

UNIVERSIDAD DE ESTUDIOS DE ROMA “LA SAPIENZA”

Facultad de Ciencias Matemáticas, Físicas y Naturales
Programa Erasmus – Ingeniería Informática

Proyecto Fin de Carrera

Reestructuración de la GUI y de los algoritmos de correlación de curvas para un sistema CBIR

Gema Segura
Universidad Carlos III de Madrid

Año Académico 2008-2009

ÍNDICE

1 INTRODUCCIÓN.....	3
2 ESTADO DEL ARTE.....	4
3 TRABAJO PREVIO	5
4 SISTEMA PROPUESTO.....	7
5 ESTRUCTURA DE LA BASE DE DATOS Y MÉTODOS DE INSERCIÓN.....	9
CONCLUSIONES.....	11

Esta tesis presenta un sistema para el reconocimiento de elefantes. Este trabajo ha sido realizado en colaboración con un grupo de investigadores del Departamento de Zoología de la Universidad La Sapienza, que trata del estudio del comportamiento de la especie. Ser capaz de identificar diferentes especímenes dentro de un grupo es estudiar las características (hábitos, movimientos, esperanza de vida, etc). Hasta la fecha, este reconocimiento se lleva a cabo mediante la observación y la comparación de las fotos tomadas en los parques africanos en diferentes períodos. Esta etapa es cada vez más difícil por el hecho de que los individuos difieren entre sí por muy pocos detalles. Una de ellas es la presencia de muescas en las orejas. La forma y posición de dichas muescas parece ser un rasgo distintivo casi como las huellas dactilares para los humanos.

El objetivo era crear un sistema que automatizase este proceso de reconocimiento. El sistema inicial fue creado por A. Ardovini. En esta segunda versión del sistema se ha aplicado un nuevo algoritmo que permite una mayor automatización del proceso. En consecuencia, algunos aspectos de la interfaz gráfica de usuario han cambiado, resultando en una interfaz más eficaz y amigable.

El sistema que se muestra aquí extrae el perfil de la oreja, identificando los cortes y, a través de una técnica de correlación de bordes son seleccionados los perfiles de la base de datos más similares a cada oreja.

El proceso de reconocimiento de un individuo se compone de cuatro pasos básicos. El primer paso consiste en la extracción del perfil del individuo a través de un operación de detección de bordes semi-automática, y la selección de las muescas de dicho perfil. Las muescas seleccionadas pasan a formar parte de un vector de consulta. En el siguiente paso todos los perfiles de la base de datos correspondiente al mismo lado de la consulta se consideran. Estos perfiles deben tener al menos una muesca en una posición similar a una de las de la entrada del perfil. El tercer paso es comparar las curvas de consulta con el borde de los perfiles previamente seleccionados. El último paso permite la extracción de una clasificación final de N individuos sobre la base de su similitud con la entrada de perfil.

En la literatura se encuentran diferentes enfoques con respecto a la correlación de bordes y casi todos hacen uso de las transformaciones de similitud. Uno de estos métodos son las técnicas de shape-signature, cuyo objetivo es encontrar la sub-curva común más larga de dos curvas 2D. Estas curvas se convierten primero en cadenas llamadas de shape-signature y se aplican técnicas de comparación de cadenas para encontrar sub-cadenas coincidentes, seguida por la comparación directa de las correspondientes sub-curvas candidatas para encontrar aquella que coincida en mayor longitud. Este método resuelve el problema de las curvas en general sin depender de un conjunto de puntos especiales en las curvas. Otro método conocido utiliza modelos ocultos de Markov (HMM). Los HMM son modelos matemáticos de procesos estocásticos. Las características de una imagen son convertidas en código y representan la entrada de estos modelos. Dichos modelos, con un enfoque probabilístico, tratará de reconocer la propia imagen.

Una especie de medida local que representa los límites físicos de los objetos fue utilizado para la metodología de la curvatura del Espacio Escala (CSS) [3]. Se utiliza una caja de luz para obtener imágenes de siluetas que se clasifican como convexas o cóncavas. Representaciones CSS se calculan para las curvas cóncavas. La representación CSS es una organización multi-escala de las características naturales, invariables de una curva. Un algoritmo de tres etapas reduce el espacio de búsqueda mediante la aplicación de la prueba de relación de aspecto CSS. Los máximos de los contornos en las representaciones CSS de los modelos que sobreviven son utilizadas para correlación rápida CSS en la segunda etapa. Por último, la tercera fase verifica la mejor coincidencia y resuelve cualquier ambigüedad a través de la determinación de la distancia entre la imagen y las curvas del modelo. La optimización de parámetros de transformación se utiliza para encontrar el mejor ajuste del objeto de entrada para el modelo correcto.

Estos métodos no están diseñados para abordar problemas de correlación entre las curvas sujetas a las transformaciones afines. Por lo tanto, no cumplen plenamente los requisitos de nuestro problema, ya que nuestras imágenes son capturadas desde diferentes puntos de vista. Sin embargo, la mayoría de los métodos de correlación de bordes se adaptan a nuestro problema, utilizando este tipo de transformaciones. Estos incluyen: el método del espacio de escala de curvatura afín, el diseñado por Cohen utilizando las curvas B-Spline y una técnica de reconocimiento de caracteres, de Flusser, que también puede aplicarse para hacer coincidir las curvas. Y, por último, el método elegido para este sistema, Chamfer Matching, que se explica en la Sección. 3.

3 Trabajo previo

En la literatura hay muy pocas aplicaciones de correlación de bordes para el reconocimiento de animales. Estos trabajos extraen de la silueta del animal algunas características especiales, que son luego procesadas para el reconocimiento.

Uno de estos sistemas se ocupa del reconocimiento de los mamíferos marinos (delfines, ballenas, lobos de mar) a través del estudio de sus aletas. Las curvas que describen las aletas se extraen de las imágenes y, a continuación son muestreadas, eligiendo los diez puntos que mejor definan cada una de ellas. La fase de correlación consiste en una transformación afín de los puntos elegidos con respecto a los puntos de la curva de la base de datos considerada. La distancia entre las dos curvas se calcula a continuación sumando las diez distancias euclidianas, cada una de las cuales se calcula entre el punto i de la curva de consulta y el punto i de la curva de la base de datos. La curva de consulta será reconocida como aquella curva de la base de datos que ofrezca la distancia mínima. Un enfoque similar fue utilizado en otro de estos sistemas para el reconocimiento de las ballenas.

El único sistema utilizado para el reconocimiento de los elefantes ha sido propuesto por un grupo de investigadores de la Universidad de Ámsterdam. Su método se basa en el estudio de todo el contorno de la oreja. Estas curvas, después de una transformación afín, se comparan con las curvas que figuran en la base de datos para su reconocimiento. Su modo consiste en muestrear todas las curvas con 500 puntos y luego calcular la distancia euclidiana.

En la versión previa de nuestro sistema, un enfoque similar fue utilizado. Pero en lugar de considerar toda la curva que representa el contorno de la oreja, el método aplicado se basó en la comparación de sólo aquellas partes del contorno de la oreja que correspondían a las muescas individuales. Después de la extracción semiautomática de los bordes, el usuario seleccionaba el punto inicial y final de cada muesca, siempre y cuando ambos puntos estuviesen conectados por una línea continua. Para que esta condición se cumpliera el usuario estaba obligado a realizar ajustes a los contornos de la imagen a través del uso de herramientas de dibujo.

La eliminación de este proceso representa el cambio más importante en lo que a facilidad de uso se refiere, ya que una característica del nuevo sistema es la de automatizar el proceso de extracción de las muescas. Por lo tanto, el proceso es más rápido, y evita la alteración subjetiva de la imagen por parte del usuario.

Volviendo a la versión anterior del sistema, después de la operación de ajuste mencionada cada curva de consulta era muestreada a 25 puntos y comparada con el resto de la base de datos para su reconocimiento. Sin embargo, este método tiene algunas limitaciones. Las fotos de los elefantes que satisfagan los supuestos de estos sistemas debe tener el contorno de la oreja limpio y sin ruido (clima, luminosidad, ramas de los árboles) para que la operación de extracción de bordes devuelva una curva continua para cada oreja. La continuidad de la curva es esencial para la toma de muestras. Con respecto a nuestras imágenes, esta hipótesis es muy restrictiva, ya que es muy poco probable obtener fotografías cuyo contorno visible de la oreja esté perfectamente limpio.

De hecho, las curvas discontinuas representan casi todos los casos. Esta es la razón por la que nuestro sistema ha adoptado un nuevo enfoque, el algoritmo de correlación de Chamfer, que se introdujo por primera vez en Barrow. et Al. "Parametric correspondence and chamfer matching: Two new techniques for image matching,". Esta técnica permite encontrar el mejor ajuste de los puntos de dos imágenes diferentes, minimizando la distancia generalizada entre ellos. En este método, dos imágenes binarias basadas en los puntos presentes de un determinado objeto se comparan. El objeto puede ser cualquiera que esté bien definido y sea visible en ambas imágenes. Para nuestro sistema estas características corresponden a las muescas presentes en los perfiles. Las dos imágenes no se tratan de forma simétrica. Una representa la transformación de distancia (DT) y la otra el pre-polígono. La imagen DT se crea a partir de la asignación a cada píxel no perteneciente a ningún borde de un valor que es una medida al píxel más cercano que sí pertenezca a un borde. Para nuestro sistema se ha definido que una muesca de consulta corresponde a un pre-polígono, y que los cortes de la base de datos para contrastar con la consulta representan las pre-distancias.

El proceso de convertir una imagen binaria a una imagen de distancia aproximada se llama la transformación de distancia (DT). El cálculo original de la distancia euclidiana exigía muchos recursos (memoria), por tanto utilizaremos una aproximación. Para el

algoritmo de correlación, es importante que la DT sea una aproximación razonablemente buena de la distancia euclidiana. Las distancias de Chamfer están definidas entre vecinos en dirección horizontal/vertical/diagonal para conjuntos de 3x3. En la imagen de pre-polígono, que corresponde a una curva de consulta, los píxeles del borde se extraen y se convierte en una lista de pares de coordenadas. Cada par es el número de fila y columna de un píxel del borde. De esta lista, los puntos que se utilizan realmente son los puntos modificados de acuerdo a los parámetros de transformación obtenidos a partir de los cortes a relacionar. El pre-polígono se superpone a la imagen de distancias. La medida de la correspondencia entre los bordes es el promedio de los valores de la imagen de distancias coincidentes con los píxeles del prepolígono. Un ajuste perfecto entre dos bordes se traducirá en una distancia de cero. La equiparación real consiste en reducir al mínimo la distancia al borde. Hay muchas variantes de medidas media de correlación, por ejemplo, la aritmética, la raíz cuadrática media y la mediana.

Aparte de este importante cambio del algoritmo en el que se basa el sistema, la mayoría de los otros procesos presentes en la versión anterior de nuestro sistema se han conservado en esta nueva versión. Por ejemplo, el sistema todavía está fuertemente basada en la posición de los cortes individuales en relación con el resto de la oreja. Para ello, un sistema de coordenadas se crea sobre la base de la posición del ojo, y un punto del perfil elegido después de la apertura de la imagen. De esta manera es posible utilizar la información de tipo contextual en la oreja, en general, y manejar la rotación, escalado y la transformación. En las secciones siguientes explicaremos esto en profundidad.

4

Sistema Propuesto

La arquitectura de nuestro sistema se compone de cinco módulos básicos: Extracción de Perfil, Selección de Muestras, Perfiles Candidatos, Correlación y Clasificación Final.

El sistema recibe como entrada la imagen con la oreja del individuo de la consulta. Esta imagen se extrae de la foto original del individuo en el momento de su apertura. Durante esta fase, además de la selección de la oreja, se solicita al usuario hacer clic en cuatro

puntos clave, es decir, el ojo del individuo, el lado extremo de la oreja, el lóbulo de la oreja y la conexión superior de la oreja con la cabeza. Las fotos deben tener la característica de representar a los elefantes de perfil, de modo que estos puntos sean visibles. Estos puntos se utilizan para identificar las posiciones asumidas por los cortes de consulta en relación a la oreja.

El primer paso del sistema es la aplicación del módulo de Extracción de Perfil. Este módulo recibe como entrada de la imagen con la oreja y después de la detección de bordes devuelve otra que contiene su perfil.

El módulo siguiente, es decir, el de Selección de Muestras, permite la extracción de los cortes del perfil obtenido anteriormente. La selección de cada una de ellas es realizada por el usuario simplemente haciendo clic en dos puntos extremos de la muestra, mientras que su extracción se realiza automáticamente. La salida de este módulo es la lista que contiene los cortes extraídos de la consulta.

Para acotar la búsqueda en la base de datos a fin de reducir el coste computacional de todo el proceso de adecuación de la muestra de consulta, se creó un algoritmo para seleccionar un subconjunto de los perfiles que se comparará con la entrada. El módulo que realiza esta operación es el módulo de Perfiles Candidatos. Cuando se busca en la base de datos, sólo los perfiles de los individuos con cortes seleccionados del mismo lado (la oreja derecha si la oreja de consulta es la derecha, izquierda, de otra manera) se extraen. Por lo tanto, la salida es un vector que almacena los perfiles del subconjunto correspondiente en la base de datos. Además, estos perfiles deben tener sus propios cortes en posiciones similares a las de la propia consulta.

La lista de las muestras de consulta junto con el vector de los perfiles descrito anteriormente, representan la entrada del siguiente módulo de Correlación. Las muestras de la consulta serán comparadas con las de cada perfil del vector de entrada y el grado de similitud se almacena dentro de un nuevo vector. El grado de similitud está dado por el cálculo de la distancia de Chamfer para el corte en cuestión.

Este vector es luego procesado por el último módulo, denominado Clasificación Final, a fin de devolver la clasificación final con los N individuos de la base de datos que más se asemejan a la consulta. El módulo también visualiza las imágenes de los individuos que forman parte de la clasificación en orden.

Estructura de la Base de Datos y Métodos de Inserción

En este capítulo se describe la estructura de la base de datos, tanto en la memoria principal como en la memoria secundaria, y los métodos de inserción en la misma. La estructura ha cambiado desde la versión anterior del sistema a éste. En la nueva versión, la carpeta *tag* ya no se utiliza. La carpeta *foto* no cambia. El principal cambio es la inclusión de una nueva carpeta *map* que contiene los archivos *.tag* que han sido generados a partir de los almacenados en *tag*. La nueva información del archivo es:

- 1) El número de muescas seleccionadas en la foto
- 2) Las coordenadas del ojo
- 3) Las coordenadas del lateral extremo de la oreja
- 4) Las coordenadas del lóbulo
- 5) Las coordenadas de la conexión de la oreja con la cabeza
- 6) Las coordenadas de los puntos extremos de cada muesca

La carga de la base de datos en memoria principal se realiza con seis estructuras, tres para el conjunto de las orejas derechas y tres para el conjunto de las orejas izquierdas. Los seis son listas de listas de longitud N, donde N es el número de individuos en la base de datos. La primera estructura contiene los puntos extremos de las muescas leídos de los archivos *.tag*. La segunda estructura contiene las cadenas que representan las rutas completas de las imágenes almacenadas en el almacenamiento secundario. La tercera estructura contiene las coordenadas del ojo, al lado extremo de la oreja, del lóbulo y de la conexión de la oreja con la cabeza, tomados de los archivos *.tag* correspondientes a cada imagen. Esto nos permitirá gestionar la información contenida en la base de datos accediendo sólo a la memoria principal, lo que limita el tiempo para la fase de lectura. En términos de las estructuras de datos que acabamos de describir, la fase que extrae los perfiles candidatos realiza una lectura de la base de datos desde el primer individuo.

Para cada imagen se lleva a cabo el control que se refiere a la comparación de la posición de la muesca k1 del perfil de la base de datos tomado en consideración con la posición de la muesca k2 del individuo de consulta.

Esto lo hacemos por cada par de muescas [consulta-BD] existente. Para ello se utiliza la información de la primera estructura, así como los puntos clave de referencia tomados de la tercera estructura. Si el resultado es positivo la fase de cotejo puede ocurrir entre dos individuos. Mientras que se realiza la correlación, se generará el mapa de Chamfer para la *j* imagen del individuo *i* a examen. Recordemos que el camino a esta imagen se almacena en la segunda estructura. La estructura que contiene las rutas de acceso a las imágenes será útil también después de la comparación para mostrar la imagen del individuo en caso de que entre en la clasificación final. Una vez que se visualiza la clasificación final, podemos elegir (si se considera apropiado) incluir el individuo de consulta en la base de datos, usando tres diferentes tipos de entradas:

1) Como una nueva foto de un individuo presente en la clasificación final. El sistema automáticamente creará el nombre estándar que se asignará a la imagen y al archivo *.tag*, que se añadirán al individuo que ya existe. Luego se crea el archivo *.tag*, y éste y la imagen se insertan en la base de datos y las estructuras de memoria principal son actualizadas.

2) Manualmente, es decir, navegando por la base de datos. El segundo tipo de integración es muy similar a la primera, pero es necesario cuando se desea insertar manualmente el individuo. Esto sucede cuando el elefante que más se parece a la consulta está en la base de datos, pero no ha sido incluido en la clasificación final. El usuario podrá navegar por la base de datos y seleccionar el directorio relativo al individuo donde colocar la imagen de consulta, después de lo cual las operaciones de actualización de las estructuras en memoria primaria y secundaria serán idénticas a las de la primera inserción.

3) Creación de un nuevo individuo, si el individuo de consulta no se parece a ninguno de los presentes en el ranking final. Este tipo de entrada permite al usuario crear un nuevo individuo. El usuario introduce el nombre del nuevo individuo y el sistema creará los subdirectorios foto y mapa, e insertará los archivos con el mismo criterio descrito anteriormente. Después de esto las estructuras en la memoria principal también serán actualizadas.

En los tres tipos de inserción la actualización de la base de datos se lleva a cabo primero en el almacenamiento secundario y después, si todo ha ido bien, las estructuras se actualizan en la memoria principal.

Conclusiones

En esta tesis un sistema de CBIR especializado para la recuperación de imágenes de interés zoológico ha sido reestructurado y mejorado. El sistema toma como entrada una imagen de una cabeza de elefante y devuelve como resultado una lista de individuos almacenados en la base de datos del sistema, ordenados de acuerdo a la similitud con la imagen de entrada. La comparación entre la imagen de consulta y la imagen de la base de datos se produce al comparar las curvas que describen las muescas presentes en el borde de las orejas de los elefantes adultos. Los zoólogos consideran la posición y forma de estas incisiones lo suficientemente características como para identificar a un individuo específico. El sistema a través de una interfaz gráfica de usuario, permite al usuario seleccionar los cortes en la foto de consulta y, posteriormente, un método completamente automático de correlación de curvas permite comparar la posición y forma de las muescas de consulta las de los individuos almacenados en la base de datos. El propósito de esta herramienta es permitir la automatización y la aceleración del proceso de reconocimiento realizado por zoólogos.

El objetivo de la tesis ha sido doble:

- Mejorar la interfaz gráfica de usuario existente, haciéndola más eficaz y flexible,
- Mejorar el algoritmo de correlación de curvas

El trabajo consistió en la modificación de un software existente (parcialmente documentado), desarrollado en Java. Esta modificación reside principalmente en la adaptación del sistema a un nuevo algoritmo para la comparación de imágenes denominado algoritmo de correlación de Chamfer, y en adaptar la aplicación a este nuevo algoritmo implementado. Para ello, ha sido necesario modificar la fase de correlación y la interfaz de usuario. En cuanto a esta última, además de añadir nuevas funcionalidades, procedí a eliminar las opciones que ya no eran necesarias debido a la mayor automatización del sistema. También realicé la modificación de algunos elementos visuales con el fin de modernizar la interfaz gráfica de usuario, haciéndola más práctica e incluyendo algunas animaciones destinadas a informar a los usuarios de ciertos procesos que tienen lugar durante el uso del sistema.

En trabajos previos, este sistema ha estado funcionando para un grupo de investigadores del Departamento de Zoología de la Universidad La Sapienza. El reconocimiento de un individuo es de primordial importancia ya que permite estudiar su comportamiento. Esta fase se basa en la comparación de fotos de muestras tomadas en los parques de África en diferentes períodos. Para complicar aún más la situación se da el hecho de que los especímenes se distinguen por muy pocos detalles, uno de los cuales es la presencia de muescas en las orejas.

El sistema extrae de un modo semi-automático el perfil del individuo a reconocer y, utilizando la técnica de coincidencia de curvas del algoritmo de Chamfer, muestra los diez individuos en la base de datos que, de acuerdo a su distancia, se acercan más a la consulta. Durante la selección de los perfiles de los candidatos en la base de datos para el cotejo, el sistema también considera la posición de cada muesca en relación con las orejas. Así, para el cálculo de la posición, el usuario puede llevar a cabo el proceso de correlación a partir de un sistema de referencia cartesiano, a elegir entre tres posibles: ojo-extremo de la oreja, ojo-lóbulo u ojo-conexión cabeza. Esta elección es necesaria porque en nuestras imágenes no siempre es claramente perceptible un mismo punto, debido al ruido o quizás a posiciones especiales que pueden tener una oreja durante la captura de la fotografía.

El sistema implementado utiliza una base de datos de 255 individuos diferentes, y nuestros experimentos se realizaron sobre un conjunto de 43 individuos. Los porcentajes para la clasificación de estas 43 consultas en los diez primeros fueron: 88,37% para el sistema de referencia de ojo-extremo de la oreja, 62,79% para el sistema de ojo-lóbulo, y 13,95% para el sistema de ojo-conexión cabeza.

El coste computacional depende del número de muescas del individuo a reconocer. Se ha optado por considerar también el entorno de los puntos extremos del corte, ya que es difícil seleccionar siempre el mismo par de puntos que delimitan la curva. Esto podría resultar en una forma diferente de reconocer una misma muesca. Esto hace que nuestro sistema sea más robusto, incluso si esto implica un aumento de los tiempos de ejecución. El tiempo promedio para la fase de reconocimiento varía en función del número de muescas que presenta las consultas individuales, y está influida también por el número de perfiles presentes en la base de datos.

El sistema final incluye el software, ejecutables, y una pequeña guía para la GUI.