta_1$,$\theta_2$), mediante la siguiente relación:

$$tan\,(\theta_s) = s \cdot tan\,(\theta_1) + (1 - {}_s)\,tan\,(\theta_2) \qquad (13)$$

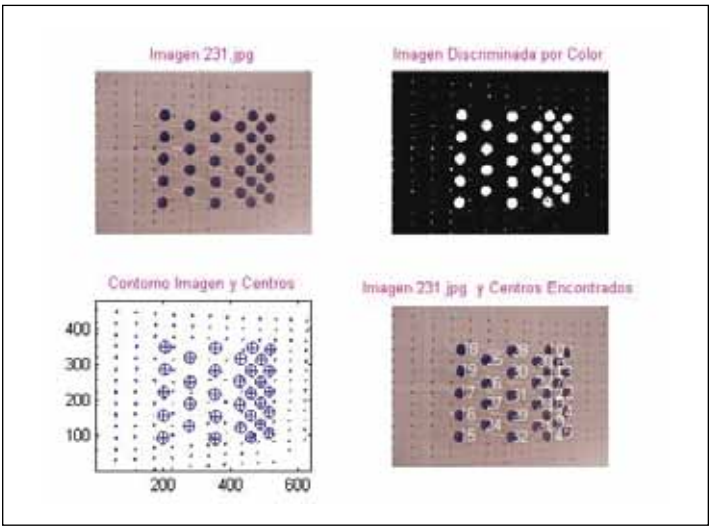
Entre tanto, para las rotaciones sobre los ejes $X$ y $Z$ se realiza una interpolación lineal de los ángulos en que se rotaron los planos imágenes originales sobre dichos ejes, para ser rectificados.

## 2. Resultados

### *2.1 Calibración*

Se trabajó con una sola cámara digital Sony MVC-FD83, calibrada en dos diferentes posiciones. Las imágenes tomadas por la cámara en dichas posiciones se toman como vistas base. La calibración se llevó a cabo usando una plantilla de calibración que posee un alto número de puntos fácilmente localizables en la imagen. En la Figura 11 se puede observar que el procedimiento comienza con la extracción de la proyección de los puntos tridimensionales, discriminando por color de las circunferencias de la plantilla en la imagen, con lo que se genera una imagen binaria a la cual se le extraen los contornos formados por la proyección de las esferas.

Figura 11. Extracción de la proyección de los puntos tridimensionales

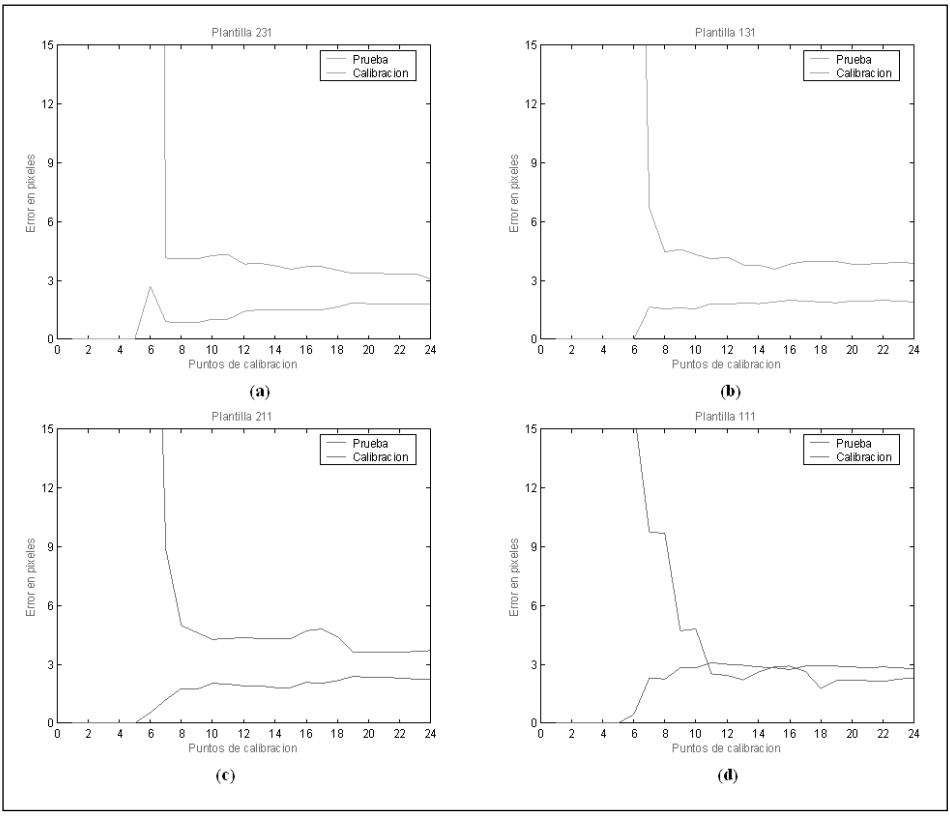



Fuente: presentación propia de los autores.

Ya con los contornos establecidos, se calcula el centro de cada contorno que define la proyección del punto tridimensional conocido en la imagen. Para validar la calibración se encontró el error cuadrático medio de la proyección en

píxeles, de los puntos 3D conocidos; estos se agruparon en puntos de prueba y puntos de calibración de forma aleatoria, a fin de observar el comportamiento del error de proyección de los puntos de calibración y de los puntos de prueba a medida que se toman más puntos para calibrar las cámaras, con una población de prueba del 20% y de calibración del 80%, sobre el total de puntos en la plantilla de calibración. Los resultados se observan en la Figura 12, en la cual se puede observar que la calibración modela adecuadamente el espacio donde se encuentran los puntos 3D conocidos, dado que el comportamiento de los errores es asintótico, tendiendo estos a un error de aproximadamente tres píxeles.

**Figura 12. Error cuadrático medio por píxel, para la calibración en las posiciones. (a) 30º izquierda, (b) 30º derecha, (c) 15º izquierda, (d) 15º derecha**




Fuente: presentación propia de los autores.

### *2.2 Rectificación*

Al realizar la transformación, la imagen rectificada resultó ser más grande que la original, por lo que se proyectó un píxel de la imagen original en varios píxeles en el plano rectificado. Por otra parte, las coordenadas de los píxeles de las imágenes rectificadas no estaban sujetas a proyectarse en una región específica de la imagen; por lo tanto, se trasladó cada imagen rectificada para visualizar los resultados.

Después de este proceso, y ante la imposibilidad de cuantificar el error del proceso de forma automática, se decidió hacer una inspección manual de las imágenes para comprobar si efectivamente las líneas epipolares en las imágenes rectificadas eran paralelas y coincidentes. En la Figura 13 se ilustran algunos resultados de esta verificación, donde se observa cómo efectivamente las imágenes rectificadas son adecuadas para el resto del proceso de síntesis.

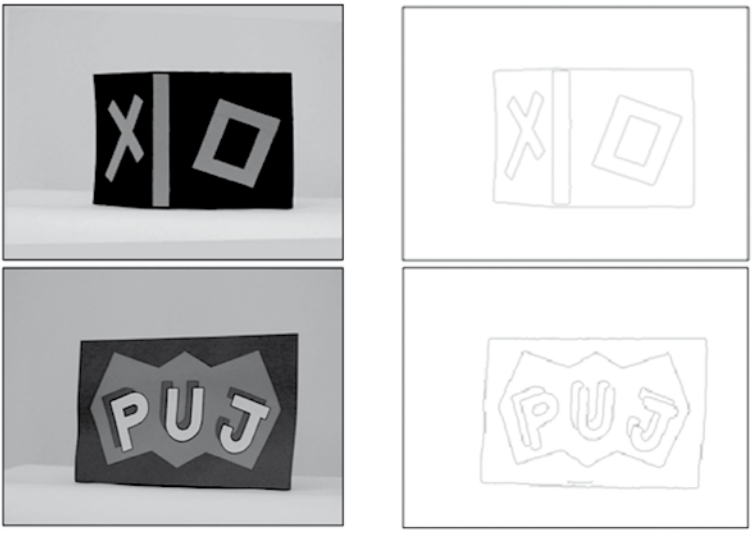
Figura 13. Rectificación obtenida con diferentes escenas


Fuente: presentación propia de los autores.

### *2.3 Correspondencia e interpolación*

Para hacer el algoritmo más eficiente es necesario extraer el menor número de bordes posible con mínima pérdida de información, por lo que se desarrolló el algoritmo de forma paramétrica, de manera que se pudieran manipular ciertos parámetros propios del proceso de extracción de bordes, para ajustar el proceso a las características propias de cada uno de los objetos, como el tamaño de las regiones o superficies y las texturas propias de las superficies. Esta manipulación permite obtener resultados adecuados del proceso de extracción de bordes para diferentes objetos. Este proceso se ilustra en la Figura 14.

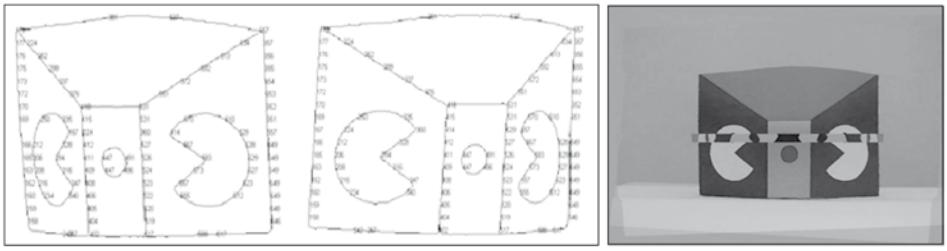
Figura 14. Extracción de bordes




Fuente: presentación propia de los autores.

Para observar un buen resultado en el proceso de interpolación es necesario que se encuentren las correspondencias de los bordes relevantes, puesto que correspondencias erróneas producen distorsiones notorias en las imágenes interpoladas. En la Figura 15 se ilustra cómo al perderse algunas características relevantes del objeto durante el proceso de extracción de bordes el proceso de correspondencia es equivocado (parte a) y, por consiguiente, la imagen finalmente sintetizada no corresponde a una vista físicamente válida (parte b).

Figura 15. (a) Correspondencia de bordes cada veinte líneas epipolares de las imágenes rectificadas. (b) Imagen interpolada rectificada en presencia de falsas correspondencias (s=0,5)

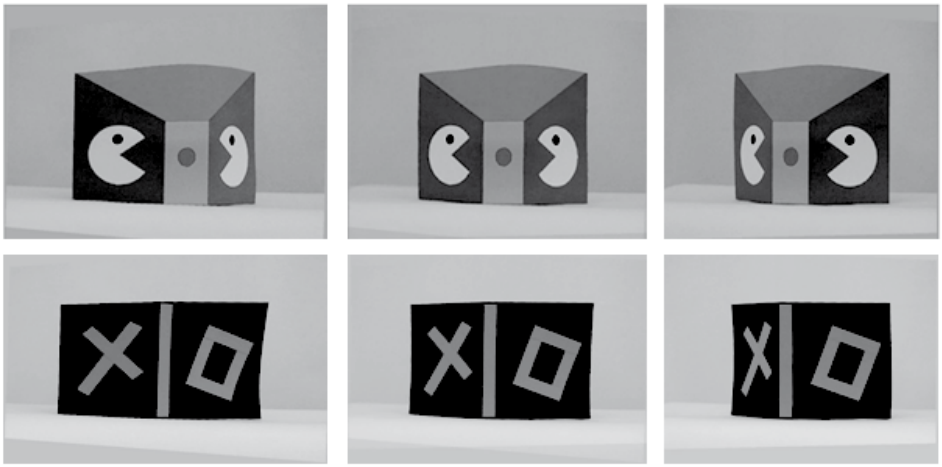


(a) (b)

Fuente: presentación propia de los autores.

Cuando se interpola con todas las correspondencias de los bordes relevantes, se obtienen resultados satisfactorios que simulan un movimiento de la cámara virtual rectificada sobre la línea base, sin cambiar su orientación. Las imágenes sintetizadas conservan las texturas en las superficies del objeto sin introducir distorsión aparente de forma. Un ejemplo de esta clase de resultados se ilustra para dos escenas simples en la Figura 16.

Figura 16. (a) Imagen izquierda rectificada (s=0), (b) imagen interpolada rectificada (s=0,5) y (c) imagen derecha rectificada (s=1)

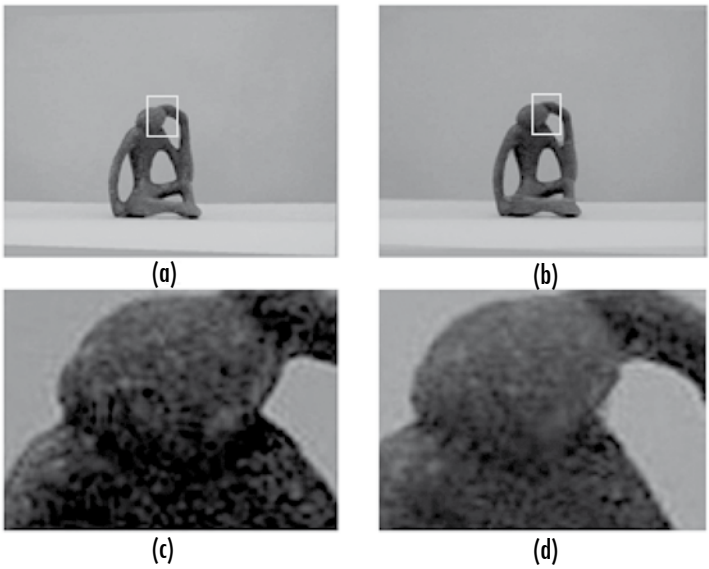


(a) (b) (c)

Fuente: presentación propia de los autores.

Aunque se tenga la correspondencia total entre bordes relevantes de la escena, no es posible sintetizar las texturas de las vistas originales en las imágenes sintetizadas de cualquier objeto, dado que la correspondencia píxel a píxel encontrada por el algoritmo de interpolación solamente es válida cuando la superficie o región delimitada por los bordes correspondientes es plana, como se muestra en Figura 17.

Figura 17. (a) Imagen rectificada izquierda original, (b) imagen interpolada (s=0,5), (c) detalle de pérdida de textura de la imagen sintetizada y (d) detalle de la textura de la imagen original

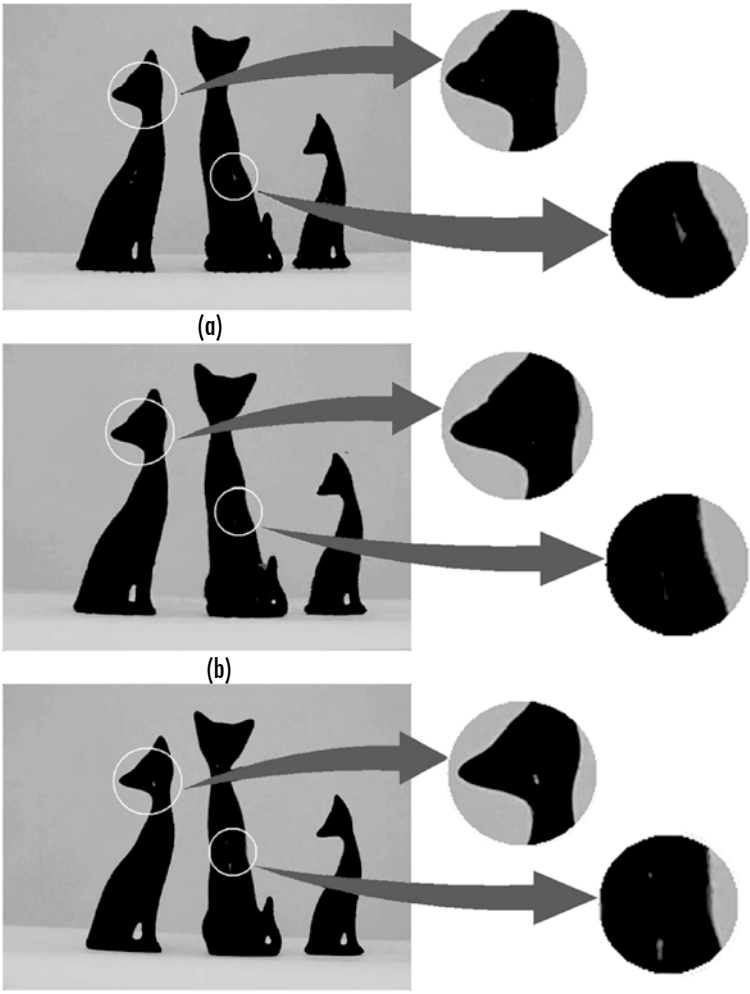



Fuente: presentación propia de los autores.

Sin embargo, cuando los brillos presentes en la imagen son parecidos y se puede establecer algún tipo de correspondencia entre ellos, aunque no estén modelados dentro del algoritmo, estos ayudan a conservar las texturas de los objetos con superficies curvas en las imágenes sintetizadas, puesto que dividen las zonas del objeto en regiones planas de menor longitud. Aunque estos brillos mejoran la apariencia de las imágenes sintetizadas cuando hay objetos con superficies curvas, estas imágenes no pueden ser consideradas físicamente válidas, pues un par de brillos parecidos no necesariamente son producidos por la misma fuente de luz.

El efecto que se presenta cuando no se puede establecer correspondencia entre los brillos presentes en la imagen es la mezcla de las texturas de la superficie o región, incluido el brillo, como se observa en la Figura 18.

Figura 18. (a) Brillos de interés presentes en la imagen izquierda, (b) brillos de interés interpolados (s=0,5) y (c) brillos de interés presentes en la imagen derecha



(a)

(b)

(c)

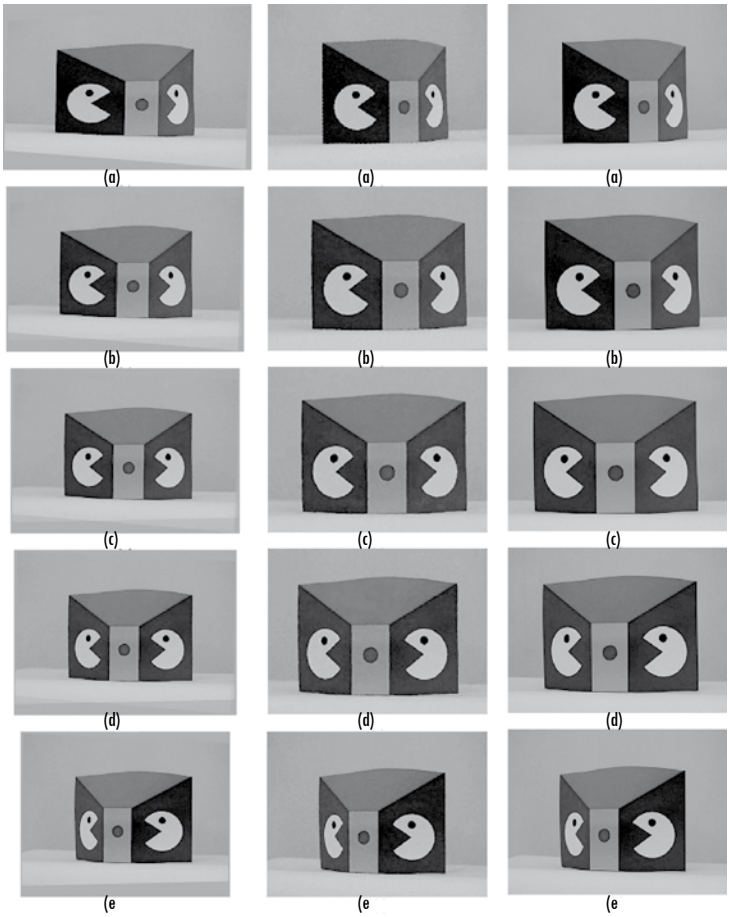
Fuente: presentación propia de los autores.

## *2.3 Derrectificación*

El método utilizado para definir el plano al que se desea derrectificar la imagen permite sintetizar imágenes que simulan un movimiento natural de la cámara virtual en línea recta, muy parecido al movimiento de la cámara real, como se observa en la Figura 19. En cuanto a la evaluación de manera cuantitativa de la validez física de las imágenes sintetizadas, fue imposible comparar las imágenes

sintetizadas con imágenes reales tomadas por la cámara, porque el montaje físico impidió reproducir con la exactitud requerida la trayectoria de la cámara virtual en el mundo real.

Figura 19. Interpolación de imágenes rectificadas (izquierda). Síntesis de imágenes (centro). (a) s=0, (b) s=0,2, (c) s=0,5, (d) s=0,8 (e) s=1. Imágenes reales (derecha). Posición (a) 30º izquierda, (b) 15º izquierda, (c) centro, (d) 15º derecha y (e) 30º derecha




Fuente: presentación propia de los autores.

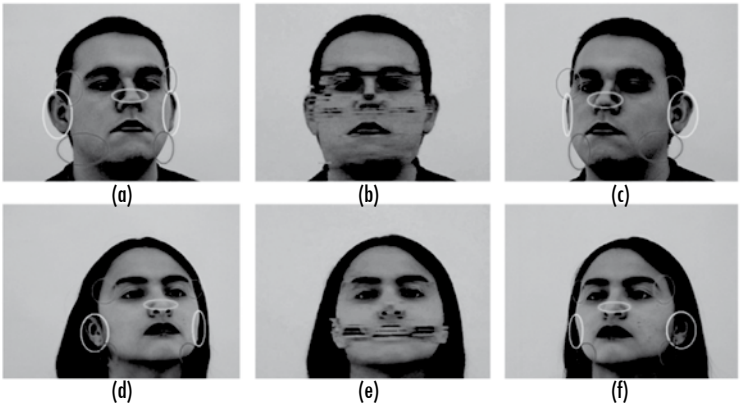
## *2.4 Evaluación general del algoritmo en rostros*

Un propósito planteado desde el principio de la investigación era evaluar las posibilidades de usar este algoritmo para corregir los problemas de postura pre-

sentes en la síntesis de voz visual (Muñoz *et al.*, 2003), por lo que el algoritmo se evaluó con rostros. Los resultados obtenidos fueron, en general, desfavorables, pues las imágenes sintetizadas se caracterizaban por distorsiones significativas, debido al gran número de correspondencias inválidas, producto de un espacio de búsqueda que no cumple con las especificaciones del sistema, porque el camino óptimo encontrado no tiene un significado físicamente válido y porque las oclusiones presentes en la imágenes son considerables.

Un ejemplo de este tipo de resultados se presenta en la Figura 20b, donde se muestran las distorsiones presentes en la imagen sintetizada en regiones como: las cejas, la patilla, la nariz, las orejas y la papada. Las regiones ocluidas que producen estas distorsiones se ilustran discriminadas por color en las figuras 20a y 20b. De forma adicional se observa cómo, cuando los rostros presentan oclusiones más drásticas, se observan distorsiones mucho más significativas.

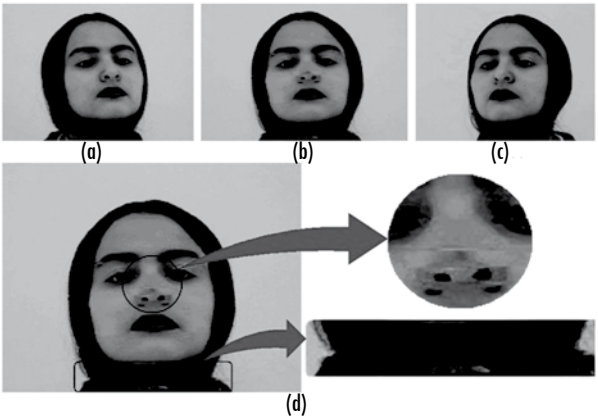
Figura 20. *Ocluart.* (a) Imagen izquierda, (b) imagen sintetizada con s=0,5, (c) imagen derecha *oclupao*, (d) imagen izquierda, (e) imagen sintetizada con s=0,5 y (f) imagen derecha




Fuente: presentación propia de los autores.

Para tratar de evaluar el algoritmo con imágenes de rostros que se acercaran más a las especificaciones del algoritmo, se trataron de disminuir las oclusiones presentes en las imágenes reales. Un ejemplo de este tipo de evaluación se ilustra en la Figura 21, donde para el rostro "Paola" se utilizó el cabello para ocultar las orejas, para disminuir el número de oclusiones presentes en las imágenes reales, con lo que se obtuvieron resultados adecuados en la gran mayoría de superficies; sin embargo, en algunas aparecieron deformidades como las ilustradas en la Figura 21d.

Figura 21. Rostro "Paola": (a) imagen izquierda, (b) imagen sintetizada con s=0,5, (c) imagen derecha, (d) detalle de distorsiones causadas por oclusión en la imagen sintetizada con s=0,5




Fuente: presentación propia de los autores.

El problema de distorsión de la nariz observado en la Figura 21d, que consiste en un ensanchamiento falso en la parte superior, no obedece a una oclusión, porque las superficies son visibles en ambas imágenes, sino que es producto de las falsas correspondencias encontradas por el algoritmo, producidas por la uniformidad de textura de la piel. Cuando las superficies tienen una textura uniforme, como las superficies de los costados de la nariz y las superficies de los pómulos, el algoritmo de correspondencias pasa por alto los bordes que delimitan el cambio de una superficie a otra; en este caso el cambio de superficie entre la nariz y el pómulo.

## 3. Conclusiones

La MPP encontrada por el proceso de calibración modela la relación existente entre el espacio tridimensional y el bidimensional, involucrada en la obtención de imágenes, por lo que se puede manipular con resultados positivos la información codificada dentro de la matriz para las demás etapas del sistema. El algoritmo implementado obtiene vistas sin distorsión aparente, de una escena a partir de dos vistas de esta, bajo la condición de que las vistas base se rectifiquen antes de realizar la interpolación y exista iluminación controlada, escenas monótonas y con texturas planas; sin embargo, no es posible evaluar de manera cuantitativa la validez física de las imágenes sintetizadas, debido a que la precisión de las medidas del montaje físico es insuficiente para reproducir con exactitud

la trayectoria de la cámara virtual en el mundo físico, lo que impide obtener imágenes válidas que puedan ser comparadas con las imágenes sintetizadas por el algoritmo a través de un error con una métrica asociada.

Adicionalmente, a medida que aumenta el ángulo entre las vistas base, el algoritmo se ve afectado en mayor medida por oclusiones, brillos y superficies no planas, presentes en los objetos. Es necesario hacer mejoras significativas al algoritmo para ser aplicado en síntesis de imágenes complejas, como el problema de corrección de postura en rostros en voz visual.

## 4. Recomendaciones

Como la posición de la cámara no debe variar del proceso de calibración al proceso de obtención de vistas, se podría calibrar la cámara con la información presente en la escena, asegurando que las MPP modelen con exactitud el espacio capturado por las vistas adquiridas y, así, den mayor flexibilidad al sistema. Para que el algoritmo permita sintetizar imágenes en un ángulo de visión más amplio, que justifique su utilización en el análisis de escenas complejas, es necesario modelar dentro del algoritmo oclusiones, brillos y superficies curvas. Para modelar cualquier tipo de superficie, y así conservar las texturas, se propone generar una correspondencia píxel a píxel. A fin de interpolar las imágenes cuando presentan oclusiones, es preciso contar con información adicional que permita modelar las superficies del objeto que estén ocluidas. Una posible manera para obtener esta información, sería tomar una tercera vista base.

## Agradecimientos

Al ingeniero Carlos Alberto Parra Rodríguez, por las discusiones sostenidas durante el curso de esta investigación, las cuales fueron de gran importancia en su desarrollo. Al Departamento de Ingeniería Electrónica de la Pontificia Universidad Javeriana, que otorgó los recursos necesarios para este proyecto de investigación. A los ingenieros Diana Paola García y Andrés Córdoba, sin los cuales no hubiera sido posible realizar el proyecto de investigación.

## Referencias

BEIER, T. y NEELY, S. Feature-based image metamorphosis. *ACM SIGGRAPH´92 Computer Graphics*, 1992, vol. 26, núm. 2, pp. 35-42.

CHEN, S. C. y WILLIAMS, L. *International Conference on Computer Graphics and Interactive Techniques: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. New York: View Interpolation for Image Synthesis, 1993, pp. 279-288.

CÓRDOBA, A.; FAJARDO A. y GARCÍA D. *Síntesis de imágenes a partir de vistas de una escena*. Bogotá: Pontificia Universidad Javeriana, 2002.

FUSIELLO, A.; TRUCCO, E. y VERRI, A. A compact algorithm for rectification of stereo pairs. *Machine Vision and Applications*, 1999, vol. 12, núm. 1, pp. 16-22.

GONZÁLEZ, J. *Visión por computador*. Madrid: Paraninfo, 2000.

INAMOTO, N. y SAITO, H. Virtual viewpoint replay for a soccer match by view interpolation from multiple cameras. *IEEE Transactions on Multimedia*, 2007, vol. 9, núm. 6, pp. 1155-1166.

KUBOTA, A.; KODAMA, K. y HATORI, Y. Deconvolution method for view interpolation using multiple images of circular camera array. *IEEE International Conference on Image Processing*, 2006, pp. 1049-1052.

MUÑOZ, M.; SOTO, C. y VIZCAYA P. Avances en síntesis de voz visual y sus aplicaciones. *VIII Simposio de Tratamiento de Señales, Imágenes y Visión Artificial*, Medellín, Colombia, 2003.

NALWA, V. *A guided tour of computer vision*. New York: Addison-Wesley, 1993.

OTHA, Y. y KANADE, T. Stereo by intra and Inter-scanline search using dynamic programming. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1985, vol. 7, núm 2, pp. 139-154.

SEITZ, S. y DYER, C. Physically-valid view synthesis by image interpolation. En: *International Conference on Computer Graphics and Interactive Techniques: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. New York, 1995, pp. 18-25.

—. View morphing: Synthetizing 3D metamorphoses using image transforms. *IEEE Transactions and Computer Graphics (SIGGRAPH'96)*, vol. 11, núm. 1, 1996, pp. 21-30.

ZH ENG, X. y WU, E. Efficient 3D image warping for composing novel views. *Proceedings of Computer Graphics International*, 2001, pp. 123-130.