

RESEARCH

Open Access

# An optical music recognition system for traditional Chinese *Kunqu* Opera scores written in *Gong-Che* Notation

Gen-Fang Chen<sup>1\*</sup> and Jia-Shing Sheu<sup>2</sup>

## Abstract

This paper presents an optical music recognition (OMR) system to process the handwritten musical scores of *Kunqu* Opera written in *Gong-Che* Notation (GCN). First, it introduces the background of *Kunqu* Opera and GCN. *Kunqu* Opera is one of the oldest forms of musical activity, spanning the sixteenth to eighteenth centuries, and GCN has been the most popular notation for recording musical works in China since the seventh century. Many *Kunqu* Operas that use GCN are available as original manuscripts or photocopies, and transforming these versions into a machine-readable format is a pressing need. The OMR system comprises six stages: image pre-processing, segmentation, feature extraction, symbol recognition, musical semantics, and musical instrument digital interface (MIDI) representation. This paper focuses on the symbol recognition stage and obtains the musical information with Bayesian, genetic algorithm, and *K*-nearest neighbor classifiers. The experimental results indicate that symbol recognition for *Kunqu* Opera's handwritten musical scores is effective. This work will help to preserve and popularize Chinese cultural heritage and to store *Kunqu* Opera scores in a machine-readable format, thereby ensuring the possibility of spreading and performing original *Kunqu* Opera musical scores.

**Keywords:** Optical music recognition; *Kunqu* Opera; *Gong-Che* Notation; MIDI; Classifiers

## 1. Introduction

*Kunqu* Opera is one of the oldest forms of opera performed in China, dating back to the end of the *Yuan* Dynasty (1271 to 1368 AD) approximately 700 years ago. It dominated Chinese theatre from the sixteenth to the eighteenth centuries (see Figure 1) and is distinct in the virtuosity of its rhythmic patterns. It has been a dominant influence on more recent forms of local opera in China and is considered the mother of all Chinese operas, such as the *Beijing* opera. In 2001, United Nations Educational, Cultural and Scientific Organization (UNESCO) proclaimed *Kunqu* Opera a Masterpiece of the Oral and Intangible Heritage of Humanity [1].

Its composers wrote the musical scores of *Kunqu* Opera in Chinese using seven or ten characters, that is, in *Gong-Che* Notation (GCN) [2]. Several *Kunqu* Opera scores exist in the ancient literature [3]. In particular,

there are three famous script sets that were popular during the *Qing* Dynasty (1636 to 1911 AD): *Jiugong Dacheng Nanbei Ci Gongpu* (A comprehensive anthology of texts and notation of the Southern and Northern opera tunes in nine modes, 1746 AD, with 4,466 musical works) [4], *Nashuying Qupu* (*Nashu* Studio Theatrical Music, 1792 AD, with more than 360 drama scripts) [5], and *Nanci Dinglv* (The Law of the Southern Word, 1720 AD, with 1,342 musical works) [6].

The risk of losing this rich cultural heritage must be urgently addressed, as *Kunqu* Opera is facing competition from mass culture, as well as a lack of audience interest since the eighteenth century because of the high level of technical knowledge required. Of the 400 arias regularly sung in opera performances in the mid-twentieth century, only a few dozen are performed now, such as *Mudan Ting* (The Peony Pavilion). Many *Kunqu* Opera works are available only as original manuscripts or photocopies, and their digitization and transformation into a machine-readable format is a pressing need.

\* Correspondence: chengenfang@gmail.com

<sup>1</sup>Department of Computer Science and Technology, Hangzhou Normal University, Hangzhou, Zhejiang 310036, China

Full list of author information is available at the end of the article



**Figure 1 Photo of the Kunqu 'Hall of Longevity'.** Playwright: Sheng Hong (1645 to 1704); actor and actress from Shanghai Kunqu Opera Troupe: Zheng-Ren Cai and Jing-Xian Zhang; performance venue: Hangzhou Theatre, China; photographer: Gen-Fang Chen; date taken: October 11, 2011.

Optical music recognition (OMR) refers to the automatic processing and analysis of images of musical notation. This process typically employs a scanner or other digital device to transform paper-based musical scores into a digital image, which is then processed, recognized, and automatically translated into a standard format for music files, such as the musical instrument digital interface (MIDI) format. OMR represents a comprehensive amalgamation of fields and methods, including musicology, artificial intelligence, image engineering, pattern recognition, and MIDI technology. OMR technology provides a way to convert paper-based scores and has numerous applications, including computer-assisted music teaching, digital music libraries, musical statistics, digital music image automatic classification, and synchronous music and audio communication.

The first published OMR work was conducted by Pruslin [7] at the Massachusetts Institute of Technology. Pruslin's system recognized a subset of Western Music Notation (WMN), primarily musical notes. Before the 1990s, many studies were performed [8,9], and research had begun to focus on handwritten formats [9-11] and ancient (or folk) music, including medieval music [12], white mensural notation [13,14], early music prints [15], Orthodox Hellenic Byzantine Music notation [16], and Greek traditional music [17]. During the same period, several commercial OMR software packages, such as Capella-scan, Optical Music easy Reader, Photo Score, Sharp Eye, Smart Score, and Vivaldi Scan, were developed. These reached the market for WMN, and their recognition rate exceeded 90% for commonly printed musical scores [18].

Some techniques and methods used in OMR technology research include projection [19], mathematical morphology

[20], neural networks [21], fuzzy theory [22], genetic algorithms [23], high-level domain knowledge [24], graph grammar [25], and probability theory [26]. Most of these methods are intended for WMN scores and can convert paper-based scores to digital form (such as a MIDI sequence). Replacing the scores' manual input mode in this way will improve the rate of musical score digitization.

In this study, an OMR system for *Kunqu* Opera musical scores is presented. The remainder of this paper is organized as follows: Section 2 describes *Kunqu* Opera musical scores, Section 3 discusses the structure of the OMR system and some preliminary operations, Section 4 provides the experimental results for selected scores using the OMR system, and Section 5 offers ideas for future research.

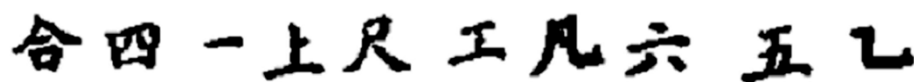
## 2. Gong-Che Notation in *Kunqu* Opera musical scores

GCN is a type of musical notation that was developed and cultivated in East Asia, flourishing mainly in ancient China. GCN was invented in the *Tang* Dynasty (618 to 907 AD) [27] and became a popular form of music notation in China during the *Song* (960 to 1279 AD), *Ming* (1368 to 1644 AD), and *Qing* (1636 to 1911 AD) Dynasties. Spanning 1,000 years of use, it is a widely accepted musical notation in East Asia and resembles WMN in Europe. Many traditional Chinese musical manuscripts have been written in GCN [28], including the *Kunqu* Opera scripts. GCN represents musical notes using Chinese characters, with the basic GCN note pitch symbols represented by ten Chinese characters, as shown in Figure 2. The meaning of these characters is presented in Table 1.

A *Kunqu* Opera GCN score contains contents such as the title, key signature, *qupai* (*qupai* is the general term for the tune names used in traditional Chinese lyrics and music writing, and every *qupai* singing tone has its own musical form, tone, and tonality), lyrics, and notes. An example of a GCN score is provided in Figure 3. Words constitute a primary component of a GCN score. GCN is scripted in the format of traditionally written Chinese: from top to bottom and from right to left. Rhythm marks are indicated to the right of the note characters.

The title of a *Kunqu* Opera script is located in the first column, and the *qupai* is located in the second column, with the key signature in the top row. The lyrics with notes appear below the *qupai*, and the notes with the rhythm marks are to the upper right of the lyrics. A frame with four borders usually encircles all the columns. Figure 4 depicts this layout for a GCN score.

Commonly, a Chinese lyric word includes several GCN notes, as shown in Figure 5. The number of pitches in a Chinese lyric word ranges from 1 to 8, and



**Figure 2** Ten GCN symbols of a pitch.

the rhythm mark number of a pitch is from 0 to 2. Thus, words, pitches, and rhythms form a complex three-dimensional (3D) spatial relationship in a GCN score, with the first dimension containing lyrics, the second dimension containing pitches, and the third dimension containing rhythm marks. The font size of a Chinese lyric word is larger than the font size of a pitch, whose font size is in turn larger than that of a rhythm mark.

GCN does not mark a note's duration but instead gives the rhythm marks at regular intervals, with each rhythm mark being indicated in the corresponding note's upper right corner. The basic GCN note rhythm mark comprises eight symbols, as shown in Table 2, and two main types of rhythms are distinguishable: 'ban', the stronger beat, and 'yan', the weaker beat.

The difficulties encountered in the processing of GCN musical scores are due to the following: (1) the complex 3D relationship between symbols in a score's image, (2) symbols being written in different sizes, shapes, and intensities, (3) the variation in relative size between different components of a musical symbol, (4) different symbols appearing connected to each other and the same musical symbol appearing in separated components, and (5) difficulties in converting the musical information in GCN scores to other styles of musical notation, such as WMN.

### 3. Description of the OMR system for Kunqu Opera musical scores

The OMR system for Kunqu Opera (KOMR), shown in Figure 6, is an off-line optical recognition system

**Table 1** Description of a GCN score's general symbols (pitch)

Symbols	Description		
	Phoneticize in simplified Chinese	Movable doh	Solfège
合	hé	5	sol
四	sì	6	la
一	yī	7	ti
上	shàng	1	do
尺	chǐ	2	re
工	gōng	3	mi
凡	fán	4	fa
六	liù	5	sol
五	wǔ	6	la
乙	yǐ	7	ti

comprising six independent stages: (1) image pre-processing, (2) document image segmentation, (3) feature extraction, (4) musical symbol recognition, (5) musical semantics, and (6) MIDI representation.

Because of the maturity of image digitizing technology [29], paper-based Kunqu Opera musical scores can easily be converted into digital images. Therefore, this paper does not discuss the image acquisition stage. The system processes a gray-level bitmap from a scanner or reads directly from image files, with the input bitmap generally being 300 dpi with 256 gray levels.

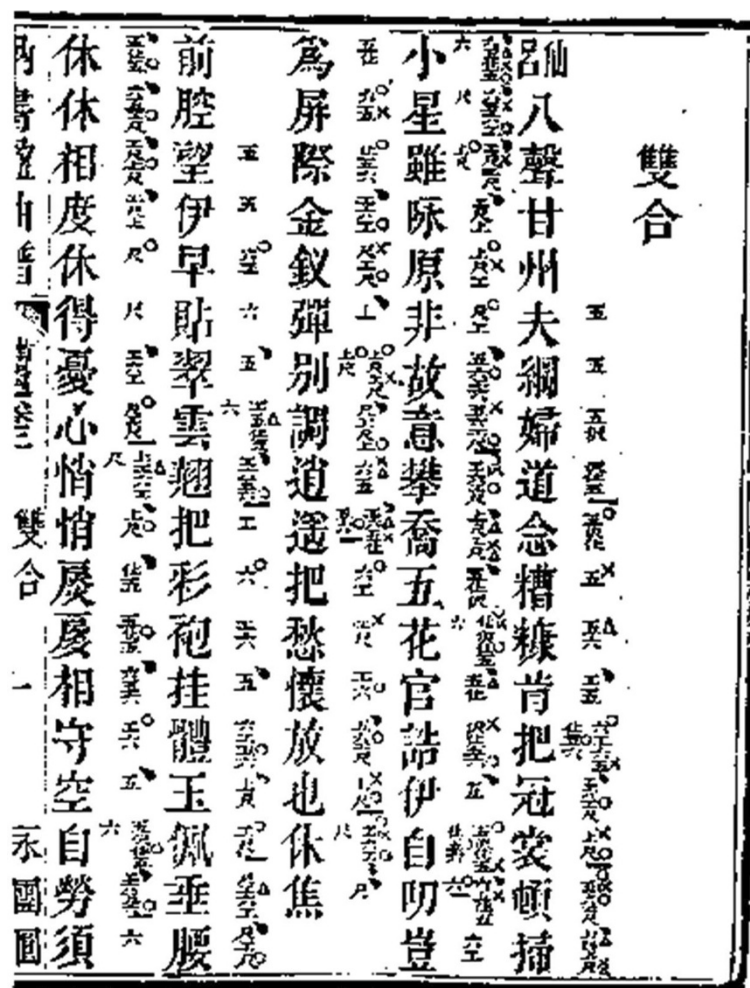
#### 3.1 Image pre-processing

Pre-processing might involve any of the standard image-processing operations, including noise removal, blurring, de-skewing, contrast adjustment, sharpening, binarization, or morphology. Many operations may be necessary to prepare a raw input image for recognition, such as the selection of an interesting area, the elimination of nonmusical elements, image binarization, and the correction of image rotation.

Many documents, particularly historical documents such as those in Kunqu Opera GCN scores, rely on careful pre-processing to ensure good overall system performance and, in some cases, to significantly improve the recognition performance. In this work, we briefly touch upon the basic pre-processing operations of Kunqu Opera GCN scores, such as binarization, noise removal, skew correction, and the selection of an area of interest. Image binarization uses the algorithm of Otsu [30], noise removal is conducted using basic morphological operations [31], and image rotation has been corrected using the least squares criterion [32]. The area of interest in a Kunqu score is surrounded by a frame (wide lines) (see Figure 3). The frame is a connected component containing the longest line in the score, so we use the Hough transform [33] to locate the longest line then delete the connected component with the longest line from the score, leaving the area of interest. Because these are common image-processing operations, we do not give a detailed description here.

#### 3.2 Document segmentation and analysis of GCN scores

Document segmentation is a key phase in the KOMR system. With the symbols in a GCN score document having been classified, this stage first segments the document into two sections, one including symbols for the



**Figure 3** Musical score sheet of Nashu Studio Theatrical Music (p. 1983).

notes and the other for the non-notes. Elements of non-note symbols, such as the title, key signature, *qupai*, lyrics, noise, and border lines of a textual framework, are then identified and removed.

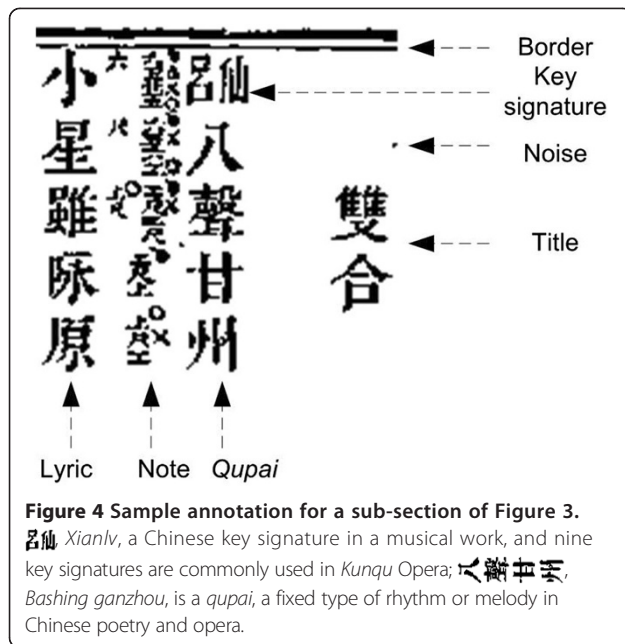
Because music is a time-based art, the arrangement of the notes is one of its most important factors. Therefore, obtaining the arrangement of the notes in a GCN score is requisite to document the segmentation. Concordant with the writing style of the GCN score, the arrangement of the notes can be organized based on high-level field knowledge of *Kunqu* Opera scores.

Several document image segmentation methods have been proposed, with the best known being *X-Y* projection [34], run length smoothing algorithm (RLSA) [35], component grouping [36], scale-space analysis [37], Voronoi tessellation [38], and the Hough transform [39]. These methods are suitable for handwritten document segmentation, such as for GCN scores, which are compatible with

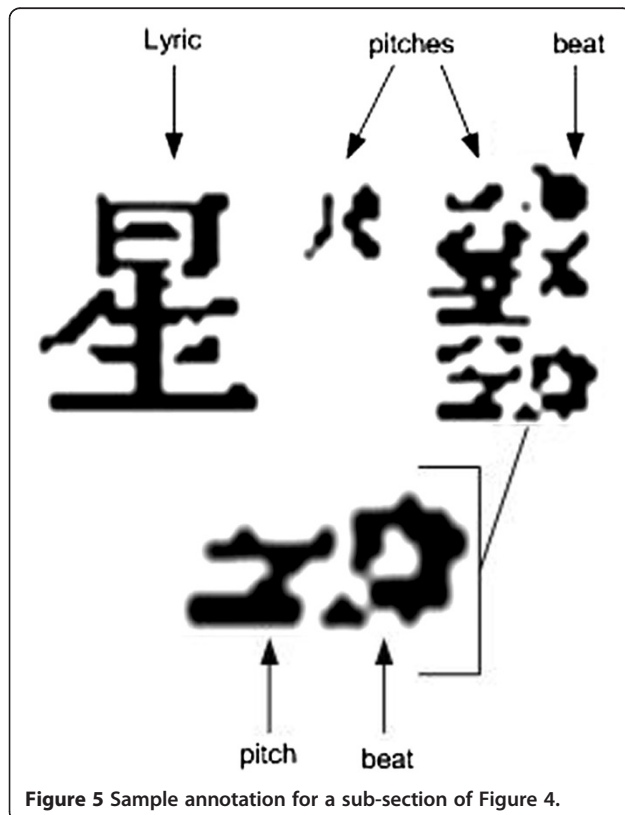
*X-Y* projection, RLSA, scale-space analysis, and the Hough transform [40].

A preliminary result of GCN score segmentation has been presented in [41]. A self-adaptive RLSA was used to segment the image according to an *X*-axis function (denoted by  $PF(x)$ ) indicating the number of foreground pixels in each column of the image. This *X*-axis function uses *X*-projection to compute the number of flex points that satisfy the following conditions:  $PF(x-1) < PF(x)$  and  $PF(x) > PF(x+1)$  or  $PF(x-1) > PF(x)$  and  $PF(x) < PF(x+1)$ . Next, the algorithm iteratively smoothes the function and analyzes the next smoothed function to ensure that the number of flex points in both functions is equal. Finally, the image is segmented into several sub-images based on the *X*-axis values for flex points in the function. To extract notes from the image, all connected components are identified using a conventional connected component labeling algorithm, and the minimum bounding box of all connected components is computed.





Next, the algorithm matches each connected component to its corresponding sub-image. According to the experimental results in [41], the rate of correct segmentation in lines is 98.9%, and the loss rate of notes is 2.7%; however, the total error rate of all notes and lyrics is almost 22%.



**Table 2** Description of GCN score's general symbols (rhythm)

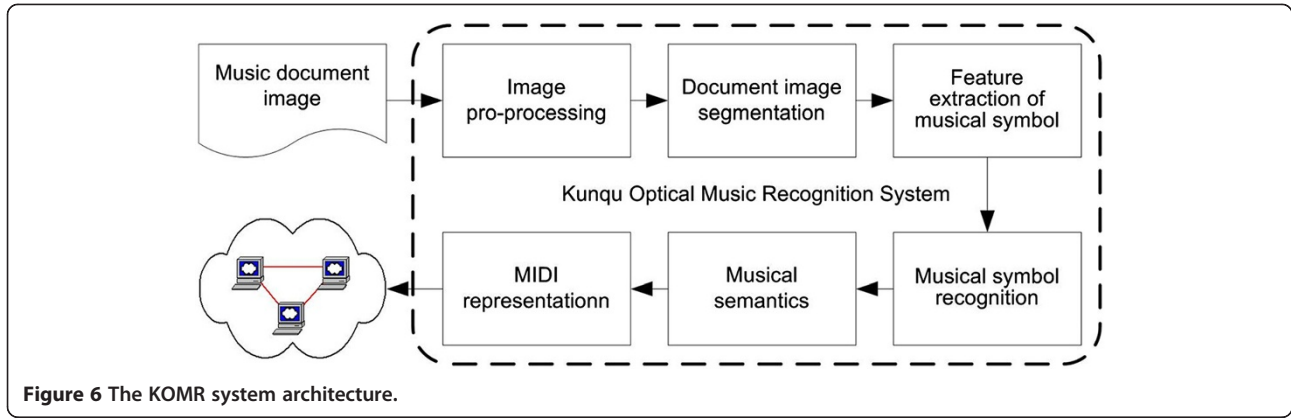
Symbols	Phoneticize in simplified Chinese	Description
↘	<i>ban</i>	A note with 'ban' starts a stronger beat
○	<i>yan</i>	A note with 'yan' starts a weaker beat
△	<i>yaoyan</i>	A note with 'yaoyan' starts a weaker beat
—	<i>yaoban</i>	Denotes a note's duration from a weaker beat to the next stronger beat
×	<i>ban</i>	A note with 'ban' starts a stronger beat
⊗	<i>zhengban</i>	Denotes the stronger beat in the middle of a bar for an 8/8 beat
L	<i>yaoban</i>	Denotes a note's duration from a weaker beat to the next stronger beat in some scores
□	<i>yaoyan</i>	A note with 'yaoyan' starts a weaker beat in some scores

### 3.3 Symbol feature extraction

Selecting suitable features for pattern classes is a critical process in the OMR system. Feature representation is also a crucial step, because perfect feature data can effectively enhance the symbol recognition rate. The goal of feature extraction is to characterize a symbol to be recognized by measurements whose values are highly similar for symbols in the same category, but different for symbols in different categories. The feature data must also be invariant during relevant transformations, such as translation, rotation, and scaling.

Popular feature extraction methods for OCR and Chinese character recognition include peripheral feature [42], cellular feature [43], and so on [44]. Because symbols are written in different sizes in a GCN score, to construct a simple and intuitive approach, four types of structural features have been used in this exploratory work based on the reference [42,43]; these are suited to feature extraction of symbols in a GCN score and to compare with the recognition rate of KOMR. An  $n \times m$  matrix featuring symbol data is used to obtain symbols of the same size as in the feature matrix. These are obtained using a grid to segment the symbols [45]. In Figure 7, a sample symbol with size  $H \times W$  is shown in sub-graph (a), and an  $n \times m$  grid is shown in sub-graph (b). In this example,  $h_0 = 0, h_n = H, h_i = \lceil \frac{H}{n} \rceil \times i, w_0 = 0, w_m = W, w_j = \lceil \frac{W}{m} \rceil \times j$ , and the four features of each symbol are given by the following equations:

$$T_1 = \begin{bmatrix} \dots, \dots, \dots \\ \dots, t_{i,j}, \dots \\ \dots, \dots, \dots \end{bmatrix}, t_{i,j} = \sum_{x=w_{i-1}}^{w_i} \sum_{y=h_{j-1}}^{h_j} f(x, y), 1 \leq i \leq m, 1 \leq j \leq n \quad (1)$$



$$T_2 = \begin{bmatrix} \dots, \dots, \dots \\ \dots, t_{ij}, \dots \\ \dots, \dots, \dots \end{bmatrix}, t_{ij} = \sum_{x=w_{i-1}}^{w_i} f(x, h_j), 1 \leq i \leq m, 1 \leq j \leq n \quad (2)$$

$$T_3 = \begin{bmatrix} \dots, \dots, \dots \\ \dots, t_{ij}, \dots \\ \dots, \dots, \dots \end{bmatrix}, t_{ij} = \sum_{y=h_{j-1}}^{h_j} f(w_i, y), 1 \leq i \leq m, 1 \leq j \leq n \quad (3)$$

$$T_4 = \begin{bmatrix} \dots, \dots, \dots \\ \dots, t_{ij}, \dots \\ \dots, \dots, \dots \end{bmatrix}, t_{ij} = \begin{cases} 0, & \text{for } 1 < i < m \text{ and } 1 < j < n \\ x, & \text{if } \sum_{k=0}^x f(k, h_j) = 0 \text{ and } f(x+1, h_j) = 1, \text{ for } j = 1 \\ W-x, & \text{if } \sum_{k=x}^{w_m} f(k, h_j) = 0 \text{ and } f(x-1, h_j) = 1, \text{ for } j = m \\ x, & \text{if } \sum_{k=0}^x f(w_i, k) = 0 \text{ and } f(w_i, x+1) = 1, \text{ for } i = 1 \\ H-x, & \text{if } \sum_{k=x}^{h_n} f(w_i, k) = 0 \text{ and } f(w_i, x-1) = 1, \text{ for } i = n \end{cases} \quad (4)$$

where  $f(x, y)$  is the value of a pixel at coordinates  $(x, y)$  in an image. If  $f(x, y) = 1$ , then the pixel is a foreground pixel; otherwise, 0 indicates a background pixel. The  $T_1$  feature is the number of pixels in each cell of the grid as a cellular feature,  $T_2$  feature appears to be the number of pixels falling on the upper edge of each grid cell,  $T_3$  is the number of pixels falling on the right edge of each grid cell, and  $T_4$  is the number to be some measure of how many background pixels there are from the edge to the first foreground pixel, if  $f(1,1) = 1$ , then  $t_{1,1} = 0$ ; if  $f(n,1) = 1$ , then  $t_{n,1} = 0$ ; if  $f(1,m) = 1$ , then  $t_{1,m} = 0$ ; if  $f(n,m) = 1$ , then  $t_{n,m} = 0$ ; and if there is no foreground pixel in the image, then  $t_{ij} = 0$ , for  $1 \leq i \leq m$  and  $1 \leq j \leq n$ .

### 3.4 Musical symbol recognition

Computer recognition of handwritten characters has been intensely researched for many years. Optical character recognition (OCR) is an active field, particularly for handwritten documents in such languages as Roman [45], Arabic [46], Chinese [47], and Indian [48].

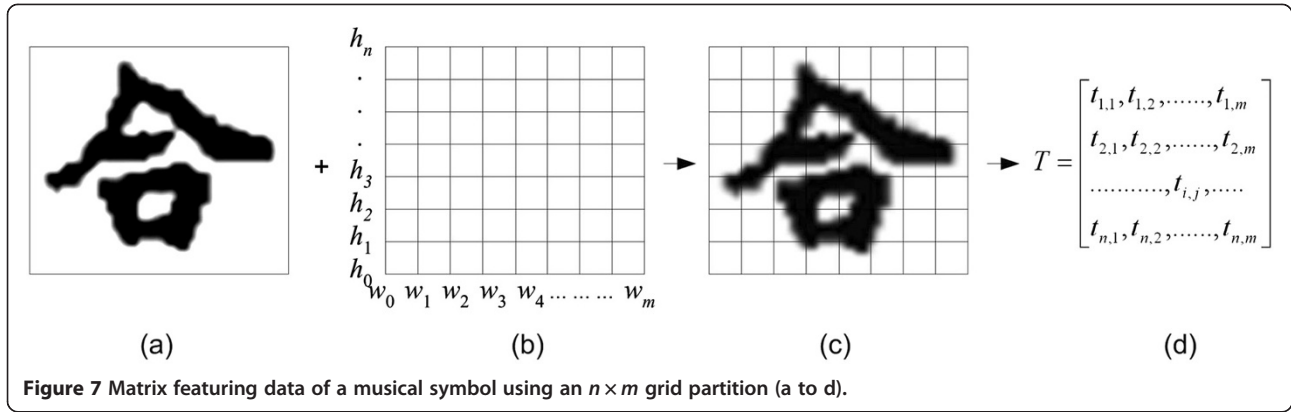
Several Chinese character recognition methods have been proposed, with the best known being the transformation invariant matching algorithm [49], adaptive confidence transform based classifier combination [50], probabilistic neural networks [51], radical decomposition [52], statistical character structure modeling [53], Markov random fields [54], and affine sparse matrix factorization [55].

The basic symbols for musical pitch (see Figure 2) in a *Kunqu* Opera GCN score are Chinese characters, but other musical pitch symbols and all rhythm symbols are specialized symbols for GCN. Thus, the methods of *Kunqu* Opera GCN score recognition refer to the techniques of both OMR and OCR. In this paper, the following three approaches to musical symbol recognition are compared.

#### 3.4.1 K-nearest neighbor

The  $K$ -nearest neighbor (KNN) classifier is one of the simplest machine learning algorithms. It classifies foregrounds based on the closest training examples in the feature space and is a type of instance-based, or lazy, learning in which the function is only approximated locally, and all computation is deferred until classification [56]. The neighbors are selected from a set of foregrounds for which the correct classification is known, a set that can be considered the training set for the algorithm, though no explicit training step is required.

In the experiment described in the manuscript, we choose half of the test musical symbols as training samples; then, for the convenience of computing, we set the value of  $K$  to 1. The feature data of each class are obtained from the samples by calculating the average feature data of all samples. Although the distance function



can also be learned during the training, the corresponding distance functions for the four features above are determined by the following equations:

$$f(S, T)_1 = \frac{m \times n - \sum_{i=1}^m \sum_{j=1}^n \rho, \rho = 1 \text{ if } s_{ij} = t_{ij}, \text{ other } \rho = 0}{m \times n} \quad (5)$$

$$f(S, T)_2 = \frac{\sum_{i=1}^m \sum_{j=1}^n |s_{ij} - t_{ij}|}{m \times n} \quad (6)$$

$$f(S, T)_3 = \frac{\sum_{i=1}^m \sum_{j=1}^n |s_{ij} - t_{ij}|}{m \times n} \quad (7)$$

$$f(S, T)_4 = \frac{m \times n - \sum_{i=1}^m \sum_{j=1}^n \rho, \text{ if } \alpha < \frac{t_{ij}}{s_{ij}} < \beta, \rho = 1, \text{ other } \rho = 0}{m \times n} \quad (8)$$

where  $f(S, T)_{1-4}$  is the distance function of the corresponding feature 1-4,  $s_{ij}$  is the feature matrix of the prototype class ( $S$ ), and  $t_{ij}$  is the feature matrix of the test sample ( $T$ ). In Equation 5, if  $s_{ij} = t_{ij}$ ,  $1 \leq i \leq m$ ,  $1 \leq j \leq n$ , then  $\rho = 1$ ; otherwise,  $\rho = 0$ .  $f(S, T)$  counts the number of unequal elements in the feature matrixes ( $S$  and  $T$ ). In Equations 6 and 7,  $f(S, T)$  calculates the sum of the difference between all corresponding elements in the feature matrixes ( $S$  and  $T$ ). Equation 8 counts the number of non-approximate elements in  $S$  and  $T$ , using the parameters  $\alpha$  and  $\beta$  as experience values. In this work,  $\alpha = 0.9$  and  $\beta = 1.1$ .

### 3.4.2 Bayesian decision theory

Bayesian decision theory is used in many statistics-based methods. Classifiers based on Bayesian decision theory are

simple, probabilistic classifiers that apply Bayesian decision theory with conditional independence assumptions, providing a simple approach to discriminative classification learning [57].

In this work, the conditional probabilities  $P(T|S_k)$ ,  $1 \leq r \leq 4$  with each of the four features for the Bayesian decision theory classifier are calculated as follows:

$$P(T|S_k)1 = \frac{\sum_{i=1}^m \sum_{j=1}^n \rho, \rho = 1 \text{ if } s_{ij}^k = t_{ij}, \text{ other } \rho = 0}{m \times n} \quad (9)$$

$$P(T|S_k)2 = \frac{\sum_{i=1}^m \sum_{j=1}^n |s_{ij}^k - t_{ij}|}{m \times n} \quad (10)$$

$$P(T|S_k)3 = \frac{\sum_{i=1}^m \sum_{j=1}^n |s_{ij}^k - t_{ij}|}{m \times n} \quad (11)$$

$$P(T|S_k)4 = \frac{m \times n - \sum_{i=1}^m \sum_{j=1}^n \rho, \text{ if } \alpha < \frac{t_{ij}}{s_{ij}^k} < \beta, \rho = 1, \text{ other } \rho = 0}{m \times n} \quad (12)$$

where  $P(T|S_k)_{1-4}$  is the conditional probability of the corresponding feature 1-4,  $S = \{S_1, \dots, S_k, \dots, S_c\}$  is the set of prototype classes,  $s_{ij}^k$  is the feature matrix of the  $k$ th class  $S_k$  in the set of prototype classes, and  $t_{ij}$  is the feature matrix of the test sample ( $T$ ). In Equation 9, if  $s_{ij}^k = t_{ij}$ , then  $\rho = 1$ ; otherwise,  $\rho = 0$ . In Equation 12,  $\alpha$  and  $\beta$  are again experience values, and in this work, we set  $\alpha = 0.9$  and  $\beta = 1.1$ .

In the experiment, the prior probability  $P(S_k)$ ,  $1 \leq k \leq c$  of different symbols  $S_k$  is not equal, and all

prior probabilities come from the prior statistics of the training dataset. For example, the beat symbol ‘↘’ has the prior probability 0.354511. If  $P(S_k|T) = \max P(S_j|T)$ ,  $T$  is classified to  $S_k$ . The Bayesian rule was used.

$P(T|S_k)rP(S_k)/\sum_{i=1}^c P(T|S_i)P(S_i)$ ,  $1 \leq r \leq 4$  and has four expressions to estimate  $P(S_k|T)$  which correspond to each of the four features.

### 3.4.3 Genetic algorithm

In the field of artificial intelligence, a genetic algorithm (GA) is a search heuristic that mimics the process of natural evolution. This heuristic is routinely used to generate useful solutions to optimization and search problems. GAs generate solutions to optimization problems using techniques inspired by natural evolution, such as inheritance, mutation, selection, and crossover. In GAs, the search space parameters are encoded as strings, and a collection of strings constitutes a population. The processes of selection, crossover, and mutation continue for a fixed number of generations or until some condition is satisfied. GAs have been applied in such fields as image processing, neural networks, and machine learning [58].

In this work, the key GA parameter values are as follows:

- Selection-reproduction rate:  $p_s = 0.5$
- Crossover rate:  $p_c = 0.6$
- Mutation rate:  $p_m = 0.05$
- Population class:  $C = 12$
- Number of individuals in the population:  $N_p = 200$
- Maximum iteration number:  $G = 300$

An individual's fitness value is determined from the following fitness function:

$$\begin{aligned} F(I_u) &= \sum_{v=1}^R f(e(h_{u,v}, s_{u,v})) = \sum_{v=1}^R f(\rho) \\ &= \sum_{v=1}^R \frac{\rho}{m \times n}, \text{ if } h_{i,j} = s_{i,j}, \text{ then } \rho \\ &= 1, \text{ else } \rho = 0 \end{aligned} \quad (13)$$

where  $I_u$  is the  $k$ th individual,  $F(I_u)$  is the fitness value of  $I_u$ ,  $R$  is the number of a gene bits in  $I_u$ ,  $h_{i,j}$  is the  $j$ th gene of  $I_u$ ,  $s_{i,j}$  is the feature matrix of the set of prototype classes corresponding to  $h_{i,j}$ , the function  $e()$  computes the value of the comparative result between  $h_{i,j}$  and  $s_{i,j}$ , and  $m$  and  $n$  are the width and length, respectively, of a symbol's grid.

### 3.5 Semantic analysis and MIDI representation

After all the stages of the OMR system are complete (see Figure 6), the recognized symbols can be employed to write the score in different data formats, such as MIDI, Nyquist, MusicXML, WEDELMUSIC, MPEG-SMR, notation interchange file format (NIFF), and standard music description language (SMDL). Although different representation formats are now available, no standard exists for computer symbolic music representation. However, MIDI is the most popular digital music format in modern China, analogous to the status of the GCN format in ancient China.

This work selected MIDI for music representation, because the musical melody in a *Kunqu* Opera GCN score provides monophonic information. Thus, the symbols recognized from the GCN score can be represented by an array melody[ $L$ ][2], with  $L$  representing the number of notes in the GCN score, the first dimension in the array representing the pitch of all notes, and the second dimension representing the duration of all notes.

Finally, the note information in melody[ $L$ ][2] can be transformed into a MIDI file using an associated coding language, such as Visual C++, and the MIDI file of the *Kunqu* Opera GCN score can then be disseminated globally via the Internet.

## 4. Experimental results

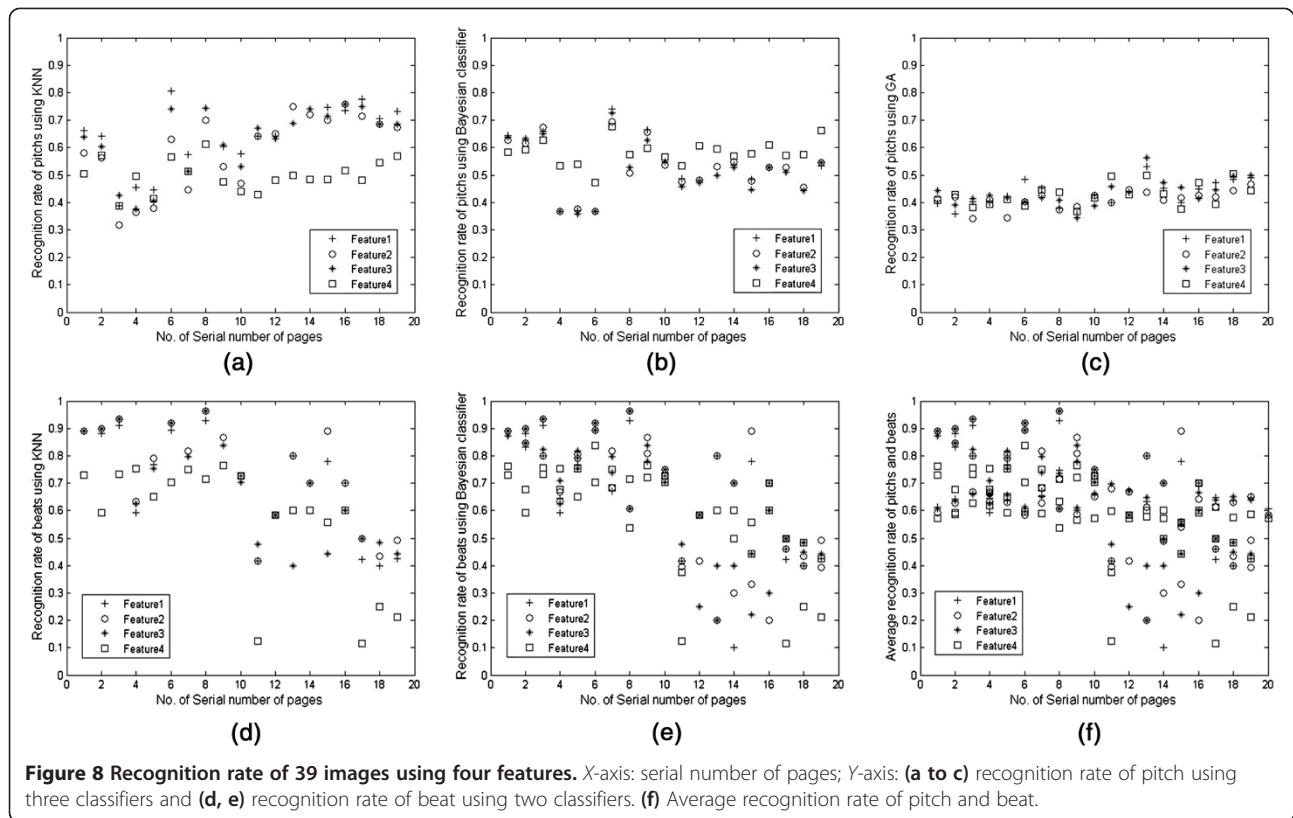
A Chinese *Kunqu* Opera GCN score may contain information as Chinese characters and musical notes, such as lyrics, title, tune, pitch, rhythm, voices, preface of music, and so on. In this experiment, a dataset was randomly selected from two *Kunqu* score books [4,5]. Because OCR for Chinese characters is actively researched [47] and OMR for musical information is a developing field, we consider only the musical note information in this experiment. This includes pitches and rhythm marks (beats) but neglects other information such as lyrics (Chinese characters).

First, all scores in the set were segmented according to [41], and then the musical note information, including

**Table 3 Information on the 39 selected images**

	Name of works	
	Nashu Studio Theatrical Music	Jiugong Dacheng Nanbei Ci Gongpu
Range of pages	pp. 1975-1996	pp. 4326-4344
Range of training pages	pp. 1975-1984	pp. 4326-4335
Range of testing pages	pp. 1985-1996	pp. 4336-4344
Serial number of testing pages	No. 1-10	No. 11 ~ 19
Number of all pages	20	19
Number of pitches in all pages	3,100	2,157
Number of beats in all pages	1,413	661





pitches and beats (rhythm marks), was decoded. Finally, the musical information was translated to a MIDI file.

#### 4.1 Training and test datasets

To estimate the efficiency of KOMR, experiments were conducted using images randomly selected from *Nashu Studio Theatrical Music* [5] and *Jiugong Dacheng Nanbei Ci Gongpu* [4]. Thirty-nine images were used. The primary information from these images is shown in Table 3 and includes the number of pages, pitches, and beats. The resolution was calculated as  $(680-720) \times (880-900)$ . Figure 3 displays a common musical score sheet from page 1983 of *Nashu Studio Theatrical Music*.

To improve the recognition rate of all classifiers, a symbol's grid was used to render a partition of resolution  $16 \times$

$16$  ( $n \times m$  above). Each symbol was then resized to  $16 \times 16$  pixels and converted to a 256-dimensional vector, a feature datum of the symbol.

To create training datasets, we randomly selected images from the available datasets. Specifically, the images on pp. 1975-1984 and pp. 4326-4335 (see Table 3) constituted the training sets, and other images were included in the test sets.

#### 4.2 Results

To evaluate the accuracy of our results, the various classifiers were compared based on their test GCN scores. The three classifiers' pitch recognition rates for the test images are shown in Figure 8, with (a) the KNN classifier (in this work,  $K = 1$ ), (b) the Bayesian classifier, and

**Table 4 Experimental statistics of the recognition rate for all pitches and all beats using features 1 to 4**

Feature	Recognition rate of pitch (%)			Recognition rate of beat (%)	
	BC	GA	KNN	BC	KNN
Feature 1	52.75	43.09	64.80	55.59	70.78
Feature 2	52.59	41.05	59.35	57.10	73.46
Feature 3	51.97	43.60	62.72	58.32	68.17
Feature 4	58.24	42.72	49.85	61.68	56.63
Average rate	53.89	42.62	59.18	58.17	67.26

(c) the GA classifier. The beat recognition rates for (d) the KNN classifier and (e) the Bayesian classifier are also shown. The recognition rate obtained by KNN is higher than that of the other two classifiers.

From the results obtained for the musical notes, shown in Table 4, it can be seen that the KNN classifier outperformed the other two classifiers. Notably, the performance of the KNN classifier was clearly superior to that of the GA classifier. The recognition rate of beats is greater than that of pitches, and the results in Figure 8 indicate that the recognition rate of the beats in *Nashu Studio Theatrical Music* is greater than that in *Jiugong Dacheng Nanbei Ci Gongpu*. This is because the musical symbols for the

beats in the first book included seven different symbols: ♪ ○ △ \_ × ✕ L, whereas there were only four different symbols in the second book: ♪ □ \_ L. Thus, each beat feature in the first book is very different from that in the second book, and the second has less discrimination. Meanwhile, Figure 8 and Table 4 suggest that the four features have little effect on the recognition rate.

Following the recognition stage, the musical information of a GCN score was saved to an array melody[L][2], which was transformed into MIDI format and stored as a MIDI file (Additional file 1). Figure 9 shows the sheet score of the MIDI file in WMN based on the recognition of musical information from Figure 3.

## 雙合

仙呂  
八聲甘州

納書櫟曲譜  
P1983

Figure 9 Sheet score of the MIDI file in WMN for Figure 3.

## 5. Conclusions

As one of UNESCO's Oral and Intangible Heritages of Humanity, *Kunqu* Opera had a deep effect on ancient Chinese entertainment and has been studied in many ways. GCN is the most widely used recording method for ancient Chinese musical information, employed by a significant number of musical works, including all *Kunqu* Opera scores.

This paper presented the six-stage KOMR system, with the key phases being image segmentation and symbol recognition. This paper focused on the recognition phase, studying the effectiveness of three classifiers to obtain musical information from the GCN score. The results indicate that a KNN classifier is most suitable for this task.

However, because the average recognition rate of KOMR is less than 90%, it is necessary to find more powerful methods that can be applied to all phases of KOMR. These problems merit further research.

This work is preliminary and tentative and will help to preserve and popularize Chinese cultural heritage. In particular, this study will assist the conversion of handwritten musical *Kunqu* Opera scores to a machine-readable format, thus ensuring the possibility of disseminating and performing the original *Kunqu* Opera musical scores.

## Additional file

**Additional file 1: Shuanghe.zip.** The file is a MIDI format file which includes musical information from Figure 3 in the main body text.

## Competing interests

The authors declare that they have no competing interests.

## Acknowledgements

This work is supported by the key project of National Social Science Foundation of China Grant No. 12AZD120, the promotion project of National Culture Science and Technology in China Grant No. [2013]718-6, HangZhou Special Funds to Support Cultural Creativity project Grant No. 93, and the planning project of educational technology research in Zhejiang Province under Grant JB001.

## Author details

<sup>1</sup>Department of Computer Science and Technology, Hangzhou Normal University, Hangzhou, Zhejiang 310036, China. <sup>2</sup>Department of Computer Science, National Taipei University of Education, Taipei 10671, Taiwan.

Received: 18 January 2013 Accepted: 14 February 2014

Published: 4 March 2014

## References

- UNESCO, Masterpieces 2001 and 2003: *Kunqu* Opera. UNESCO Physics Web (2008). <http://www.unesco.org/culture/ich/en/RL/00004>. Accessed 12 Oct 2012
- Z Chen, *Introduction to Gong-Che Notation*, 1st edn. (Huayue's Music Publishing House, Beijing, 2006), pp. 1–2
- Y-H Wang, *Musical Notation of Chinese Traditional Music*, 1st edn. (Fujian Education Publishing House, Fuzhou, 2006), pp. 245–246
- X Zhou, J Zhou, X Xu, W Wang, T Zhu, Y Xu, *A comprehensive anthology of texts and notation of the Southern and Northern opera tunes in nine modes*. new, edn (Shanghai Guji Publishing House, Shanghai, 1995), pp. 4326–4344
- T Ye, *Nashu Studio Theatrical Music*. new, edn (Shanghai Guji Publishing House, Shanghai, 1995), pp. 1975–1996
- S Lv, X Yang, H Liu, S Tang, *The Law of the Southern Word*. new edn. (Shanghai Guji Publishing House, Shanghai, 1995), pp. 1–10
- D Pruslin, *Automatic recognition of sheet music*, Dissertation (Massachusetts Institute of Technology, Cambridge, 1966)
- I Fujinaga, *Adaptive optical music recognition*, Dissertation (Faculty of Music, McGill University, 1996)
- K Ng, Optical music analysis for printed music score and handwritten music manuscript, in *Visual Perception of Music Notation: On-Line and Off-Line Recognition*, ed. by SE George (Idea Group Publishing, Pennsylvania, 2003), pp. 108–127
- J Wolman, J Choi, S Asgharzadeh, J Kahana, Recognition of handwritten music notation, in *Proceedings of the International Computer Music Conference* (San Jose, 9–13 Nov 1992), pp. 125–127
- J Caldas Pinto, P Vieira, M Ramalho, MO Mengucci, P Pina, F Muge, Ancient music recovery for digital libraries, in *Proceedings of the 4th European Conference on Research and Advanced Technology for Digital Libraries. Lecture Notes in Computer Science* (Springer, Heidelberg, 2000), pp. 245–249
- W McGee, P Merkley, The optical scanning of medieval music. *Comput. Hum.* **25**(1), 47–53 (1991)
- LJ Tardón, S Sammartino, I Barbancho, Optical music recognition for scores written in white mensural notation. *EURASIP J. Image Video Process.* (2009). doi: 10.1155/2009/843401
- NP Carter, Segmentation and preliminary recognition of madrigals notated in white mensural notation. *Mach. Vis. Appl.* **5**(3), 223–230 (1992)
- L Pugin, JA Burgoyne, I Fujinaga, Goal-directed evaluation for the improvement of optical music recognition on early music prints, in *Proceedings of the 7th ACM/IEEE Joint Conference on Digital Libraries (JCDL'07, Vancouver, 1 April 2007)*, pp. 303–304
- VG Gezerlis, S Theodoridis, Optical character recognition of the Orthodox Hellenic Byzantine Music notation. *Pattern Recogn.* **35**(4), 895–914 (2002)
- A Pikrakis, Recognition of isolated musical patterns in the context of Greek traditional music, in *Proceedings of the Third IEEE International Conference on Electronics, Circuits, and Systems*, vol 2nd edn. (IEEE Xplore Digital Library, Rodos, 1996), pp. 1223–1226
- K Ng, J Barthelemy, B Ong, I Bruno, P Nesi, CIMS: coding images of music sheets. Leeds University PhysicsWeb (2005). <http://music.leeds.ac.uk/people/kia-ng>. Accessed 20 June 2005
- I Fujinaga, *Optical music recognition using projections*, Dissertation (Montreal University, Canada, 1988)
- G Chen, *The study of optical music recognition and prototype system*, Dissertation (Zhejiang University, China, 2003)
- IG Ratiu, Self organizing neural network and direction of arrival estimation technique MUSIC 2D applied in target radiolocation recognition, in *Proceedings of the 4th WSEAS International Conference on Computer Engineering and Applications* (Stevens Point, 2011), pp. 231–236
- HS Park, JO Yoo, SB Cho, *A context-aware music recommendation system using fuzzy Bayesian networks with utility theory. Lecture Notes in Computer Science*. vol. 4223 (Springer, Heidelberg, 2006), pp. 970–979
- I Yoda, K Yamamoto, Automatic construction of recognition procedures for musical notes by genetic algorithm, in *International Association for Pattern Recognition Workshop on Document Analysis Systems* (World Scientific, Hackensack, 1995), pp. 153–159
- MV Stuckelberg, C Pellegrini, M Hilario, An architecture for musical score recognition using high-level domain knowledge, in *Proceedings of the Fourth International Conference on Document Analysis and Recognition*, 2nd edn. (Ulrm, 1997), pp. 813–818
- S Baumann, G Res, A simplified attributed graph grammar for high-level music recognition, in *Proceedings of the Third International Conference on Document Analysis and Recognition*, 2nd edn. (Montreal, 1995), pp. 1080–1083
- G Taubman, *Music hand: a handwritten music recognition system*, Dissertation (Brown University, 2005)
- YL Yang, *Chinese Ancient Music Historical Manuscript* (People's Music Publishing House, Beijing, 2004), pp. 258–259
- Z Tong, J Gu, Y Zhou, X Sun, *Chinese Traditional Musicology* (Fujian Education Publishing House, Fuzhou, 2003), pp. 73–75
- A Rebelo, G Capela, JS Cardoso, Optical recognition of music symbols. *UDAR* **13**, 19–31 (2010)
- N Otsu, A threshold selection method from gray-level histogram. *IEEE. Trans. Syst. Man Cybern.* **9**, 62–66 (1979)

31. L Najman, H Talbot, *Mathematical Morphology* (Wiley, Hoboken, 2010), pp. 12–13
32. YH Liang, ZY Wang, A skew detection method for 2D bar code images based on the least square method, in *2006 International Conference on Machine Learning and Cybernetics, Dalian*, 2006, pp. 3974–3977. doi:10.1109/ICMLC.2006.258793
33. RC Gonzalez, RE Woods, SL Eddins, *Digital image processing using MATLAB* (Prentice Hall, 2004), pp. 393–403
34. G Nagy, S Seth, Hierarchical representation of optically scanned documents, in *Seventh International Conference on Pattern Recognition* (Montreal, 1984), pp. 347–349
35. FM Wahl, RG Wong, KY Casey, Block segmentation and text extraction in mixed text/image documents. *Comput. Graph. Image Process.* **20**, 375–390 (1982)
36. M Feldbach, KD Tonnies, Line detection and segmentation in historical document analysis and recognition. *IJDAR* **14**, 743–747 (2011)
37. R Manmatha, JL Rothfeder, A scale space approach for automatically segmenting words from historical handwritten documents. *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(8), 1212–1225 (2005)
38. K Kise, A Sato, M Iwata, Segmentation of page images using the area Voronoi diagram. *Comput. Vis. Image Underst.* **70**(3), 370–382 (1998)
39. RD Duda, PE Hart, Use of the Hough transform to detect lines and curves in pictures. *Commun. ACM* **15**(1), 11–15 (1972)
40. N Nikolaou, M Makridis, B Gatos, N Stamatopoulos, N Papamarkos, Segmentation of historical machine-printed documents using adaptive run length smoothing and skeleton segmentation paths. *Image Vis. Comput.* **28**, 590–604 (2010)
41. G-F Chen, W Zhang, H Cui, Extracting notes from Chinese *Gong-Che* notation musical score image using a self-adaptive smoothing and connected component labeling algorithm. *Int. J. Adv. Comput. Technol.* **4**(1), 86–95 (2012)
42. TH Hildebrandt, WT Liu, Optical recognition of handwritten Chinese characters: advances since 1980. *Pattern Recogn.* **26**, 205–225 (1993)
43. RI Oka, Handwritten Chinese-Japanese character recognition by using cellular feature. *Proc. Int. Conf. Pattern Recogn.* **2**, 783–785 (1982)
44. OD Trier, AK Jain, T Taxt, Feature extraction methods for character recognition—a survey. *Pattern Recogn.* **29**(4), 641–662 (1996)
45. AW Senior, AJ Robinson, An off-line cursive handwriting recognition system. *IEEE Trans. Pattern Anal. Mach. Intell.* **20**, 309–321 (1998)
46. A Amin, Off-line Arabic character recognition: the state of the art. *Pattern Recogn.* **31**(5), 517–530 (1998)
47. CL Liu, F Yin, DH Wang, QF Wang, Online and offline handwritten Chinese character recognition: benchmarking on new databases. *Pattern Recogn.* **46**, 155–162 (2013)
48. S Basu, N Das, R Sarkar, M Kundu, M Nasipuri, DK Basu, A hierarchical approach to recognition of handwritten Bangla characters. *Pattern Recogn.* **42**, 1467–1484 (2009)
49. CW Liao, JS Huang, A transformation invariant matching algorithm for handwritten Chinese character recognition. *Pattern Recogn.* **23**(1), 1167–1188 (1990)
50. X Lin, X Ding, M Chen, R Zhang, W Youshou, Adaptive confidence transform based classifier combination for Chinese character recognition. *Pattern Recogn. Lett.* **19**, 975–988 (1998)
51. RD Romero, DS Touretzky, RH Thibadeau, Optical Chinese character recognition using probabilistic neural networks. *Pattern Recogn.* **30**(8), 1279–1292 (1997)
52. D Shi, RI Dampier, SR Gunn, Offline handwritten Chinese character recognition by radical decomposition. *ACM Transac. Asian Lang. Info. Process.* **2**(1), 27–48 (2003)
53. IJ Kim, JH Kim, Statistical character structure modeling and its application to handwritten Chinese character recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **25**(11), 1422–1436 (2003)
54. J Zeng, ZQ Liu, Markov random field-based statistical character structure modeling for handwritten Chinese character recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **30**(5), 767–780 (2008)
55. J Tan, X Xie, WS Zheng, JH Lai, Radical extraction using affine sparse matrix factorization for printed Chinese characters recognition. *Int. J. Pattern Recognit. Artif. Intell.* **26**(3) (2012). doi:10.1142/S021800141250005X
56. T Cover, P Hart, Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* **13**(1), 21–27 (1967)
57. R Duda, P Hart, *Pattern Classification and Scene Analysis* (Wiley, New York, 1973)
58. W Gong, Z Cai, L Jia, H Li, A generalized hybrid generation scheme of differential evolution for global numerical optimization, international journal of computational intelligence and applications. *Int. J. Comp. Intel. Appl* **35**(10), 560–567 (2011). doi: 10.1142/S1469026811002982

doi:10.1186/1687-4722-2014-7

**Cite this article as:** Chen and Sheu: An optical music recognition system for traditional Chinese *Kunqu* Opera scores written in *Gong-Che* Notation. *EURASIP Journal on Audio, Speech, and Music Processing* 2014 **2014**:7.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)