Discotective

UNIVERSITY OF MICHIGAN
Michigan Engineering

EECS 452
DSP Design Laboratory

An optical music recognition project developed by
Katie Bouman, Brad Campbell, Michael Hand, Tyler Johnson, Joe Kurleto

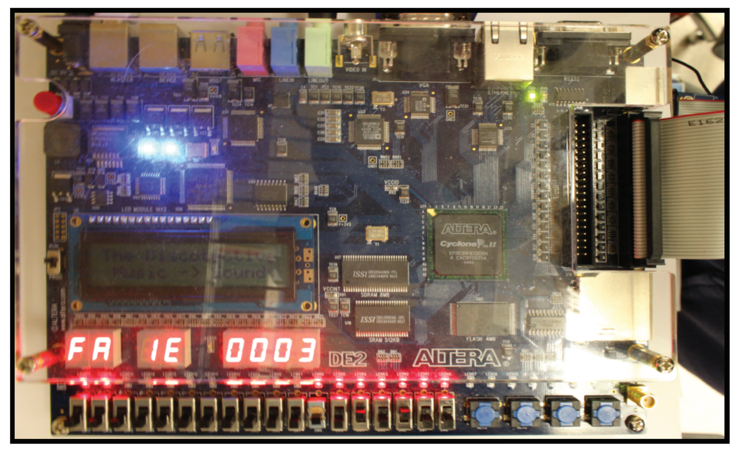
Winter 2011
Design Expo

INTRODUCTION

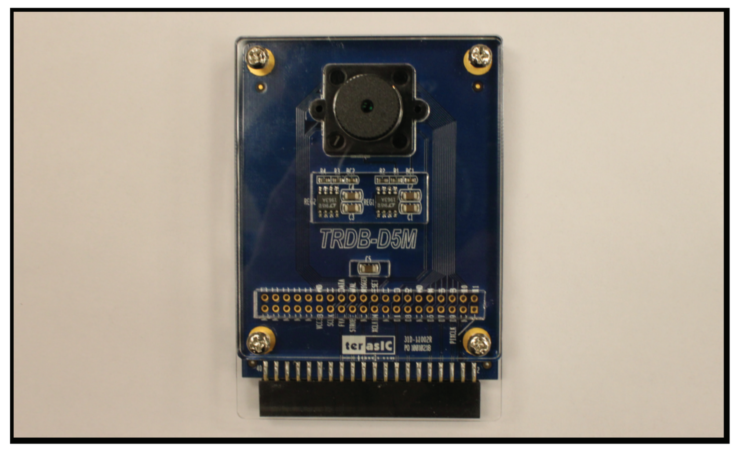
When learning to read sheet music, many students have difficulty when they do not know how the music should sound. The Discotective, an embedded music recognition system, presents a solution to this problem through image processing. A user simply takes a picture of sheet music, and after a series of processing steps, the Discotective plays the song.

HARDWARE

The Discotective is built on an Altera DE2 FPGA. A 50 MHz softcore processor has been downloaded to the FPGA, making the system compatible with C programs. In addition, a 5 megapixel camera from Terasic captures images, which are used for processing and can be displayed on a VGA monitor. The final audio is output to speakers via the board's audio codec.

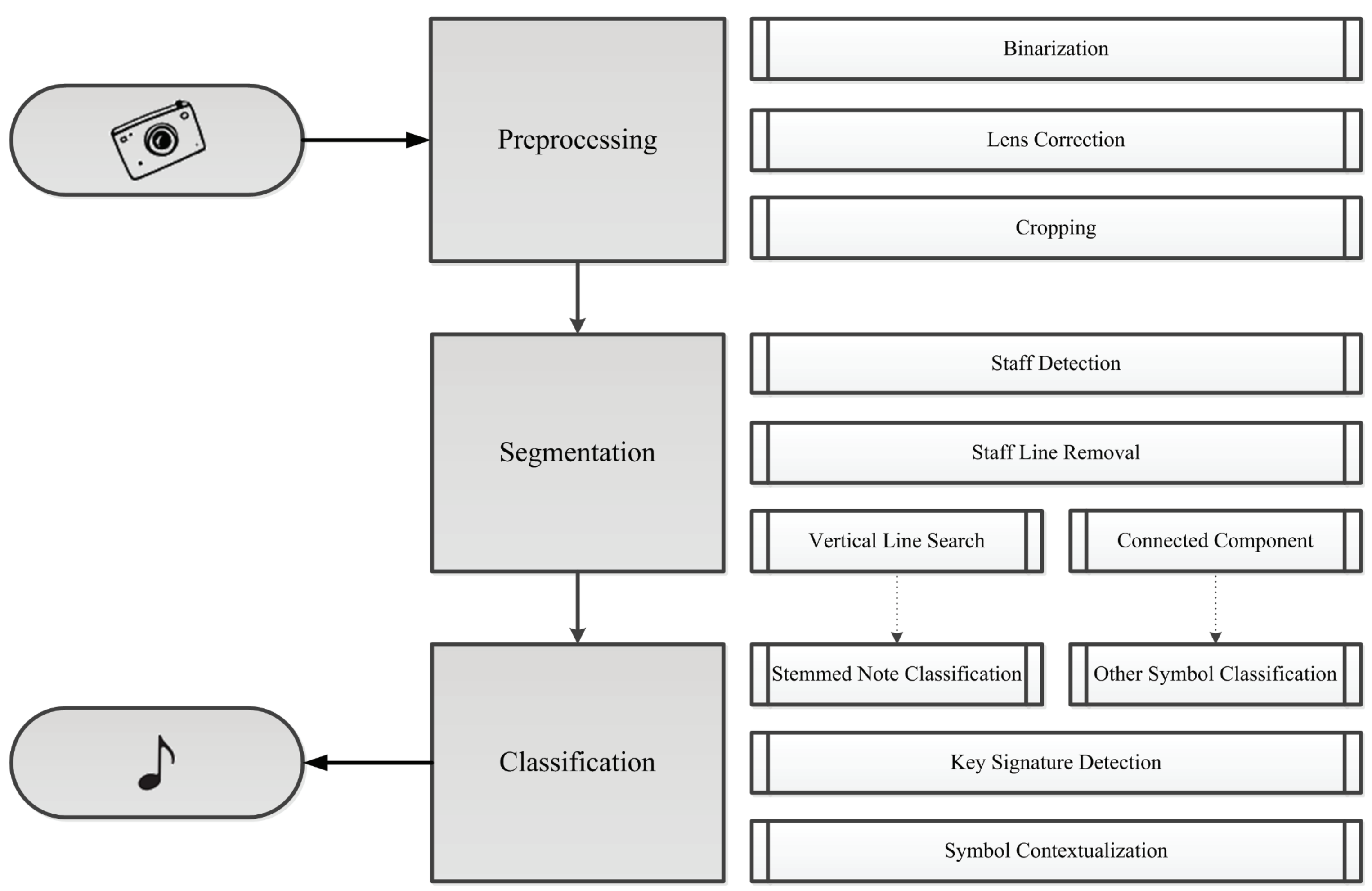


Altera DE2 FPGA



Terasic 5 megapixel camera

SYSTEM OVERVIEW



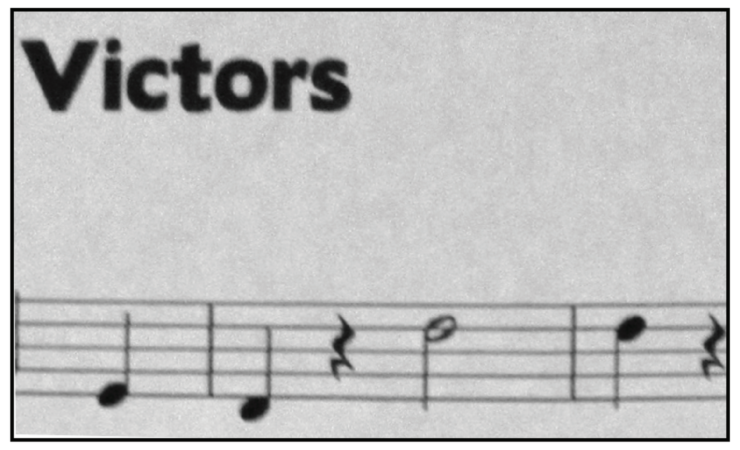
DESIGN DETAILS

I. Image Acquisition

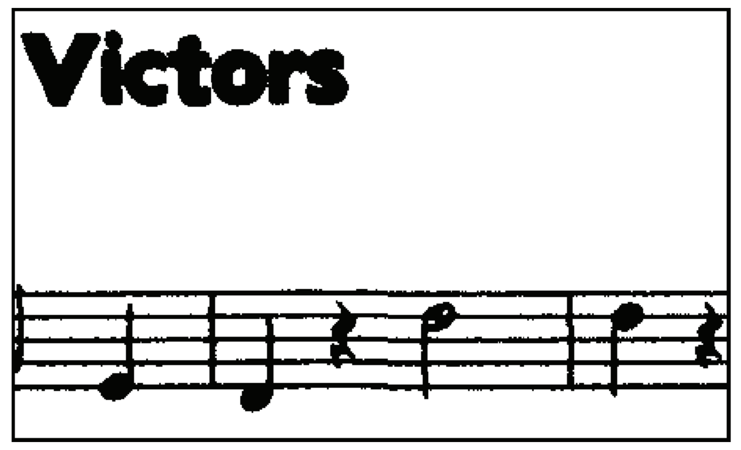
The Discotective first captures a 5 megapixel image, which is displayed on the VGA in real-time.

II. Preprocessing

Next the image is prepared for further processing. To ease memory and computational demands, the image is binarized. A simple transformation is also applied to correct for lens distortions. Because of the embedded system's limited capabilities, more robust MATLAB prototype algorithms have been streamlined to work well on the FPGA.



Original grayscale image

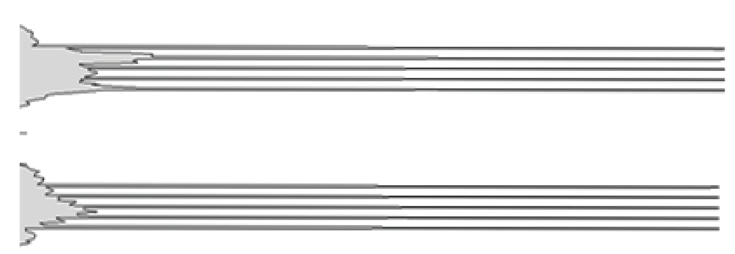


Preprocessed image

III. Segmentation

After preprocessing, a top-down approach is used to isolate smaller symbols. Image projections are computed to find horizontal lines and detect individual staffs. Afterward, stafflines are removed, followed by stemmed notes, leaving an image of unconnected musical symbols. These symbols are then segmented using connected-component analysis.

Unsegmented staffs



Horizontal projection of image

DESIGN DETAILS

IV. Classification

The Discotective takes a tree-based approach to classification. Stemmed notes, which are easy to detect, are identified first. Then additional symbols, such as accidentals, dots, and rests are classified by extracting certain features.

Some useful features:

- Symbol height
- Symbol width
- Presence of vertical lines
- Proximity to notes
- Proximity to stafflines
- Black to white pixel ratio

V. Output

The Discotective produces sound output using direct digital synthesis. In essence, a sinusoidal look-up table is used to produce pure tones. The frequency of each tone can be tuned depending on the amount of samples used in each sinusoidal period. Additional tonal qualities can be achieved by adding harmonic frequencies or adjusting the tone's amplitude at specific times.

PERFORMANCE

Overall, the Discotective performs well with simple music, such as that played by a middle school band. Due to the limits of our design, however, many symbols are difficult to classify. Slurs or ties connected to notes, for example, cannot be segmented properly with connected-component analysis. More complicated symbols, such as chords, repeat signs, accents, or text will not be properly recognized by the classifier.

The hardware implementation constraint introduces further limitations. The slow softcore processor prohibits the use of more advanced binarization and deskew preprocessing steps, which reduce the accuracy of later algorithms. Nevertheless, the device still performs well under proper conditions.

CONTACT

Group members may be reached at discotective@umich.edu.