

Realizando Realidad Aumentada

Dr. Luis Gerardo de la Fraga

E-mail: fraga@cs.cinvestav.mx
Departamento de Computación
Cinvestav

21 de mayo, 2015

1. Áreas de estudio relacionadas con *realidad aumentada*
2. Definición de *realidad aumentada*
3. Cómo poner una cámara en OpenGL
4. Calibración de la cámara
5. Conclusiones

Las áreas principales ^a

^aFoley, van Dam, Feiner, Hughes, Computer Graphics: principles and practice, 2000, Addison Wesley.

1. **Graficación:** Trata la síntesis de escenas con objetos reales o imaginarios a partir de sus modelos computacionales.
2. **Procesamiento de Imagen:**
 - 2.1 **Realzado de imagen**
 - 2.2 **Detección y reconocimiento de patrones**
 - 2.3 **Análisis de escenas**
 - 2.4 **Visión por computadora:** Reconstrucción de un modelo 3D de una escena a partir de varias imágenes 2D.

1. Procesamiento de imágenes
 - 1.1 Compresión de datos
2. Graficación
 - 2.1 Realidad virtual
 - 2.2 Geometría computacional
 - 2.3 Interfaces hápticas
 - 2.4 Interfaces hombre-máquina
 - 2.5 Modelos deformables
3. Visión
 - 3.1 Geometría proyectiva
 - 3.2 Realidad aumentada (usa las tres áreas)
 - 3.3 Análisis numérico
 - 3.4 Procesamiento paralelo (GPUs)

Nuevas áreas emergentes

1. Fotografía computacional^a
 2. Visualización científica^b
 3. Visualización de información^b
 4. Análisis visual^b
- ▶ (a) Special issue on Computational Photography, IEEE Computer, August 2006.
 - ▶ (b) IEEE Computer, May 2013, vol 46, number 5

Realidad aumentada

- ▶ La realidad aumentada nos permite superponer imágenes generadas por computadora sobre ambientes pertenecientes al mundo real.
- ▶ Este proceso proporciona a los usuarios la interacción con contenidos digitales a través de la manipulación de objetos reales, o bien, complementa el mundo real mediante objetos virtuales, creando la sensación de que el ambiente ha sido aumentado.

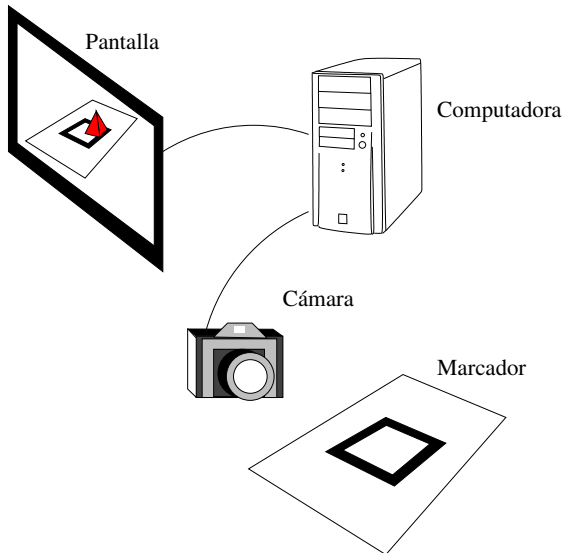
IEEE Computer, July 2012. What's real about Augmented reality?

Tipos de realidad aumentada

1. Adición de contenidos digitales
2. Adición de objetos virtuales

Aunque (1) es teóricamente posible, existen mejores formas de localizar objetos para desplegar contenidos digitales, por ejemplo, en redes WIFI [IEEE Signal Processing Magazine 22 (4) (2005)]

El escenario

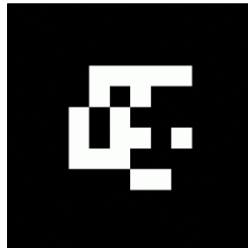


Marcadores

- ▶ Necesitamos marcadores para calibrar la cámara e identificar el objetivo virtual que se usará.
- ▶ Hemos trabajado con dos tipos de marcadores: de plantilla y de referencia



(a) Marcador de plantilla



(b) Marcador de referencia

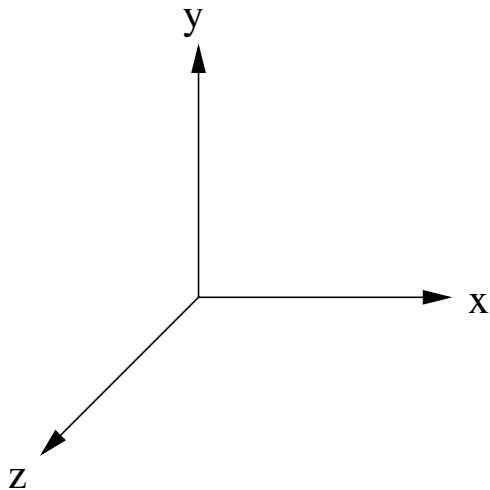
Identificación de un marcador

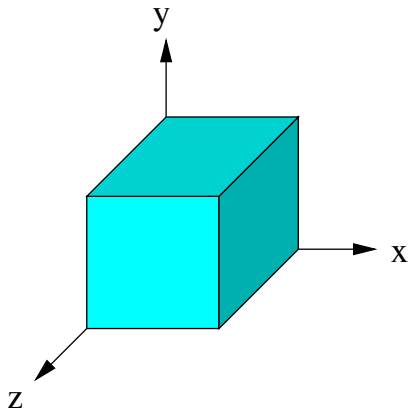
Se utilizan técnicas de procesamiento de imagen:

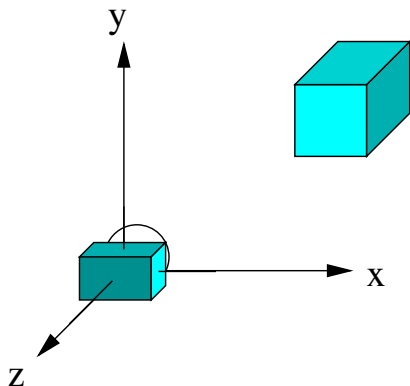
1. Segmentar
2. Extraer y etiquetar todas las componentes conectadas
3. Encontrar primero un vértices y después los otros tres
4. Ajustar las aristas a líneas usando análisis de componentes principales
5. Extraer la imagen central del marcador

Objetos virtuales

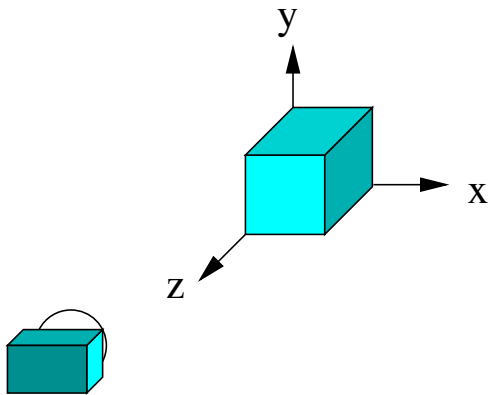
- ▶ Agrega objetos virtuales para enriquecer el contenido que ofrece una cámara
- ▶ Hay que crear los objetos virtuales
- ▶ Tienen que dibujarse en la misma geometría del mundo real







$$\begin{aligned}\mathbf{P}_2 &= T(t_1, t_2, t_3)R\mathbf{P}_1 \\ &= \begin{bmatrix} I & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} R & \mathbf{0} \\ \mathbf{0}^T & 1 \end{bmatrix} \mathbf{P}_1 = \begin{bmatrix} R & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \mathbf{P}_1\end{aligned}$$



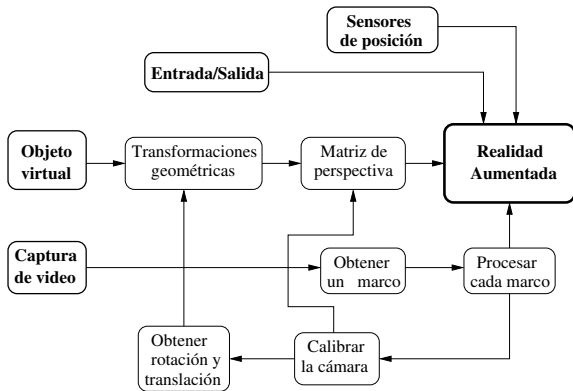
$$\left[\begin{array}{c|c} R & \mathbf{t} \\ \hline \mathbf{0}^T & 1 \end{array} \right] = \left[\begin{array}{c|c} R & \mathbf{0} \\ \hline \mathbf{0}^T & 1 \end{array} \right] \left[\begin{array}{c|c} I & -R^T \mathbf{t} \\ \hline \mathbf{0}^T & 1 \end{array} \right]$$

donde la cámara se sitúa en el punto

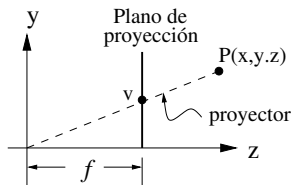
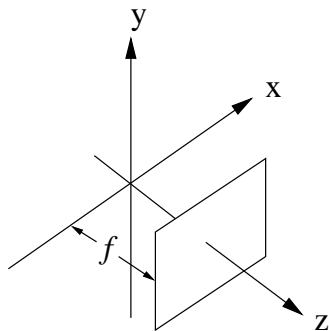
$$\mathbf{c} = [c_x, c_y, c_z]^T = -R^T \mathbf{t} \text{ y } \mathbf{t} = -R\mathbf{c}.$$

- ▶ Geométricamente es equivalente mover el objeto o mover la cámara.
- ▶ Pero en OpenGL todas las transformaciones deben afectar a los objetos
- ▶ Para dibujar los objetos en OpenGL, una cámara real no existe y no debe moverse.
- ▶ Se usa la función `gluLookAt()` que calcula la transformación aplicada en los objetos según una posición (virtual) de la cámara.

Realidad aumentada



Proyección en perspectiva



Por triángulos semejantes:

$$\frac{v}{f} = \frac{y}{z}$$
$$v = \frac{fy}{z}$$

Las ecuaciones:

$$u = \frac{xy}{z}, \quad y \quad v = \frac{fy}{z},$$

pueden escribirse en forma matricial como:

$$\lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

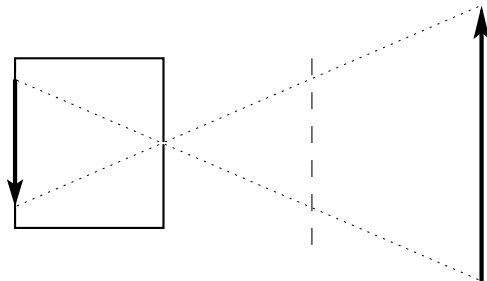
Usando solo coordenadas homogeneas:

$$\lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

Cambiando el centro óptico de la cámara a (u_0, v_0) :

$$\lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

Modelo de la cámara oscura (1/2)



Usando la matriz de rotación R y el vector de translación \mathbf{t} de la cámara con respecto al sistema de coordenadas del mundo:

$$\lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & \alpha & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

$$\lambda \mathbf{p} = K[R|\mathbf{t}]\mathbf{P}$$

$$\lambda \mathbf{p} = K[R|\mathbf{t}]\mathbf{P}$$

$$\begin{aligned}\lambda \mathbf{p} &= KR[I|R^T \mathbf{t}]\mathbf{P} \\ &= KR[I|-\mathbf{c}]\mathbf{P}\end{aligned}$$

donde $\mathbf{c} = -R^T \mathbf{t}$ es la posición de la cámara con respecto al sistema de coordenadas global.

R tiene tres grados de libertad, que puede verse como tres ángulos de rotación alrededor de los ejes:

$$R = R_z(\theta_3) \cdot R_y(\theta_2) \cdot R_z(\theta_1)$$

Checar

http://en.wikipedia.org/wiki/Euler_angles

Objetos que pueden utilizarse como marcadores

1. Planos

Z. Zhang, A Flexible New Technique for Camera Calibration, IEEE Trans on Patt Anal & Mach Intel (22) 11, 2000, p. 1330–1334.

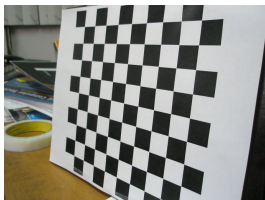
2. Dos círculos concéntricos

J.S. Kim, P. Gurdjos and I.S. Kweon, Geometric and Algebraic Constraints of Projected Concentric Circles and Their Applications to Camera Calibration, IEEE Trans. PAMI, vol 27, 4, Apr 2005, pp. 637-642.

3. Cuboides

M. Wilczkowiak and E. Boyer and P. Sturm, Camera Calibration and 3D Reconstruction from Single Images Using Parallelepipeds, Proceedings of the 8th International Conference on Computer Vision, Vancouver, Canada, Jul. 2001, p. 142–148. <http://perception.inrialpes.fr/Publications/2001/WBS01a>

- ▶ Los métodos anteriores de calibración son fáciles de usar
- ▶ Todos se basan en aplicar restricciones en la matriz de rotación:
 - ▶ La matriz de rotación es ortogonal: $R^{-1} = R^T$ y los vectores columna (y renglón) son unitarios.
 - ▶ Esto también quiere decir que si $R = [\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3]$:
 - ▶ $\mathbf{r}_i \cdot \mathbf{r}_j = 0$, o $\mathbf{r}_i^T \mathbf{r}_j = 0$, para $i, j = \{1, 2, 3\}$ e $i \neq j$
 - ▶ $\|\mathbf{r}_1\| = \|\mathbf{r}_2\| = \|\mathbf{r}_3\|$



$$\lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 & \mathbf{t} \end{bmatrix} \begin{bmatrix} x \\ y \\ 0 \\ 1 \end{bmatrix}$$

$$\lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

$$\lambda \mathbf{p} = \mathbf{H} \mathbf{P}$$

Z. Zhang, A Flexible New Technique for Camera Calibration, IEEE Trans on Patt Anal & Mach Intel (22) 11, 2000, p. 1330–1334.

Si $H = [\mathbf{h}_1 \ \mathbf{h}_2 \ \mathbf{h}_3]$, tenemos

$$[\mathbf{h}_1 \ \mathbf{h}_2 \ \mathbf{h}_3] = \lambda K[\mathbf{r}_1 \ \mathbf{r}_2 \ \mathbf{t}].$$

Sabiendo que \mathbf{r}_1 y \mathbf{r}_2 son ortogonales, tenemos:

$$\mathbf{h}_1^T K^{-T} K^{-1} \mathbf{h}_2 = 0$$

$$\mathbf{h}_1^T K^{-T} K^{-1} \mathbf{h}_1 = \mathbf{h}_2^T K^{-T} K^{-1} \mathbf{h}_2.$$

$$\begin{aligned} \omega &= K^{-T} K^{-1} = \\ &= \begin{bmatrix} \frac{1}{f} & 0 & 0 \\ 0 & \frac{1}{f} & 0 \\ \frac{-u_0}{f} & \frac{-v_0}{f} & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{f} & 0 & \frac{-u_0}{f} \\ 0 & \frac{1}{f} & \frac{-v_0}{f} \\ 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{f^2} & 0 & \frac{-u_0}{f^2} \\ 0 & \frac{1}{f^2} & \frac{-v_0}{f^2} \\ \frac{-u_0}{f^2} & \frac{-v_0}{f^2} & \frac{u_0^2}{f^2} + \frac{v_0^2}{f^2} + 1 \end{bmatrix} = \\ &= \begin{bmatrix} \omega_{11} & 0 & \omega_{13} \\ 0 & \omega_{11} & \omega_{23} \\ \omega_{13} & \omega_{23} & \omega_{33} \end{bmatrix} \end{aligned}$$

Se puede establecer un sistema de ecuaciones homogéneo de la forma:

$$V\mathbf{b} = 0,$$

donde $\mathbf{b} = [\omega_{11}, \omega_{13}, \omega_{23}, \omega_{33}]^T$.

Con una sola imagen, se puede asumir que (u_0, v_0) está en el centro de la imagen, y estimar una solución al sistema de ecuaciones usando DVS.

$$\lambda \mathbf{p} = H \mathbf{P}$$

Los puntos \mathbf{p} y \mathbf{P} los conocemos y podemos establecer un sistema de ecuaciones homogéneo de la forma:

$$A \mathbf{h} = 0$$

Los valores de los puntos en la imágenes deben normalizarse:

$$\begin{aligned} T_1 \mathbf{p} &= H' T_2 \mathbf{P} \\ \mathbf{p}' &= H' \mathbf{P}' \end{aligned}$$

Para recuperar H hacemos:

$$\begin{aligned} T_1^{-1} T_1 \mathbf{p} &= T_1^{-1} H' T_2 \mathbf{P}, \\ \mathbf{p} &= T_1^{-1} H' T_2 \mathbf{P}', \end{aligned}$$

$$\text{y } H = T_1^{-1} H' T_2.$$

La normalización de los datos es a media cero y d.e. 1.

El método lineal para calcular f , u_0 , v_0 , \mathbf{r}_1 , \mathbf{r}_2 y \mathbf{t} se puede mejorar aplicando un estimador de máxima verosimilitud, minimizando:

$$\sum_{i=1}^n \sum_{j=1}^m \|\mathbf{p}_{ij} - \hat{\mathbf{p}}(K, R_i, \mathbf{t}_i, \mathbf{P}_j)\|^2$$

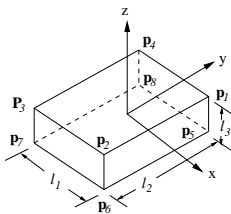
para n imágenes y m puntos por imagen.

- ▶ Se basa en el trabajo de Wilczkowiak (2001) ¹
- ▶ Es posible reconstruir un cuboide a partir de una sola imagen.

¹M. Wilczkowiak and E. Boyer and P. Sturm, Camera Calibration and 3D Reconstruction from Single Images Using Parallelepipeds, Proceedings of the 8th International Conference on Computer Vision, Vancouver, Canada, Jul. 2001, p. 142–148. <http://perception.inrialpes.fr/Publications/2001/WBS01a>

Modelo parametrizado del cuboide

Con el siguiente modelo parametrizado de un cuboide:



El punto \mathbf{P}_1 es igual a $\tilde{\Lambda}(1, 1, 1, 1)^T$, donde $\tilde{\Lambda}$ es igual a

$$\tilde{\Lambda} = \begin{bmatrix} l_1/2 & 0 & 0 & 0 \\ 0 & l_2/2 & 0 & 0 \\ 0 & 0 & l_3/2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

l_1 , l_2 , y l_3 son las tres longitudes de los lados del cuboide.

Los puntos en la imagen $\mathbf{p}_i = [u_i, v_i, 1]^T$, con $i = 1 \dots 8$, satisfacen la ecuación:

$$\begin{bmatrix} \alpha_1 u_1 & \alpha_2 u_2 & \cdots & \alpha_8 u_8 \\ \alpha_1 v_1 & \alpha_2 v_2 & \cdots & \alpha_8 v_8 \\ \alpha_1 & \alpha_2 & \cdots & \alpha_8 \end{bmatrix} = \tilde{X} \cdot \begin{bmatrix} 1 & 1 & \cdots & -1 \\ 1 & -1 & \cdots & -1 \\ 1 & 1 & \cdots & -1 \\ 1 & 1 & \cdots & 1 \end{bmatrix}$$

La matrix \tilde{X} puede estimarse a partir de seis puntos en la imagen

Una estimación por mínimos cuadrados de \tilde{X} para el sistema homogéneo de ecuaciones resultante se obtiene usando la descomposición en valores singulares (DVS).

La matriz de proyección \tilde{X} captura toda la información geométrica contenida en la imagen de proyección del cuboide.

\tilde{X} es la matriz de proyección de tamaño 3×4 la cual está definida hasta un factor de escala como:

$$\tilde{X} \sim M \cdot \tilde{\Lambda} \sim K \cdot [R|\mathbf{t}] \cdot \tilde{\Lambda}$$

Considerando la relación $X \sim K \cdot R \cdot \Lambda$, entonces $K^{-1}X \sim R \cdot \Lambda$, elevando esta última al cuadrado y considerando que $R^T R = I$, se obtiene:

$$X^T \cdot K^{-T} \cdot K^{-1} \cdot X \sim \Lambda^T \cdot \Lambda$$

Tenemos tres incógnitas en K , que son los valores para f , u_0 y v_0 . De $\omega = K^{-T} \cdot K^{-1}$ resulta:

$$\omega = \begin{bmatrix} \omega_{11} & 0 & \omega_{13} \\ 0 & \omega_{11} & \omega_{23} \\ \omega_{13} & \omega_{23} & \omega_{33} \end{bmatrix}$$

que se resuelve aplicando las siguientes restricciones lineales, donde X_i es la i -ésima columna de X :

1. $X_i^T \cdot \omega \cdot X_j = 0$.
2. Para cada relación de largos de la caja $r_{ij} = l_i/l_j$:
 $X_i^T \cdot \omega \cdot X_i - r_{ij}^2 X_j^T \cdot \omega \cdot X_j = 0$.

puede establecer un sistema de ecuaciones homogéneo y obtenerse la estimación para ω usando DVS

$$R = \frac{1}{\lambda} K^{-1} X \Lambda^{-1}$$

R no cumple las propiedades de una matriz ortogonal ($R^{-1}R = R^T R = I$), por lo tanto para cumplir estas propiedades se forzan las condiciones de ortogonalidad usando de nuevo DVS.

\mathbf{t} se determina de:

$$\begin{aligned}\tilde{X} &= \lambda K \cdot [R | \mathbf{t}] \tilde{\Lambda}, \text{ y} \\ \lambda [R | \mathbf{t}] &= K^{-1} \tilde{X} \tilde{\Lambda}^{-1}\end{aligned}$$

$\lambda = 1/\|\mathbf{r}_1\| = 1/\|\mathbf{r}_2\| = 1/\|\mathbf{r}_3\|$, donde \mathbf{r}_i es la i -ésima columna de la matriz R .

Pasos necesarios en un sistema de realidad aumentada

- Paso 1 Procesamiento de imagen.** Involucra la adquisición de las imágenes, el filtrado de ellas para mejorar sus condiciones, y segmentar el marcador en cada marco del video.
- Paso 2 Visión.** Consiste en la *calibración de la cámara y la obtención de los parámetros de orientación y posición* para la vista correspondiente en cada imagen.
- Paso 3 Visualización.** Agregar los objetos virtuales y dibujarlos.

Conclusiones (2/2)

1. El objetivo de visión es la reconstrucción tridimensional de la superficie de objetos, a partir de imágenes bidimensionales. Esto es, nos permite reconstruir la estructura de la superficie del objeto.
2. Un sistema de realidad aumentada usa técnicas de procesamiento de imágenes, de visión y de graficación.
3. Para visión, es necesario tener conocimiento de algebra lineal y análisis numérico.
4. La parte más difícil de visión es la calibración de la cámara, esto es, encontrar la transformación proyectiva que realiza la cámara entre el mundo 3D y las imágenes en 2D.

Ofrecemos la **maestría** y **doctorado** en
Ciencias de la Computación.

- ▶ Página del Departamento: <http://www.cs.cinvestav.mx>
- ▶ Página personal: <http://cs.cinvestav.mx/~fraga>
- ▶ Correo-e: fraga@cs.cinvestav.mx