

Laboratorio de Métodos Numéricos - Primer Cuatrimestre 2014

Trabajo Práctico Número 2: Tu cara me suena

Introducción

A menos de cincuenta días del inicio de la gran fiesta del fútbol mundial, el Equipo de Desarrollos de Métodos Numéricos ha recibido un pedido reciente por parte del Comité de Seguridad de la Organización del Mundial Brasil 2014. Informaciones de último momento aseguran que una gran cantidad de *barras* se están organizando para *copar* las tribunas de los (aún no terminados) estadios del país organizador. Debido al retraso en la implementación del sistema FIFA Plus por cuestiones técnicas, el evento será cancelado a menos que nuestro equipo logre desarrollar un software de reconocimiento facial efectivo en los próximos 20 días.

El Comité Organizador ha puesto a nuestra disposición todos sus recursos a fin de evitar lo que sería la mayor vergüenza en la historia de la entidad. Entre otras cosas, se dispone de una base de datos de fotografías digitalizada de todas aquellas personas que alguna vez hayan ingresado a algún estadio de fútbol en algún país del planeta. El objetivo del trabajo es desarrollar un software que, tomando esta información como entrenamiento, pueda detectar en tiempo real si los asistentes a alguno de los partidos corresponde a un *barra*.

Como instancias de entrenamiento, se tiene un conjunto de N personas, cada una de ellas con M imágenes (eventualmente) distintas sus rostros en escala de grises del mismo tamaño y resolución. Cada una de estas imágenes sabemos a qué persona corresponde.

Para $i = 1, \dots, n$, sea $x_i \in \mathbb{R}^m$ la i -ésima imagen de nuestra base de datos almacenada por filas en un vector, y sea $\mu = (x_1 + \dots + x_n)/n$ el promedio de las imágenes. Definimos $X \in \mathbb{R}^{n \times m}$ como la matriz que contiene en la i -ésima fila al vector $(x_i - \mu)^t / \sqrt{n-1}$, y

$$A = U\Sigma V^t$$

a su descomposición en valores singulares, con $U \in \mathbb{R}^{n \times n}$ y $V \in \mathbb{R}^{m \times m}$ matrices ortogonales, y $\Sigma \in \mathbb{R}^{n \times m}$ la matriz diagonal conteniendo en la posición (i, i) al i -ésimo valor singular σ_i . Siendo v_i la columna i de V , definimos para $i = 1, \dots, n$ la *transformación característica* del dígito x_i como el vector $\mathbf{tc}(x_i) = (v_1^t x_i, v_2^t x_i, \dots, v_k^t x_i) \in \mathbb{R}^k$, donde $k \in \{1, \dots, m\}$ es un parámetro de la implementación. Este proceso corresponde a extraer las k primeras *componentes principales* de cada imagen. La intención es que $\mathbf{tc}(x_i)$ resuma la información más relevante de la imagen, descartando los detalles o las zonas que no aportan rasgos distintivos.

Dada una nueva imagen x de una cara, que no se encuentra en el conjunto inicial de imágenes de entrenamiento, el problema de reconocimiento consiste en determinar a qué persona de la base de datos corresponde. Para esto, se calcula $\mathbf{tc}(x)$ y se compara con $\mathbf{tc}(x_i)$, para $i = 1, \dots, n$, utilizando un criterio de selección adecuado.

Enunciado

Se pide implementar un programa en C o C++ que lea desde archivos las imágenes de entrenamiento correspondientes a distintas personas y que, utilizando la descomposición en valores singulares y el número de componentes principales k mencionado anteriormente, calcule la transformación característica de acuerdo con la descripción anterior. Se debe proponer

e implementar al menos un método que, dada una nueva imagen de una cara, determine a que persona de la base de datos corresponde utilizando la transformación característica.

Con el objetivo de obtener la descomposición en valores singulares, se deberá implementar el método de la potencia con deflación para la estimación de autovalores/autovectores. En este contexto, la factibilidad de aplicar este método es particularmente sensible al tamaño de las imágenes de la base de datos. Por ejemplo, considerar imágenes en escala de grises de 100×100 píxeles implicaría trabajar con matrices de tamaño 10000×10000 . Una alternativa es reducir el tamaño de las imágenes, por ejemplo, mediante un submuestreo. Sin embargo, es posible superar esta dificultad en los casos donde el número de muestras es menor que el número de variables. Se pide desarrollar las siguientes sugerencias y fundamentar como utilizarlas en el contexto del trabajo.

- Dada una matriz y su descomposición en valores singulares $A = U\Sigma V^t$, encontrar la descomposición en valores singulares de A^t . Cómo se relacionan los valores singulares de A y A^t ?
- Dada la descomposición en valores singulares de A , expresar en función de U , Σ y V las matrices A^t , $A^t A$ y AA^t . Analizar el tamaño de cada una de ellas y deducir como relacionar las respectivas componentes principales. Combinar con el item anterior para el cómputo de los componentes principales.

En base a este análisis, se pide desarrollar una herramienta alternativa que permita trabajar bajo ciertas condiciones con imágenes de tamaño mediano/grande.

Junto con este enunciado se provee una base de datos de imágenes correspondiente a 41 personas, con 10 imágenes por cada una de ellas. Esta base de datos se encuentra disponible en dos resoluciones distintas: 92×112 y 23×28 píxeles por cada imagen. La segunda corresponde a un submuestreo de la base original. En relación a la experimentación, se pide como mínimo realizar los siguientes experimentos:

- Analizar para cada una de las variantes qué versión de la base de datos es posible utilizar, en base a requerimientos de memoria y tiempo de cómputo.
- Para cada una de las variantes propuestas, analizar el impacto en la tasa de efectividad del algoritmo de reconocimiento al variar la cantidad de componentes principales considerados. Estudiar también como impacta la cantidad de imágenes consideradas para cada persona en la etapa de entrenamiento.
- En caso de considerar más de una posibilidad para determinar a que persona corresponde una nueva cara, considerar para cada una la mejor configuración de parámetros y compararlas entre ellas.

El objetivo final de la experimentación es proponer una configuración de parámetros/métodos que obtenga resultados un buen balance entre la tasa de efectividad de reconocimiento de caras, la factibilidad de la propuesta y el tiempo de cómputo requerido.

Programa y formato de archivos

Se deberán entregar los archivos fuentes que contengan la resolución del trabajo práctico. El ejecutable tomará tres parámetros por línea de comando, que serán el archivo de entrada, el archivo de salida, y el método a ejecutar (0 método estándar, 1 método alternativo).

Asumimos que la base de datos de imágenes se encuentra organizada de la siguiente forma: partiendo de un directorio raíz, contiene un subdirectorio por cada una de las personas donde se encuentran las imágenes de entrenamiento. El nombre de las imágenes será un número y su extensión una archivo `.pgm` (ejemplo: `1.pgm`, `2.pgm`, etc.). De esta forma se puede saber a que persona corresponde cada imagen simulando un etiquetado de las mismas.

Para facilitar la experimentación, el archivo de entrada con la descripción del experimento sigue la siguiente estructura:

- La primera línea contendrá el *path* al directorio que contiene la base de datos, seguido de 5 números enteros que representan la cantidad de filas y columnas de las imágenes de la base, cuantas personas (p) y cuantas imágenes por cada una de ellas ($nimgp$), y cuántas componentes principales se utilizarán en el experimento (k). A continuación de muestra un ejemplo de una base de datos de imágenes de 112×92 , con 41 sujetos, 5 imágenes por sujeto y tomando 15 componentes principales.

```
../data/ImagenesCaras/ 112 92 41 5 15
```

- A continuación, el archivo contendrá p líneas donde en cada una de ellas se especificará la carpeta correspondiente a la p -ésima persona, seguido de $nimgp$ numeros enteros indicando las imágenes a considerar para el entrenamiento. La siguiente línea muestra como ejemplo 5 imágenes (2, 4, 7, 8 y 10) a considerar para la persona `s10`.

```
s10/ 2 4 7 8 10
```

- Finalmente, se especificará un número n_{test} de imágenes (preferentemente no contenidas en la base de imágenes) para las cuales se desea identificar a quien pertenecen. Cada una de ellas se especificará en una nueva línea, indicando el *path* al archivo seguido del número de individuo al que pertenece (relativo a la numeración establecida en el punto anterior). La siguiente línea muestra un ejemplo de una imagen a testear para el sujeto número 1:

```
../data/ImagenesCaras/s1/9.pgm 1
```

El archivo de salida obligatorio tendrá el vector solución con los k valores singulares de mayor magnitud, con una componente del mismo por línea. Adicionalmente, en caso de considerarlo conveniente, se podrán agregar archivos adicionales opcionales con más información que la descripta en la presente sección.

Junto con el presente enunciado, se adjunta una serie de scripts hechos en `python` y un conjunto instancias de test que deberán ser utilizados para la compilación y un testeo básico

de la implementación. Se recomienda leer el archivo `README.txt` con el detalle sobre su utilización.

Fecha de entrega

- *Formato electrónico:* Jueves 15 de Mayo de 2014, hasta las 23:59 hs., enviando el trabajo (informe+código) a `metnum.lab@gmail.com`. El subject del email debe comenzar con el texto [TP2] seguido de la lista de apellidos de los integrantes del grupo.
- *Formato físico:* Viernes 16 de Mayo de 2014, de 17:30 a 18:00 hs.