

КОМП'ЮТЕРНА ТЕХНОЛОГІЯ ДЛЯ РОЗПІЗНАВАННЯ ТИПОГРАФСЬКИХ НОТНИХ ТЕКСТІВ¹

COMPUTER TECHNOLOGY FOR PRINTED NOTES RECOGNITION

Шлезінгер М.І.², Савчинський Б.Д.³, Анохіна М.О.

Міжнародний науково-навчальний центр ЮНЕСКО/МПІ інформаційних технологій та систем.
пр. Академіка Глушкова 40, Київ т.266-62-08

Анотація.

Описано принципи побудови та експериментальну перевірку комп'ютерної технології для розпізнавання типографських нотних текстів. Технологія базується на новітніх досягненнях у структурному розпізнаванні складних зображень, що складені з великої кількості взаємопов'язаних елементів. Врахування контекстних залежностей між символами нотного тексту дозволяє суттєво покращити якість розпізнавання.

Abstract.

Design principles and experimental testing of computer technology for printed notes recognition are described. Technology is based on new advances in structural recognition of complex objects consisting of large amount of mutually connected components. Taking into account context dependences between symbols of note text allows to increase substantially recognition rate.

1. Вступ

Розпізнавання нотних текстів, окрім свого безсумнівного культурологічного значення, має широкий спектр застосувань при створенні електронних музичних бібліотек, автоматизації рутинної ручної праці при переписуванні партитури на окремі партії, транспонуванні музичних творів та в багатьох інших практичних роботах.

Інтерес до проблеми розпізнавання нотних текстів стимулюється також її внутрішнім науковим змістом. Зображення нотного тексту є яскравим прикладом об'єктів із складною, але цілком визначеною внутрішньою структурою, що дозволяє застосувати новітні методи структурного розпізнавання. Науковим та прикладним проблемам розпізнавання нотних текстів присвячені відомі роботи [1,2,3]. Повнішу бібліографію містить [4].

У наступному розділі описані основні ідеї, на яких базується розроблена технологія розпізнавання. У заключному розділі подаються результати експериментальної перевірки застосованих методів та висновки.

2. Структура зображення сторінки нотного тексту. Неформальний погляд.

Процеси, які формують технологію розпізнавання, базуються на вирішенні задач наступного класу.

Нехай X — це множина всіх можливих зображень, E — відома її підмножина, яка далі розуміється як множина певних ідеальних, еталонних зображень; $f: E \times X \rightarrow R$ — це задана функція така, що $f(e, x)$ характеризує, наскільки реальне зображення x відрізняється від ідеального зображення e .

¹ Робота була виконана в рамках української державної програми «Образний комп'ютер» і підтримувалась програмою DAAD уряду Німеччини.

² schles@image.kiev.ua

³ bogdan@image.kiev.ua

Вирішувана задача полягає в тому, щоб для заданої множини E та функції $f : E \times X \rightarrow R$ побудувати алгоритм, який для кожного вхідного зображення x вказує еталонне зображення e^* з множини E , що найменше відрізняється від x :

$$e^* = \arg \min_{e \in E} f(e, x). \quad (1)$$

Опишемо множину E ідеальних зображень нотних текстів. Дану множину доречно визначити за допомогою уявного процесу малювання таких зображень. Це відомий і широкоживаний у структурному розпізнаванні спосіб, коли множина об'єктів задається за допомогою генеративної моделі (див. [5]).

Перший етап полягає в послідовному розбитті чистого аркуша на горизонтальні смуги двох типів. Смуги першого типу визначають ті місця на папері, де будуть знаходитись рядки нотного тексту. Смуги другого типу – це проміжки між рядками. Таким чином, наслідком першого етапу є певний ескіз майбутнього зображення. Смуги другого типу – це та частина зображення, яка вже не буде змінюватися, в той час, як смуги першого типу визначають ті місця, в яких щось має домалюватися на наступних етапах.

На другому етапі оброблюється кожна смуга першого типу, побудована на першому етапі. Перш за все на кожній такій смузі малюється зображення п'яти ліній нотного стану. Подальше оброблення смуги полягає в тому, що по своїй довжині, тобто в горизонтальному напрямку, вона розбивається на послідовність прямокутників, що щільно прилягають один до одного. Ці прямокутники також поділяються на два типи. В прямокутниках другого типу малюється зображення, яке вже не буде змінюватися, в той час, як прямокутники першого типу позначаються певним чином для визначення того, що зображення в них буде створене на третьому етапі.

В прямокутниках другого типу малюються зображення елементарних музичних символів, тобто тих, що не складаються з інших, більш простих. Це такі символи, як ключі, паузи, тактові риси, позначення тактового розміру та інші. До складу таких символів входить ще особливий символ, що означає проміжок між власне музичними символами. Саме тому ми говоримо, що прямокутники, на які розбивається рядок нотного тексту, щільно прилягають один до одного.

Прямокутники першого типу призначені для музичних символів, що складаються з більш простих, елементарних. Це акорди, які являють собою вертикальну послідовність нот. Кількість різних акордів величезна в порівнянні з кількістю простих музичних символів, про які йшлося в попередньому абзаці.

Третій етап полягає в тому, що в прямокутниках першого типу, створених на другому етапі, генерується зображення акордів. Під акордом розуміється складний об'єкт, що за певними правилами складається з таких елементарних музичних символів, як ноти (двох типів: чорні та білі), штилі, прапорці, що позначають тривалість акорду тощо.

Вказаний процес породження зображень нотних текстів ілюструє, яким чином визначається множина ідеальних зображень, на якій базується формальна постановка задачі їх розпізнавання (див. (1)).

Якби йшлося про розпізнавання лише ідеальних зображень, то воно би зводилось до відновлення послідовності дій, результатом яких є саме те зображення, яке треба розпізнати. Справді, знаючи цю послідовність, неважко визначити послідовність акордів, як сукупність звуків, що звучать одночасно, та тривалість кожного акорду. Але через неминучу неідеальність реальних зображень їх розпізнавання вимагає вирішення оптимізаційної задачі

(1). Це значить, що треба знайти таку послідовність дій, яка приводить до створення зображення, що найменше відрізняється від вхідного.

Послідовний характер створення зображення робить природним побудову алгоритму розпізнавання як послідовне відтворення етапів його створення у тому ж порядку, в якому зображення створювалось. Це значить, що спочатку, на основі певних розумних евристичних міркувань, визначають місцеположення окремих рядків нотного тексту. Потім на кожному виявленому рядку визначають місцеположення окремих музичних символів, включаючи акорди. Нарешті, аналізують кожний виявлений акорд з метою визначення звуків, що його складають та їх тривалості. Такий поширений підхід (дивись, наприклад, [1,3,4]) має дуже серйозні недоліки. Вони полягають в тому, що на кожному етапі розпізнавання приймається кінцеве рішення, яке може бути і хибним, але не може бути виправлене на наступних етапах. Більш за це, хибне рішення на якомусь етапі неминуче призводить до помилок на наступних етапах.

Таких недоліків не має алгоритм, що базується на точному вирішенні задачі (1). Хоча кількість зображень у множині E експоненційно залежить від довжини музичного тексту, задача (1) може бути точно вирішена без повного перегляду всіх зображень у цій множині. Таке вирішення задачі (1) базується на сучасних методах структурного розпізнавання (див.[5]). Саме завдяки послідовному характеру генерування ідеальних зображень вони можуть розглядатися як речення у певних формальних мовах, схожих на контекстно-вільні формальні граматики Хомського [6]. Однак, специфіка нашої прикладної задачі полягає в тому, що музичний текст не є послідовністю елементарних символів, яка впорядкована в одному напрямку. Елементарні музичні символи розташовані один відносно іншого як в горизонтальному, так і в вертикальному напрямках. Структурний аналіз таких складних утворень вимагає використання більш загальних конструкцій, ніж контекстно-вільні мови та граматики Хомського. Це так звані двовимірні контекстно-вільні граматики та мови, що визначені та досліджені у [5]. Там же наведений загальний алгоритм вирішення задачі (1) у випадку, коли E є двовимірною контекстно-вільною мовою. Дослідження специфіки зображень музичних текстів як об'єктів машинного аналізу дозволяє сконструювати алгоритм розпізнавання, що є на порядок ефективнішим, ніж відомий загальний алгоритм.

Саме такий ефективний алгоритм був використаний для розпізнавання типографських нотних текстів. Отримані результати обговорюються в наступному розділі.

3. Демонстрація результатів

Експериментальна перевірка розробленого методу мала на меті верифікацію основних ідей, на яких базується розроблений алгоритм. Тому експерименти виконувались на простих, але реальних музичних текстах, а не на текстах довільної складності. Тексти були взяті із збірника музичних вправ для учнів початкових класів музичної школи. Ці нотні тексти задовольняють наступним обмеженням:

- акорди в межах одного такту утворюють послідовність, а не дві послідовності, як це є при записі верхнього та нижнього голосу на одному нотному стані;
- нотні тексти не містять ліг;
- акорди зображені окремо один від одного, тобто не утворюють груп, поєднаних ребрами;
- нотні тексти не містять форшлагів, динамічних позначень (таких, як форте, піано, крещендо) та інших позначень, що визначають настрій, темп, постановку пальців і т. д.

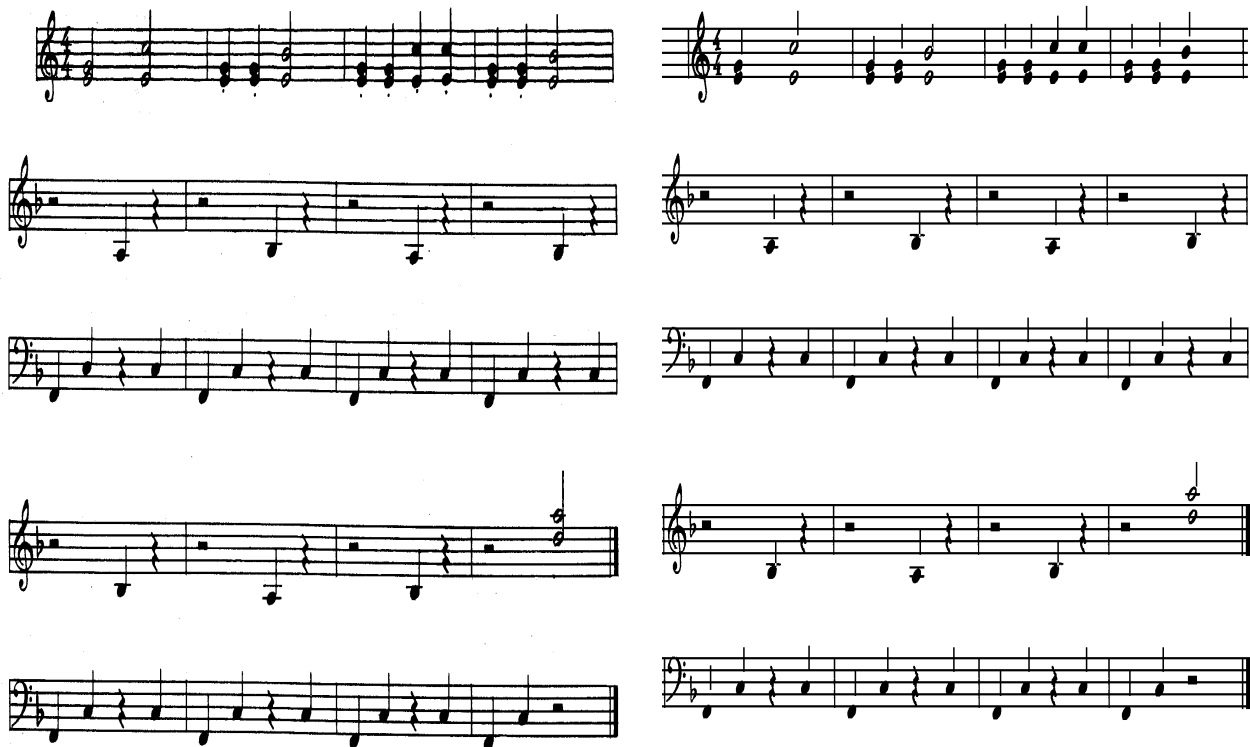


Рис. 1 Один з тих нотних текстів, на яких проводилось тестування.

Рис. 2: Результат розпізнавання тексту, зображеного на рис.1.

Приклади таких зображень наведені на рис.1. Негативні результати експериментів на зображеннях такого класу вимагали би перегляду усієї концепції, покладеної в основу розробленої технології. Позитивні результати дозволяють стверджувати, що подальше вдосконалення програми, тобто розширення класу музичних текстів, вимагатиме лише кількісного вдосконалення технології: збільшення кількості елементарних символів, ускладнення граматики тощо.

Тестування розроблених методів проводилось на 15 зображеннях сторінок нотного тексту, кожна з яких містила 3 - 4 нотні рядки. Якість розпізнавання характеризується тим, що в середньому із 100 музичних символів 3 були розпізнані неправильно. Положення та кількість нотних рядків на всіх розглянутих зображеннях визначались без помилок. Допущені ж помилки можуть бути легко пояснені і так само легко усунені в рамках розроблених методів. На рис.1 зображено один з прикладів нотних текстів, на яких проводилось тестування. Результати розпізнавання показані на рис.2. Правильно були розпізнані кількість і положення нотних рядків та послідовності символів нотного тексту в усіх нотних рядках, крім першого, де припущено дві помилки: два половинні акорди були розпізнані як четвертні.

Для порівняння якості роботи алгоритмів ті ж 15 тестових зображень були подані на вхід відомої комерційної програми розпізнавання нотних текстів MIDISCAN for Windows версії 3.0. При їх розпізнаванні за допомогою цієї програми 15 музичних символів із 100 розпізнавались неправильно. Оскільки ж MIDISCAN не містить опису алгоритмів, які використовуються ним для розпізнавання, то щодо пояснення допущених помилок можуть існувати лише здогадки.

Література

1. T. Beran, T. Macek. Recognition of printed music score. In: *MLDM'99*, LNAI 1715, pp. 174-179. Springer-Verlag 1999.

2. B. Coūasnon, B. Rétif. Using a grammar for a realible full score recognition system. In: *Proc. of the International Computer Music Conference*, Canada, Sept.1995.
3. J.C. Pinto, P. Viera, M. Ramalho, M. Mengucci, P. Pina, F. Muge. Ancient music recovery for digital libraries. In: J. Borbinha, T. Baker (Eds.): *ECDL 2000*, LNCS 1923, pp. 24-34, 2000. Springer-Verlag, 2000.
4. D. Blostein, H.S. Baird. A critical survey of music image analysis. In: *Structured Document Image Analysis*, ed. par Springer-Verlag, pp.405-434. Eds. H.S. Baird, H. Bunke, K. Yamamoto, 1992.
5. Schlesinger M.I., V. Hlavač. Deset prednasek z teorie statistickeho a strukturniho rozpoznavani. Vydavatelstvi CVUT, Praha, 1999.
6. А. Ахо, Дж. Ульман. Теория синтаксического анализа перевода и компиляции. изд-во МИР, Москва, 1978.