

Reconocimiento Automático de Partituras (Junio 2010)

Iván López Espejo, Jonathan Prados Garzón

El objetivo del presente trabajo es el de proporcionar un reconocimiento automatizado de partituras muy sencillas no manuscritas presentadas en algún formato de imagen.

I. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

EN ESTE principal apartado del texto desarrollaremos la solución propuesta al problema de reconocimiento óptico musical. Las diferentes etapas del algoritmo son las descritas a continuación.

A. Preprocesamiento

En este primer punto adaptamos la partitura para su tratamiento. Una vez obtenemos la imagen de la misma, supuesta su entrada en el modelo de color RGB, la transformamos a niveles de gris para seguidamente invertirla. La misión de la inversión es la de un tratamiento más sencillo mediante la contemplación de la información contenida en la partitura como niveles de gris relativamente grandes frente al fondo tendente a nivel cero. Finalmente, aplicamos binarización para que los píxeles de interés se mantengan en blanco frente al fondo negro.

B. Corrección de la inclinación

Esta etapa está incluida en el preprocesado de la imagen y, como su nombre indica, consiste en corregir la inclinación que pueda presentar la imagen de partida. Este paso es de vital importancia dado que los métodos de segmentación y reconocimiento desarrollados no operarían correctamente ante la presencia de este defecto.

El método seguido para llevar a cabo la corrección de la inclinación se basa principalmente en la transformada de Hough. En primer lugar se aplica dicha transformada con una resolución en la componente Θ de 0.05° . El espacio de parámetros de salida usado es el polar. La ventaja que aporta (aparte de los problemas que nos evitamos con las líneas verticales) es que obtenemos directamente una aproximación del ángulo de inclinación de la imagen. El porqué de optar por el uso de la transformada de Hough es por las características inherentes de los datos de entrada que vamos a emplear. Dado que estamos ante la imagen de una partitura musical que presenta numerosos pentagramas, los máximos de la transformada de Hough se corresponderán con las líneas que componen dichos pentagramas. En consecuencia, uno de los máximos de dicha transformada nos proporcionará la ecuación de la recta en coordenadas polares asociada a una de las líneas de los pentagramas y, a partir de la misma, podremos obtener directamente una estimación del ángulo de inclinación. Posteriormente, por si acaso la imagen se presenta aún algo torcida, se opta por un ajuste de tanteo. Este consiste en aplicar inclinaciones a la imagen en los márgenes de error que tenemos tras efectuar el paso de la transformada de Hough. La realimentación para determinar si las correcciones que

estamos realizando están surtiendo efecto o no (y en definitiva para obtener convergencia hacia la solución), se basan en el estudio de la proyección horizontal de histograma de la imagen resultado.

C. Segmentación de pentagramas

A continuación segmentamos la imagen separando pentagramas y título de la partitura. Esto lo logramos con ayuda de la proyección horizontal de histograma. Dado que el fondo es completamente negro, las zonas inter-pentagrama se caracterizarán en la proyección horizontal de histograma por una secuencia continua de ceros, por lo que, mediante un algoritmo de búsqueda de ceros basado en comparación con el valor anterior y posterior, establecemos el punto medio entre dos pentagramas para que estos sean separados.

D. Segmentación de símbolos

Ahora, para cada uno de los pentagramas, así como para el título, se procede a segmentar cada uno de los símbolos que componen la línea con la ayuda de la proyección vertical de histograma de cada una de las líneas por separado. En el caso de un pentagrama, en su proyección vertical de histograma se podrá observar un nivel mínimo de continua correspondiente a la contribución de las cinco líneas del pentagrama. Este nivel, que conocemos como el mínimo no-cero, es detectado para cada uno de los pentagramas individualmente, de modo que, con un algoritmo similar al anteriormente presentado para la segmentación de la proyección horizontal de histograma, segmentamos los símbolos de cada una de las líneas detectadas (con la diferencia del uso de un algoritmo de búsqueda del mínimo no-cero comentado). Estos límites intersimbólicos podrían aprovecharse para la segmentación directa del elemento en la imagen. No obstante, se prefirió realizar una caracterización del símbolo mediante el estudio del contorno de la proyección vertical de histograma del mismo, ya que podemos limitar su extensión mucho más fácilmente respecto de la extensión del símbolo en el interior del pentagrama (comentar en este sentido que la caracterización del símbolo mediante el estudio de seis momentos estadísticos arrojó peores resultados respecto del método finalmente implementado).

E. Reconocimiento de símbolos

A continuación se procede a la clasificación de cada uno de los símbolos detectados según el siguiente método. Ya hemos dicho que realmente estamos segmentando la proyección de histograma vertical del símbolo. Pues bien, con esta información hacemos lo siguiente:

- Eliminación de la componente de continua detectada como el mínimo no-cero anteriormente comentado.
- Normalización de los valores de la proyección vertical de histograma del símbolo.
- Obtención de 24 coeficientes del módulo de la FFT de la proyección vertical de histograma del símbolo.
- Comparación con cada uno de los patrones base almacenados por clase en la base de datos. La selección de la clase a la cual pertenece el símbolo se realiza en base a la minimización del error lineal calculado como la acumulación de las diferencias en valor absoluto entre cada par de coeficientes de los módulos de las FFTs del símbolo en estudio y del patrón de la base de datos con el cual se compara.

El estudio del símbolo en base a la FFT se basa, junto con la normalización, en la importancia de caracterizar con un número estándar de coeficientes el símbolo, de modo que se contemple la propiedad de la invarianza a la escala.

La función implementada en el prototipo asume que el símbolo que se estudia se va a encontrar en la base de datos, por lo que se selecciona mediante minimización del error y no se establece ningún umbral necesario tipo tolerancia para poder clasificarse o no.

Otra limitación importante es que la función también presupone que el título sólo ocupa la primera línea segmentada mediante la proyección horizontal de histograma. Esto se ha pensado así debido a la forma de segmentar un símbolo dentro de un pentagrama. Recordemos que el símbolo se lleva sobre el nivel cero (eliminación de la componente de continua introducida por las cinco líneas del pentagrama). Pues bien, el programa hace una excepción sobre el cálculo del nivel mínimo de continua para la primera línea ya que, si se selecciona un mínimo no-cero en la línea del título, podemos perder y distorsionar la información no proveyendo un resultado coherente.

Comentar que sobre los cinco sencillos ejemplos provistos se obtienen unos buenos resultados con este método de clasificación.

Símbolos detectables		
A	B	C
D	E	F
G	H	I
L	M	N
Ñ	O	P
R	S	U
Z	Clave de sol	Compas de 3/4
Compás de 4/4	Silencio de blanca	Negra
Blanca	Redonda	Barra divisoria y doble barra final

Tab. 1. Símbolos con patrón correspondiente en la base de datos.

Los símbolos contemplados en la base de datos son los recogidos en la tabla 1. Notar que, para mayor precisión en la selección, por cada símbolo se incluyen varias

caracterizaciones en función de si este puede aparecer reflejado, invertido (piénsese en una figura como una blanca), etc.

El reconocimiento de más símbolos aparte de los incluidos en la base de datos sería inmediatamente extensible, pudiendo cubrir con este método partituras de mayor complejidad si fuese necesario.

Tampoco se ha tenido en cuenta el reconocimiento de armaduras, por simplicidad (sólo se admiten partituras cuya tonalidad sea de *Do Mayor* o su relativo menor *La menor*), ni alteraciones accidentales. Por tanto, las partituras de test están todas transportadas a la tonalidad de *Do Mayor*.

F. Detección de tono

Una vez tenemos segmentada la imagen y se han reconocido las figuras musicales, se procede a la detección de tono donde se requiera. El método seguido para la detección de tono está basado en el uso de la proyección horizontal de histograma. Dado el trozo de la imagen, hallamos su proyección horizontal (sobre el eje *y*). A continuación se sigue un método en el que se determina tanto la posición de las líneas del pentagrama como las líneas virtuales (se pueden definir 13 líneas desde el *Sol3* al *Re7*, 5 visibles y 8 virtuales). Lo que hacemos es recorrer la proyección horizontal hasta determinar las líneas del pentagrama. Como estas estarán representadas en blanco y ocuparán todo el trozo de la imagen, se corresponderán con máximos. Así determinamos sus posiciones y las distancias de separación entre cada par de líneas. Haciendo un promedio de la distancia entre pares de líneas adyacentes visibles, podemos obtener una estimación de la separación (aplicando redondeo por estar en un espacio discreto) entre líneas del pentagrama. Con esta distancia estimada podemos determinar las posiciones de las 8 líneas virtuales restantes.

Una vez determinadas todas las posiciones, vamos recorriendo la proyección de histograma situándonos en las líneas estimadas y en los puntos medios entre ambas. En cada posición haremos la suma de todos los elementos del histograma en un entorno igual a la mitad de la separación calculada.

Finalmente, en aquella posición (de las 26 posibles, líneas y puntos medios entre dos líneas) donde se obtenga una suma mayor, se considera que se encuentra la cabeza de la figura musical.

II. NOTAS IMPORTANTES SOBRE LA EJECUCIÓN

Posterior a la ejecución del fichero *demo.m*, el resultado del reconocimiento de la partitura se almacena en un archivo en el mismo directorio y llamado *Resultado.txt*. En él figura la secuencia de caracteres reconocidos y, al lado de cada figura musical correspondiente a una nota, la frecuencia correspondiente al tono detectado.

Mediante la siguiente expresión se puede asociar la frecuencia a la nota musical para corroboración inmediata:

$$f = 440e^{((\log_2(\frac{f}{32.7})+1)-4)+(\frac{n-10}{12}))\log(2)} \Rightarrow$$

$$\begin{aligned} \Rightarrow \log\left(\frac{f}{440}\right) &= \left(\left(\left\lfloor \log_2\left(\frac{f}{32.7}\right) + 1 \right\rfloor - 4\right) + \left(\frac{n-10}{12}\right)\right) \log(2) \Rightarrow \\ \Rightarrow \log\left(\frac{f}{440}\right) &= \log(2) \left\lfloor \log_2\left(\frac{f}{32.7}\right) + 1 \right\rfloor - 4 \log(2) + \left(\frac{n-10}{12}\right) \log(2) \Rightarrow \\ \Rightarrow n &= 10 + \frac{12}{\log(2)} \left[\log\left(\frac{f}{440}\right) - \log(2) \left[\left\lfloor \log_2\left(\frac{f}{32.7}\right) + 1 \right\rfloor - 4 \right] \right]. \end{aligned}$$

n	Nota
1	Do
2	Do#
3	Re
4	Re#
5	Mi
6	Fa
7	Fa#
8	Sol
9	Sol#
10	La
11	La#
12	Si

Tab. 2. Correspondencia entre el número n y la nota musical.

La anterior tabla relaciona el número natural n obtenido de la aplicación de la anterior expresión en función de la frecuencia con la nota en cuestión. Por otro lado, la octava en la que dicha nota se encuentra también se puede obtener a partir de

$$o = \left\lfloor \log_2\left(\frac{f}{32.7}\right) + 1 \right\rfloor.$$

REFERENCIAS

- [1] X. Fernández Hermida y C. Sánchez-Barbudo y Vargas, "Development of an Optical Music Recognizer".
- [2] A. Sánchez, J.J. Pantrigo y J. Ignacio Pérez, "Extracción de Líneas Melódicas a partir de Imágenes de Partituras Musicales".