TNM034 - Advanced Image Processing

HT 2018

Project:

Optical Music Recognition

TNM034 – Advanced Image Processing

In the course TNM034 - Advanced Image Processing, the students will learn advanced image processing and current research issues, mainly by carrying out a project task in the area. After the course the students will be able to:

- plan how to solve the project goals.
- describe the project context and the theories behind the solution.
- identify and implement a particular solution to the problem.
- compile a report over the project, describing the implementation and the algorithms used.

The main target for the students is to assimilate the literature and independently solve and implement the given course task. The course consists of lectures and practical laboratory work. The lectures will give an overview of the topic and some background and theory related to the given project. The laboratory work is devoted to the accomplishment of the software program. The project work is performed in groups of 2 to 4 students.

Knowledge in writing Matlab programing and the basic course on image processing (TNM087) are prerequisites for participation in the course.

The list below shows examples of various projects over previous years. The projects have been chosen because of their interesting content of advanced image processing:

- Restoration of image sequences
- Verification of fingerprints
- ”Snakes” and ”Live Wire”
- Stitching and image assembly
- QR-code processing
- Face recognition

The project task 2018: Optical Music Recognition

This project course will be devoted to pattern recognition in the context of optical reading of printed sheet music, or more specifically, to implement software for Optical Music Recognition. The title page shows an example of captured sheet music. The task is simply to extract the musical content from such images, a task that will require solutions to multiple sub-problems, such as pre-processing, segmentation, matching, feature extraction and decision theory.

Introduction to Optical Music Recognition

Optical character recognition (OCR) has been commercially available for many years, and was first introduced in the mid 1950’s for reading credit card imprints. The commercial use accelerated when several postal services and banks started using it for mail sorting and reading. Parallel to the development of OCR, the first studies in *Optical Music Recognition* (OMR) took place. The first commercial products for optical music recognition were introduced in the beginning of the 1990’s, and today OMR is included in many popular music programs.

Theoretically, recognition of printed sheet music is very much resembling optical character recognition, OCR. However, there are several additional problems making OMR a more challenging problem. In OCR, the input is usually a page of text in several rows of words and characters. Essentially, it is a one-dimensional output in the form of a string of characters. The relationship between the objects in the image (the characters) is only the lateral position, relating previous and next object. In sheet music there is not a simple ordering of the objects. The pitch of a note is dependent on its position in relation to the staffs. Several notes can appear at the same position within a score, which makes it even more difficult. On top of this, there are a large number of modifiers in the form of dotted notes, pitch and tempo changing characters.

The problem of pattern recognition is often described as a stepwise process on an increasingly level of abstraction. First, preprocessing of the raw data is carried out, to improve the possibilities for the following recognition. Noise removal and photometric transformations are examples of this process. Tentative measurements for establishing size and rotation usually follow. Most often the image that is analyzed contains many objects that have to be separated, using segmentation of symbols. Since the sheet music may include text and other symbols, it has to be established whether they are part of the music or not. Finally the extracted notes are compressed into a symbolic description of the music.

Carrying out the project work

The work consists of reading applicable literature, testing ideas and finally carrying out the programming according to your own ideas. The lectures will provide an overview of the area of optical music recognition, as well as some theoretic background related to different problems that need to be addressed in the project. The aim is to give you an understanding of the general problem and some ideas of different ways to approach it, not to tell you "how to do it". The approach to solve the problem must be decided by you and your group. There are few restrictions on how to do it. The project group should have four members maximum. All group members should take part in the work in all aspects, and be able to describe the design of the software and the methodical approach.

In order to facilitate the testing of your program you have to use a certain functional form. Below you can find the skeleton of how your program should be designed:

```
%%%%%%%%%%%%%%%%%%%%%%%%%%
function strout = tnm034(Im)
%
% Im: Input image of captured sheet music. Im should be in
% double format, normalized to the interval [0,1]
%
% strout: The resulting character string of the detected notes.
% The string must follow the pre-defined format, explained below.
%
% Your program code…
%%%%%%%%%%%%%%%%%%%%%%%%%%
```

Your program must be written using Matlab. In the Matlab environment you will find many valuable functions that you may call in your program. Other program libraries are not allowed. You can use any number of functions, but the format of the calling function should adhere to the skeleton above. The testing of your program will be done automatically, in a batch run, with a large number of input mages. Therefore your code should not include any pop-up windows or any kind of interaction, which will interrupt the testing.

Besides the program, you are required to deliver a report that contains a description of the problem, the chosen methodology for how to solve the problem and a description of your program implementation. The report can be written in English or Swedish.

The output format

The output from the program should be a character string representing the detected notes. An important limitation is that you should detect and register *quarter notes* (filled notes without a flag), and *eight notes* (notes with one flag) only. Different quarter notes and eight notes must be described with the labels given in Figures 1 and 2 respectively. Rows are recorded from left to right, and a "line break" on the musical sheet is simply marked with the character '*n*'. If chords (doubled notes, etc.) are present, you should register each note separately, from top to bottom. All other notes or signs, like breaks, accidentals etc., should be ignored.

1/4

Figure 1. The encoding of the crotchets (quarter notes) using uppercase letters.

1/8

Figure 2. The encoding of the quavers (eights notes) using lowercase letters.

A correct character array notation (an example of what your program might return) of an example note sheet is given in Figure 3. Study the image and the output string and make sure you understand the connection between them.

Figure 3. The correct output siring is: C3E3F2d3c3b2a2F2a2a2a2G2E2B2E2d3a2b2nB2d3c3b2a2.

When working in Matlab, a simple way to add characters to a string is to use the following syntax:

```
strout=''; %An empty string
strout=[strout,'D0']; %Add the note D0 to the end of the string
```

Course organization

All students registered to the course will have access to the course space on Lisam. Here we will share lectures notes and relevant literature, and also announcements regarding the course. You will also find a number of training images of music sheets, that you can use to test your code. The images with the suffix ´**s**´ refer to scanned images, while the images with the suffix ´**c**´ are captured with a digital camera, and may require additional preprocessing to correct for non-uniform illumination, geometric transformations, etc.

Under the course space on Lisam, it is also possible to discuss problems and solutions between the groups, and to share interesting references. Helping each other and discussing solutions to problems within the project is allowed, and encouraged. However, sharing and directly copying code between groups is not allowed.

Examination

The course will give 6 hp with the grades 3, 4 or 5, where grade 5 requires an individual oral exam.

Grades are given according to the evaluation of your report, the implementation of the software and the performance of the implemented methods. The following requirements must be fulfilled:

- **Grade 3**: A report describing your approach, including references to used methodology. Your program should work according to the instructions for sheets with single notes (no chords), captured by a scanner. The images will have good registration and may be slightly rotated (e.g. *im1s* and *im3s* in the training set).

- **Grade 4**: Your report must further include discussions on alternative approaches to solve the problem. Your program should work reasonably well also for sheets with multiple notes (chords) and images captured by a digital camera. The images may be slightly out of focus, contain non-uniform illumination and geometric transformations (rotation, perspective, barrel distortion).

When evaluating the performance of your program, we will also use images of sheet music that is not included in the training set, but with the same level of difficulty. The number of students in each group will also be considered during the evaluation.

Students that fulfill the requirements for grade 4 will be given the opportunity to take an individual oral exam for grade 5. During the oral exam (approximately 20 minutes), the students will be asked to explain their work, motivate the selected methods and relate them to alternative solutions, and to answer questions to demonstrate a good understanding of the topic. The oral exam is given in English or Swedish.

Literature

The lectures will partly be based on selected research papers, which will be made available on the course space on Lisam. They are by no means the only background material for the project, but should rather be seen as a starting point for further exploration of the literature. A search for “optical music recognition” on Google Scholar (http://scholar.google.com) returns 217 000 hits. For pattern recognition in general, there is a huge amount of available literature that is relevant for this project, so you will not have to worry about finding relevant reading material.

Deadline

The program code and the report must be uploaded, using the submission function on Lisam, no later than Sunday **December 16, 2018**. Don’t forget to include the identities of all the members of the group!

Good luck!

Daniel Nyström (examiner)