國 立 清 華 大 學

碩 士 論 文

自動光學檢測系統的演算法改良

An Algorithm improve on AOI System

系所別：資訊系統與應用研究所碩士班

學號姓名：105065527 呂昊叡 (Hao-Jui Lu)

指導教授：韓永楷 博士 (Prof. Wing-Kai Hon)

中 華 民 國 107 年 7 月

誌謝

感謝…

Acknowledgements

I’m glad to thank…

摘要

在工業檢測的場合中，自動化影像辨識技術日漸成為一個重要的應用，而現成的影像處理軟體與演算法通常有授權費過高，且準確度與分析速度並不符合產線的預算與需求，對此問題，我們希望可以找到一個低成本的解決方案，同時滿足產線對於分析速度的需求。本研究所實作出的產品檢驗流程可以粗略分為兩步驟，第一步驟為樣板設定，此步驟會紀錄標準的產品特徵；第二步驟為樣本檢驗，此步驟會將樣本與第一步驟所記錄下的樣本進行比對，並判斷此背光鍵盤是否有瑕疵或故障。本論文主要討論自動化光學檢測系統及分析演算法的設計架構與分析過程中的演算法比較並加以改良。並在最後將嘗試過的各種方法在產線的標準下進行比較。

Abstract

The purpose of optical music recognition is to develop a computer program that is able to understand the musical score, which is invented for human beings to annotate melody. A score is usually stored as an image. Therefore, a recognition system must retrieve musical information from a set of pixels. This dissertation deals with two major issues: preprocessing and recognition. Preprocessing aims at dividing the input image into several slices that can be processed independently and handling the defects in the printing step. The goal of preprocessing is to simplify the subsequent recognition stage. Afterward, recognition on a staff image is the core of this dissertation. The implementation is based on template matching and the support vector machine. For real score images, the present algorithm works well. The design of the present algorithm brings a different perspective to optical music recognition. First, the preprocessing uses *random sample consensus* (RANSAC) as a part of staff detection. Such randomness makes it meaningful to repeat the same operation; by comparing the results between different iterations, consensus-based correction provides possibility of finding symbols that other existing stable algorithms cannot find. Secondly, the algorithm is based on the *divide and conquer* concept, which means the subtasks have little correlation, and hence the algorithm can be readily parallelized.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

High-tech tools are prevalent nowadays and many of our daily are now routinely performed with computers. People write articles with computers; people draw diagrams with computers; people, of course, design programs with computers. Among our various usages of computers, one of them is music composition. For the purpose of storing and visualizing musicians' creation, the standard western musical score, which contains information pertaining to how a piece of music should be played, has been used for hundreds of years and around the globe. However, the score was designed for human beings instead of computers, and most of scores are scanned and stored as images, which means nothing but lots of pixels for computers. In other words, these scores are not yet symbolically represented. Therefore, the concern of this dissertation is *optical music recognition* (OMR), which refers to the development of methods that automatically convert score images into their symbolic representation.

1.2 Goal

Design a software that converts a score image (.png / .jpeg / .bmp / .pdf) into its symbolic representation encoded in a format that is readable by a computer such as MusicXML.

1.3 Divide and Conquer

1.3.1 Definition

Fig. 1.1 shows the concepts of *divide and conquer* (D&C). D&C is an algorithm design paradigm that breaks a complex problem into a couple of relatively simple subproblems, to *divide*, then solves them respectively, to *conquer*. Before conquering, the problem will be divided recursively until it is simple enough to be processed. Finally, the solutions to the subproblems will be merged as those to the original problem.

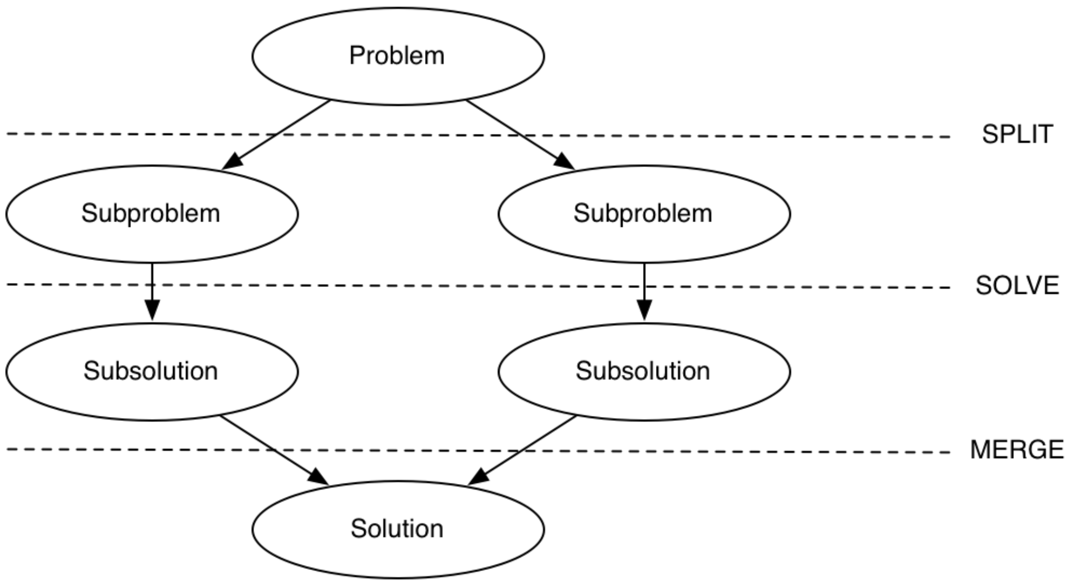

Figure 1.1: A diagram showing how divide and conquer works.

1.3.2 Main Contribution of This Dissertation

Reducing the Difficulty of Problems

Due to characteristics of D&C, all problems that can be accurately split are expected to be solved. For this dissertation, particularly, if the function detecting staves is reliable, then we can analyze arbitrarily complicated scores.

Independence of Subproblems

Typically, a score contains something useless for recognition such as the metadata of the song, lyrics, and even printed defects. By partitioning the original images into subimages where each contains only one staff, the amount of noisy information can be reduced and interference between staves is eliminated. Therefore, the detection tasks are independent between different staves.

Parallelism

Nowadays, a processor usually has multiple cores, and lots of computational tasks are implemented to be executed with parallel programs. In D&C algorithm, the functions solving split subproblems are identically designed. With high independence and similar operations between subproblems, it is a good strategy to process them simultaneously. In other word, the original problem is suitable to be solved with *SIMD (Single-Instruction-Multiple-Data)* parallel programs.

Chapter 2

Overview of OMR

In this section, previous works of OMR are mentioned. Preprocessing (binarization, staff profiling, staff detection, and staff removal) and recognition (symbol segmentation, symbol classification) are included.

2.1 Binarization

In recognition of printed scores, the color information, namely R/G/B or R/G/B/A vectors, is not useful. Instead, only the intensity information is considered for recognition, so gray-scaled images are always used as the raw input. Furthermore, people always determine if each pixel is background (white) or foreground (black) in advance, and hence the binarization is included in most applications of OMR.

In Pinto's research [1], two kinds of binarization methods were introduced depending on whether the binarization threshold is locally adjustable. The simplest way is applying a constant threshold to all pixels in the image, which is called *global thresholding*. The global threshold can be obtained by finding a value that maximizes the variance [2] between foreground and background pixels, preserves the most edge information [3], or maximizes the similarity between the binarized image and the original image [4, 5]. However, it cannot be expected that the intensity in different small regions is constant over the document, and a constant threshold might not work at a different intensity level. In particular, near the boundary of a page in a book, the image might show a gradient-like difference in terms of the average intensity as compared to the region far from the book spine (Fig. 2.1). To deal with such situations, the choice of the threshold should be determined by local information (nearby pixels) [6], which is called *local thresholding*. In general, global thresholding is easier to be implemented, while local thresholding is more adaptive and robust.

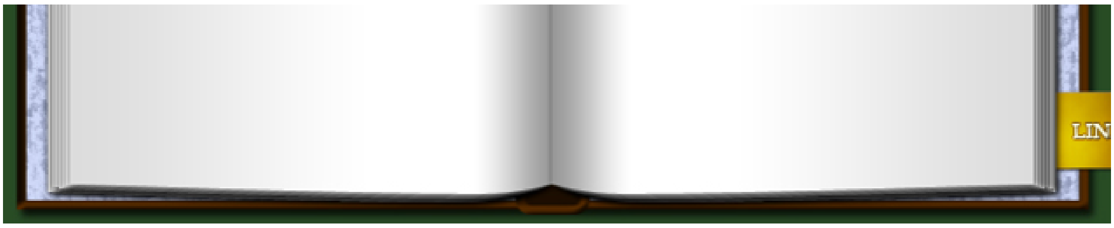

Figure 2.1: Example of the gray-scale image near the book spine.

2.2 Staff Detection and Removal

Dalitz et al. [7] introduced a systematic way for testing the staff removal algorithms. A dataset was generated from a set of ideal score images with the deformation methods listed in Table. 2.1. The deformation algorithms and the CVC-MUSCIMA dataset are made openly available by Forns et al. [8].

Deformation	**Type**	**Parameter Description**
Curvature	deterministic	height/width ratio of sine curve
Typeset Emulation	both	gap width, maximal height and variance of vertical shift
Line Interruptions	random	interruption frequency, maximal width and variance of gap width
Thickness Variation	random	Markov chain stationary distribution and inertia factor
y-variation	random	Markov chain stationary distribution and inertia factor
Degradation	random	emulating local distortions suggested by Kanungo et al. [**?**]
White Speckles	random	speckle frequency, random walk length and smoothing factor

Table 2.1: Deformation Methods.

References

[1] T. Pinto, A. Rebelo, G. Giraldi, and J. S. Cardoso, “Music score binarization based on domain knowledge,” *Pattern Recognition and Image Analysis - 5th Iberian Conf. (IbPRIA)*, pp. 700–708, 2011.

[2] O. Nobuyuki, “A threshold selection method from gray-level histograms,” *IEEE Trans. Systems, Man and Cybernetics*, vol. 9, pp. 62–66, 1979.

[3] Q. Chen, Q.-s. Sun, P. A. Heng, and D.-s. Xia, “A double-threshold image binarization method based on edge detector,” *Pattern Recognition*, vol. 41, pp. 1254–1267, 2008.

[4] L.-K. Huang and M.-J. J. Wang, “Image thresholding by minimizing the measures of fuzziness,” *Pattern Recognition*, vol. 28, pp. 41–51, 1995.

[5] D.-M. Tsai, “A fast thresholding selection procedure for multimodal and unimodal histograms,” *Pattern Recognition Letters*, vol. 16, pp. 653–666, 1995.

[6] J. Bernsen, “Dynamic thresholding of grey-level images,” in *Proc. the 8th. Int. IEEE Conf. CAD Systems in Microelectronics (CADSM)*, pp. 1254–1267, 2005.

[7] C. Dalitz, M. Droettboom, B. Pranzas, and I. Fujinaga, “A comparative study of staff removal algorithms,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 30, pp. 753–766, 2008.

[8] A. Forns, A. Dutta, A. Gordo, and J. Llads, “Cvc-muscima: A ground-truth of handwritten music score images for writer identification and staff removal,” *Int. J. on Document Analysis and Recognition*, vol. 15, pp. 243–251, 2012.