

Base de datos de imágenes de gestos manuales para una interface Tabletop

Andoni Beristain Iraola
25 de febrero de 2009

Informe Interno GIC-UPV/EHU-RR-2009-02-25



eman ta zabal zazu



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

Este documento presenta la creación de una base de datos de imágenes binarias donde se muestran secuencias de imágenes correspondientes a la ejecución de 3 gestos manuales denominados agarrar, señalar y pasar página.

Base de datos de imágenes de gestos

1.	<u>INTRODUCCIÓN</u>	3
2.	<u>CREACIÓN DE LA BASE DE DATOS</u>	4
2.1.	<u>SELECCIÓN DE POSES SIGNIFICATIVAS</u>	4
2.1.1.	<u>GESTO AGARRAR</u>	4
2.1.2.	<u>GESTO PASAR PÁGINA</u>	5
2.1.3.	<u>POSE SEÑALAR</u>	6
2.2.	<u>CREACIÓN DE MUESTRAS DE CADA UNA DE ESTAS POSES SIGNIFICATIVAS</u>	6
3.	<u>DESCRIPCIÓN DE LA BASE DE DATOS</u>	9
3.1.	<u>FORMATO DE NOMBRE DE FICHEROS</u>	9
3.2.	<u>FORMATO DE LAS IMÁGENES</u>	9



1. Introducción

El marco de creación de esta base de datos experimental de imágenes es el trabajo sobre el diseño de interfaces tipo Tabletop, esto es, interfaces de manipulación de información multimedia sobre un dispositivo de visualización con las dimensiones y configuración de una mesa de trabajo. La base de datos se crea para el estudio de sistemas de clasificación y reconocimiento de los gestos manuales que los usuarios pueden hacer para manipular la información y su visualización en la Mesa Interactiva Multimedia (MIT: Multimedia Interface Tabletop).

Este documento describe el proceso de creación y las características de una base de datos de imágenes binarias de manos mostrando varias poses diferentes. Se han definido dos gestos dinámicos, uno de ellos con dos variantes, y una pose. Los gestos son agarrar objeto (pinzar y no pinzar) y pasar página. Y la pose corresponde a señalar un objeto con el dedo índice. El reconocimiento de los gestos dinámicos implica el reconocimiento de una secuencia de imágenes, mientras que el reconocimiento de la pose sólo implica el reconocimiento de una imagen clave que captura el gesto.

Cada uno de estos gestos está definido por un conjunto de poses clave. Esta base de datos contiene secuencias de imágenes correspondientes a la animación de cada uno de los gestos.

El objetivo inicial de la base de datos es servir como banco de pruebas para la hipótesis de trabajo de que el esqueleto o medial-axis puede ser utilizado como vector de características adecuado para el reconocimiento y distinción de cada una de las poses definidas.

Esta base de datos puede utilizarse libremente para investigación bajo la licencia GPL. Si se utiliza en trabajos publicados agradeceremos que se referencie este informe interno y las publicaciones que irán apareciendo en la página web del grupo de Inteligencia Computacional relativas a esta base de datos.



2. Creación de la base de datos

El proceso de construcción de la base de datos ha seguido los siguientes pasos:

- Selección de las imágenes o secuencias de imágenes más significativas para cada uno de los gestos y las poses de interés.
- Creación de muestras de cada una de estas poses significativas variando de forma aleatoria los parámetros de síntesis de la imagen.

2.1. Selección de poses significativas

En este apartado se muestran y describen todas las poses y gestos relevantes para nuestra investigación. Estas poses y gestos están caracterizados por una o más imágenes clave obtenidas de las de las animaciones de cada gesto.

Todas las poses se corresponden a gestos realizados con la mano izquierda.

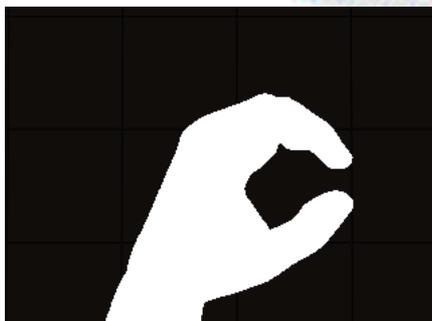
2.1.1. Gesto agarrar.

El gesto agarrar consiste en cerrar la mano, como en el caso de agarrar un vaso. Tiene dos versiones: la primera termina cuando el índice y el pulgar llegan a tocarse y la segunda cuando se acercan lo suficiente. Por ello hemos definido 3 imágenes clave para este gesto:

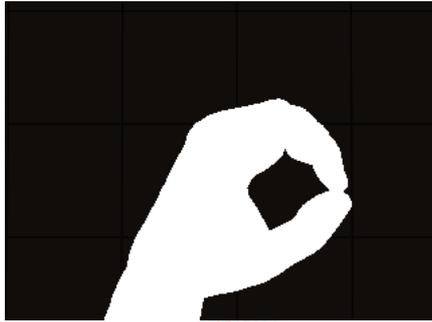
- Mano empezando a agarrar: AGARRAR_A



- Mano en posición de agarrar sin pinzar: AGARRAR_B



- Mano en posición de agarrar pinzando (los dedos índice y pulgar se tocan) :
AGARRAR_C



2.1.2. Gesto pasar página

El gesto de pasar página consiste en ondear la mano de derecha hacia izquierda o viceversa, con la mano extendida y ladeada ligeramente durante el gesto. El gesto de pasar página está formado por 3 imágenes clave:

- Mano iniciando el gesto de pasar página: PPAGINA_A



- Mano entre las dos páginas: PPAGINA_B



- Mano al final de pasar página: PPAGINA_C



2.1.3. Pose señalar

Señalar es una pose y no un gesto, es decir consiste en una forma estática concreta de la mano y no implica movimiento alguno. Su imagen clave característica es:

- Imagen clave de señalar: SENALAR



2.2. Creación de muestras de cada una de estas poses significativas

Para las poses básicas empleamos la herramienta Poser® de la compañía e-Frontier que nos permite crear la animación de un gesto manual y crear imágenes de muestra de dicho gesto con variaciones de la mano en cuanto a grosor, tamaño, posición y orientación. Se han creado animaciones que corresponden a los dos gestos dinámicos completos, que son pasar página y agarrar, y la pose adecuada para señalar. En el caso de la pose señalar, se ha generado una animación en la que la mano ondea de derecha a izquierda. Todos los gestos se han realizado empleando únicamente la mano izquierda, ya que el vector de características final deberá ser invariante a rotación y escala.

Los dos gestos tienen tres poses significativas cada uno. Todas las animaciones creadas en Poser tienen una longitud de 30 imágenes. En el caso de los gestos se han definido 3 fotogramas clave que corresponden a las 3 poses. Los fotogramas clave han sido el 1, el 15 y el 30. Y para el caso de la pose señalar se han creado también 3 fotogramas clave en las mismas posiciones con la diferencia de que la variación entre ellas se limita a un ligero desplazamiento de la mano de izquierda a derecha.

Los pasos para realizar la animación de la versión básica de cada gesto han consistido en lo siguiente:

- Seleccionar el modelo de mano del apartado Additional figures.



- Posicionar la cámara principal de forma que enfoque la mano como en las imágenes de ejemplo.
- Modificar la pose de la mano hasta adoptar la pose inicial del gesto y crear un fotograma clave en la posición 1 de la barra de tiempo.



- Avanzar al fotograma 15 y crear otro fotograma clave con la pose adecuada a la mitad del gesto.
- Avanzar al último fotograma 30, el último y añadir un fotograma clave con la pose final del gesto.
- Modificar las propiedades de los materiales y la iluminación de la escena para obtener imágenes en blanco y negro. Posteriormente se han modificado las propiedades de las texturas del fondo y de la mano para que las imágenes renderizadas de los gestos obtenidas sean binarias, en lugar de ser en color. De esta manera se ahorra el cómputo de la iluminación y las sombras por parte de la herramienta Poser y además se reduce el evita el proceso de binarización posterior.



Las imágenes resultantes han sido almacenadas en formato BMP de 3 canales en escala de grises con una resolución de 320 x 240.

Después se ha automatizado el proceso de creación de muestras a través de las facilidades que ofrece Poser. Esta herramienta permite la automatización de la mayoría de operaciones realizables a través de la interfaz gráfica por medio de unas clases específicas sobre la base del lenguaje de scripting Python. Se ha creado un script para cada gesto, de forma que puedan crearse variantes de cada gesto base y se almacenen como muestras de imagen. Solamente se



almacenan las imágenes pares de cada animación, que tiene una longitud de 30 imágenes. La razón básica consiste en que para nuestra aplicación es interesante que haya una diferencia apreciable entre imágenes consecutivas.

La creación de cada muestra empleando un gesto base sigue el siguiente esquema:

- Para cada parámetro Global del modelo de la mano (Escala x, escala y, escala z)
 - Modificar de forma aleatoria en el rango definido para cada gesto.
- Modificar la rotación de la muñeca de forma aleatoria en el rango definido.
- Modificar la inclinación de la muñeca de forma aleatoria en el rango definido.
- Modificar el grado de contracción del pulgar.
- Modificar el grado de agarre general de todos los dedos de la mano.
- Modificar la separación entre los dedos de la mano.
- Para cada parte de la mano.
 - Modificar el grosor de la parte de la mano de forma aleatoria. Todas las partes se modifican en la misma medida.

Todas las modificaciones se realizan sobre el primer fotograma de la animación del gesto, de manera que los cambios realizados afectan a todo el gesto.



3. Descripción de la base de datos

Actualmente la base de datos está formada por 3 directorios con 3000 imágenes en cada directorio. Los directorios son AgarrarVerticeI, ApuntarI y PPaginaI, y respectivamente almacenan los gestos agarrar, señalar y pasar página para gestos realizados con la mano izquierda.

En cada directorio se encuentran 200 secuencias de 15 imágenes cada una. Cada secuencia de imágenes muestra una animación del gesto que contiene dicho directorio.

3.1. Formato de nombre de ficheros

La nomenclatura de los ficheros sigue este formato:

Nombre_directorio_Nº de secuencia_Nº imagen.bmp

- Nombre_directorio: es el nombre del directorio en el que se encuentra la imagen, que corresponde al nombre del gesto. Los valores son: AgarrarVerticeI, ApuntarI y PPaginaI.
- Nº de secuencia: Número identificador de la secuencia. Comienza con el valor 0 y termina en el valor 199.
- Nº imagen: Número de frame en la secuencia de imágenes.

Ejemplo: AgarrarVerticeI_3_4.bmp. Esta imagen corresponde al gesto agarrar, que se encuentra en el directorio AgarrarVerticeI. La imagen corresponde al frame 4 de la secuencia de imágenes 3.

3.2. Formato de las imágenes

Tipo: Windows Bitmap (bmp).

Tamaño: 320 * 240.

Color: 1 canal escala de grises.

Resolución: 71 ppp.

Profundidad de color: 4 bits.

Las imágenes son binarias con el fondo negro y la mano en blanco y están almacenadas en el formato Windows Bitmap (extensión BMP) para Windows en escala de grises y profundidad de gris de 4 bits. El tamaño de la imagen es de 320*240 píxeles y la resolución tanto vertical como horizontal es de 71 ppp.



Grupo de Inteligencia Computacional 2009

Autor:

Andoni Beristain Iraola.

beristainandoni@yahoo.es



