Machine Vision and Applications (1992) 5:223–230

Machine Vision
and Applications

Segmentation and Preliminary Recognition of Madrigals Notated in White Mensural Notation*

Nicholas P. Carter
Departments of Physics and Music, University of Surrey, Guildford, Surrey GU2 5XH United Kingdom

Abstract: An automatic music score-reading system will facilitate applications including computer-based editing of new editions, production of databases for musicological research, and creation of braille or large-format scores for the blind or partially-sighted. The work described here deals specifically with initial processing of images containing early seventeenth century madrigals notated in white mensural notation. The problems of segmentation involved in isolating the musical symbols from the word-underlay and decorative graphics are compounded by the poor quality of the originals which present a significant challenge to any recognition system. The solution described takes advantage of structural decomposition techniques based on a novel transformation of the line adjacency graph which have been developed during work on a score-reading system for conventional music notation.

Key Words: structural pattern recognition, printed music, white mensural notation

1 Introduction

The availability of an automatic music score-reading system will facilitate applications including computer-based editing of new editions, production of databases for musicological research, and creation of braille or large-format scores for the blind or partially sighted. The availability of electronic databases of music representational language encodings will also hasten the development of point-of-sale printing systems for sheet music and open up possibilities of electronic distribution. For background information regarding the field of acquisition, representation and reconstruction of printed music by computer, the reader is referred to Carter et al. (1988) and Hewlett and Selfridge-Field (1991).

Automatic recognition of printed music has been the subject of research since the late 1960s (when hardware limitations restricted progress—see Pruslin 1967 and Prerau 1970). More recent work at Waseda University on the WABOT-2 keyboard-playing robot used mask-matching implemented in hardware to read nursery song sheets (Matsushima 1985). A team at Osaka University aims to produce an overall "music information processing system" which attempts automatic transcription (producing a score from a soundtrack) and sentiment identification, in addition to recognition of piano scores (Katayose 1989). Other work in score recognition is being undertaken at the University of Ottawa, using projection profiles (Fujinaga 1988) and at University College Cardiff, using simple, localised measurements aimed at producing a low-cost solution to the problem of automatic acquisition for publishers and engravers (Clarke 1988).

The work described here moves beyond the scope of conventional music notation and deals with one particular form of early notation, i.e., seventeenth century white mensural notation. An overview of the note and rest symbols used in white mensural notation, together with their modern equivalents, can be found in Figure 1. In this context the advantages of an automatic score-reading system are supplemented by the facility to convert such obsolete notation into conventional notation ready for publication. Related research into recognition of an even earlier form of notation has been described by McGee and Merkley (1991) in discussing their work with medieval music.

Address offprint requests to: Nicholas P. Carter, Departments of Physics and Music, University of Surrey, Guildford, Surrey GU2 5XH, United Kingdom.

*This research was undertaken with support from Oxford University Press.

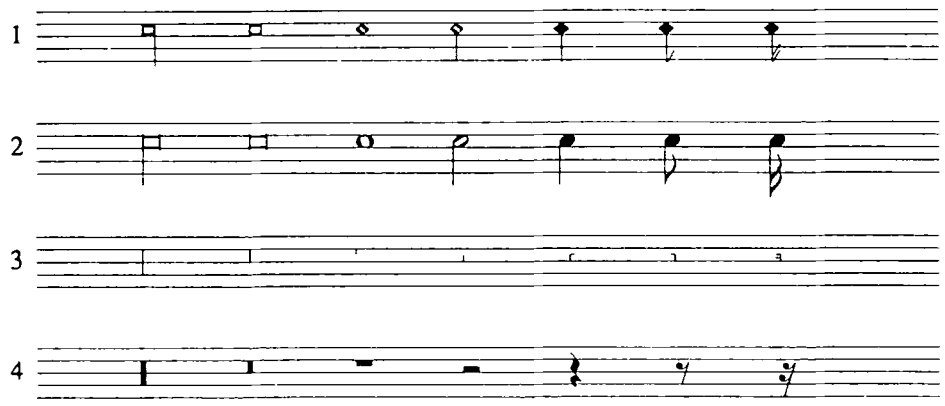

Figure 1. An overview of the note and rest symbols used in white mensural notation (staves 1 and 3), together with their modern equivalents (staves 2 and 4).

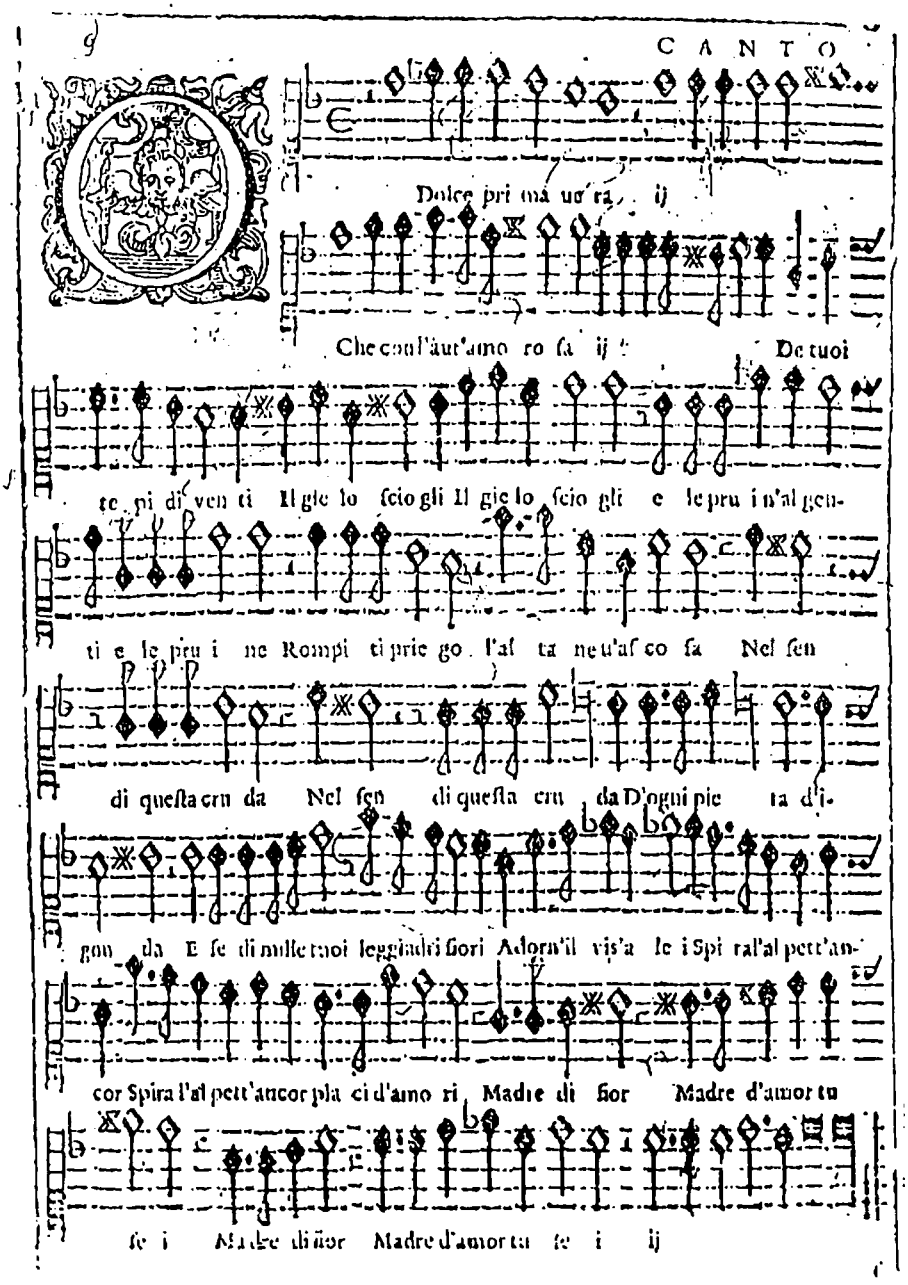

Figure 2. An example of seventeenth century white mensural notation.

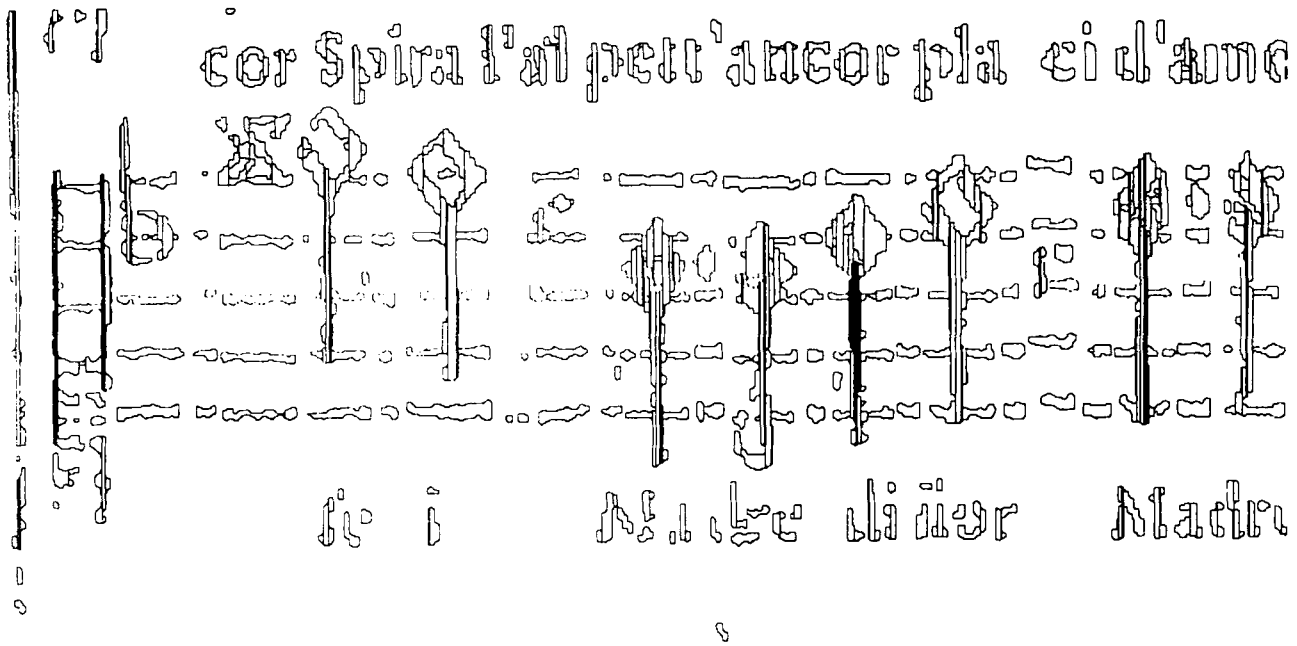

Figure 3. Outlines of the sections which are the nodes of the transformed line adjacency graph for the bottom left-hand corner of Figure 2.

2 Image Acquisition and Structural Decomposition

Binary images are acquired using a conventional flatbed CCD-based scanner at a resolution of 300 dots per inch. Some of the images had previously been transmitted by fax, so sampling had already taken place at a lower resolution and some additional degradation occurred (Figures 2 and 7). Figure 9 is a more typical example of an image, from a different publication to those of the other figures, with less noise and degradation.

The first operation on each image involves the production of an original transformation of the line adjacency graph (LAG), the details of which have been described elsewhere (Carter 1989; Carter and Bacon 1992), in connection with work on a recognition system for conventional music notation. Subsequent processing is based on manipulation of the nodes of the transformed LAG (termed "sections") which correspond to particular regions of adjacent black pixels (Figure 3). This avoids further operations at the pixel level, thus speeding processing, and also provides a first approximation to the structurally significant portions of the image, such as horizontal line fragments. By making use of attributes of the sections such as area, aspect ratio, and average thickness, significant tolerance of image distortions is built into the system. For example, in order to reduce the noise present in the image, all sections with area $\leq$ five pixels are removed. In order to achieve recognition of the text when this has been isolated from the original image, some of these so called "noise" sections may have to be restored, as they are in fact structurally important features of the constituent letters.

After removal of noise sections, groups of connected sections are formed into "objects" by a depth-first traversal of the transformed line-adjacency graph. Due to the method of production used in creating the original seventeenth century scores, where each symbol would have been printed using a separate piece of type, a fragmented appearance results and as a consequence the objects formed by the above process normally consist of a single symbol. In terms of score recognition, this is the principal difference (apart from the symbol designs as shown in Figure 1) between these images and conventional music notation. In the latter case, it can normally be assumed that the stavelines will run continuously for the majority of their extent across the page, whereas with the madrigal scores the stavelines consist of a large number of fragments.

3 Isolation of Ornamental Graphic and Text

The pages of notation consist of three main elements: a number of five-line staves with overlaid musical symbols (including notes, rests, and clefs), word underlay (lyrics) and an ornamental graphic situated at the top left of each page. In order to isolate the ornamental graphic, the largest object in the top left hand quarter of the image is found. Then,

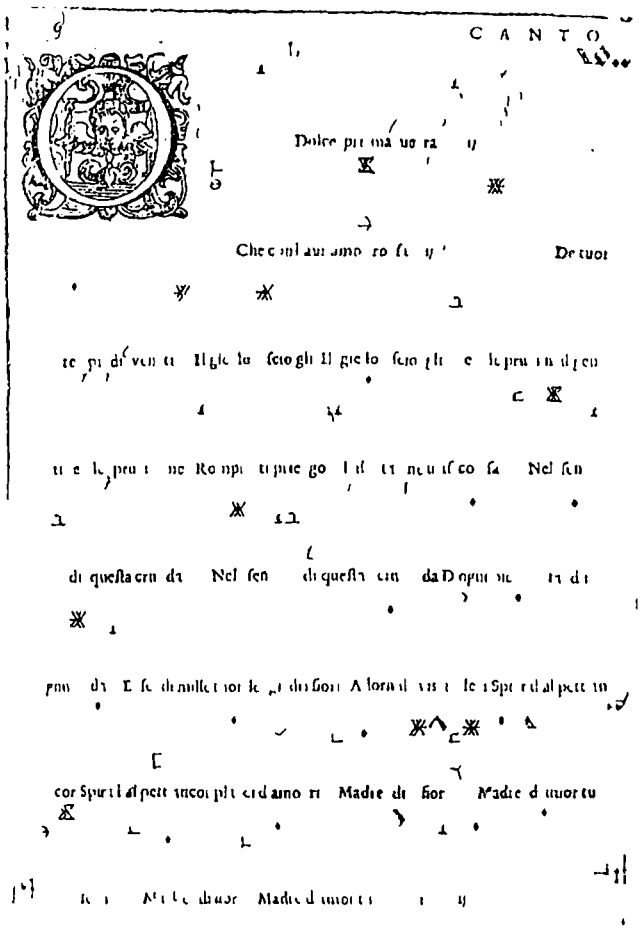

Figure 4. Potential text objects, the ornamental graphic, and some shadow lines extracted from Figure 2.

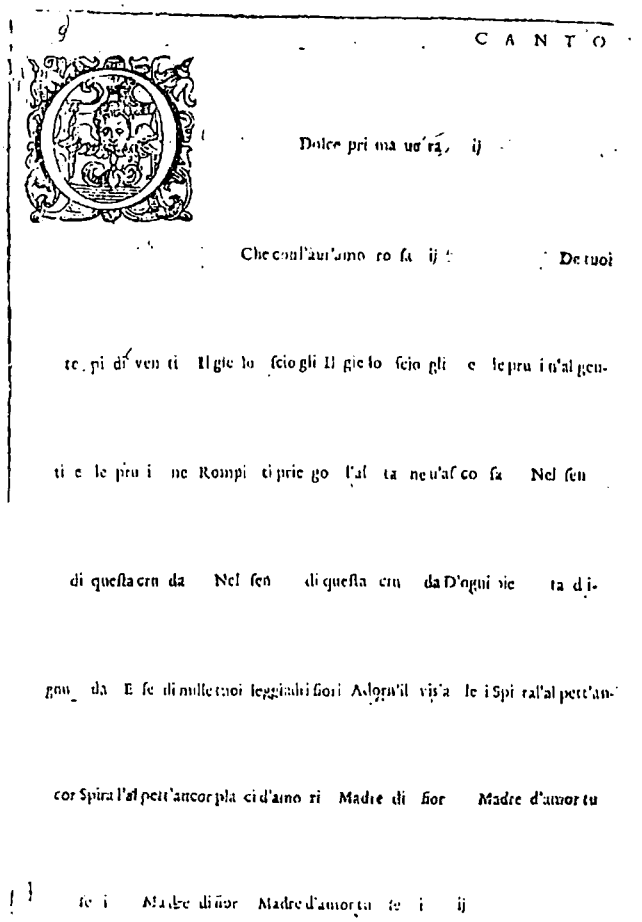

Figure 5. The final selection of text extracted from Figure 2.

both it and all other objects which overlap the region delineated by the graphic object's lowest and rightmost extremes and the top and leftmost edges of the page, are removed. This is necessary in order to include all fragments of the ornamental graphic (which is not guaranteed to be a single-connected component) and some of the shadow lines present at the borders of the image.

The next operation makes use of a maximum line thickness threshold which is set to 2.5 × the most common section thickness. This relies on the fact that most sections are horizontal line fragments which make up the stavelines. The inter-staveline spacing (subsequently referred to as α) is then calculated by finding the minimum spacing between each section with average thickness less than the maximum line thickness and a horizontally-overlapping section with similarly limited thickness. The most common spacing between these lines is taken as the inter-staveline spacing. Potential text objects are categorized as those which satisfy the following criteria:

$$\text{width} < 1.5 \times \alpha,$$
$$\text{maximum line thickness} < \text{height} < 1.5 \times \alpha$$
$$\text{and width/height} < 2.$$

The potential text objects extracted by this method from Figure 2 are shown in Figure 4. The potential text objects are then linked together into strings by testing for vertical overlap between neighbouring objects. Strings of less than four characters are removed as erroneous. Also, text strings which are vertically spaced by less than the maximum line thickness value are combined.

Objects with height of between three and ten times α and width greater than the maximum line thickness are then classfied as "significant" objects. These objects are then examined in pairs and where they overlap vertically, the extent covered by the two objects is recorded to build up a picture of stave location. Horizontal stave dividing lines are then situated where there are gaps in this vertical extent array which are wider than the maximum line thickness. The text strings are then matched to these

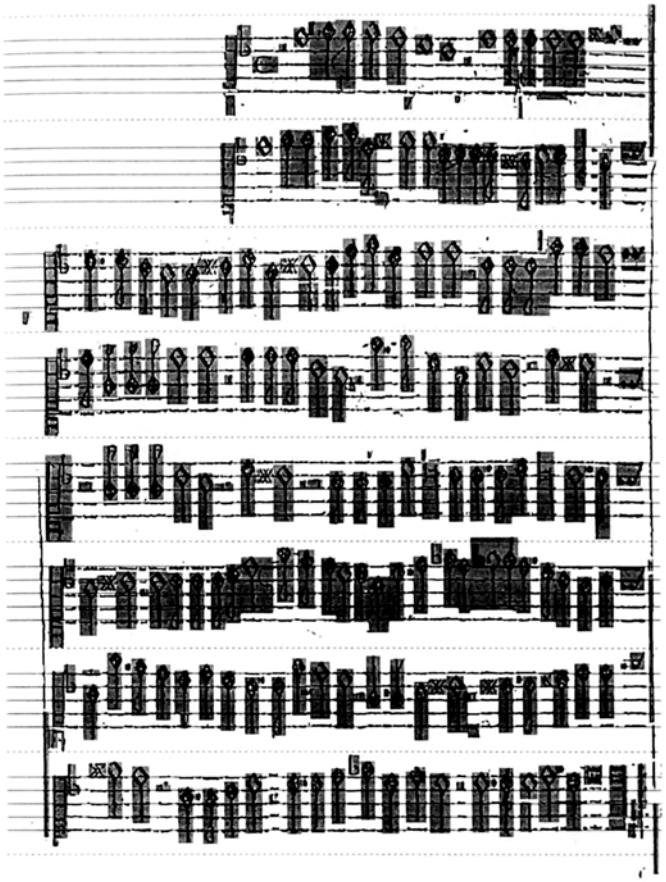

Figure 6. Processed output for Figure 2. The dotted lines are the stave separation lines; solid lines are the staveline approximations; the bounding rectangles of large and small objects are shaded with horizontal and vertical lines respectively.

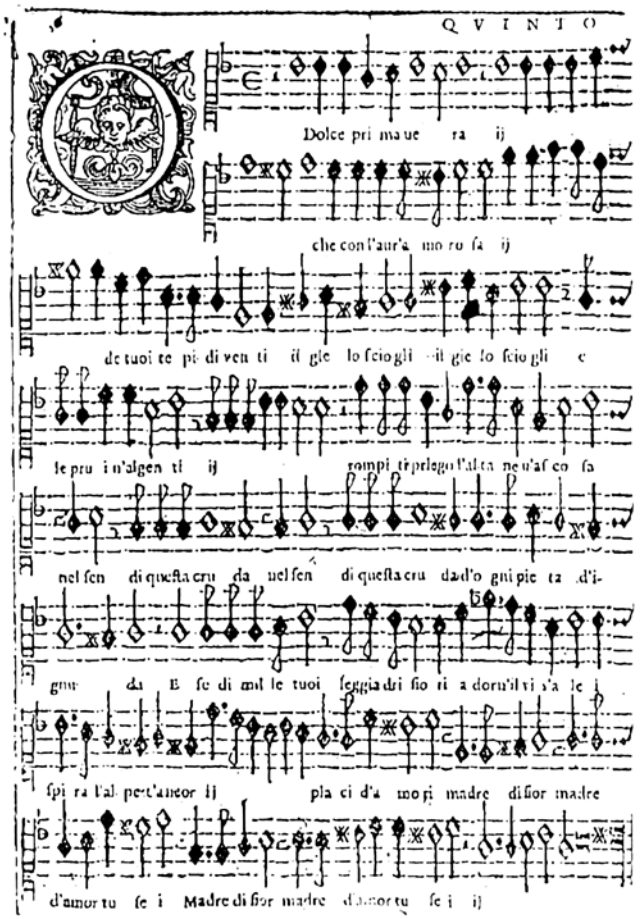

Figure 7. A further example of seventeenth century white mensural notation.

stave dividing lines and any which are left over are removed. All objects which are finally to be categorized as part of the word underlay are found by locating all objects which are of appropriate size and overlap a horizontal strip of height $1.5 \times \alpha$ centered on a stave dividing line. The results of this process can be seen in Figure 5. All remaining objects are then allocated to a particular stave depending on their vertical position relative to the stave dividing lines.

4 Location of Stavelines

Identifying the location of stavelines is made difficult by the fragmented nature of the image and also by steps in the stavelines caused by the vertical displacement of some of the original engraving blocks. Currently, a global straight line fit is performed based on the location and length of near-horizontal line fragments between each pair of stave dividing lines. The current technique needs to be improved to take account of local distortions such as the "steps" described above, and also curvature in the original (i.e., bowing of the stave). The results of fitting five horizontal lines equally spaced by the value for α can be seen in Figure 6. This information will be used when note pitches are being established, and so needs to be accurate enough to enable distinction between notes superimposed on a line or in a space.

5 Preliminary Musical Symbol Analysis

The first processing operation divides all objects associated with the current stave into 'large' (height $\geq 3 \times \alpha$) and 'small' ($0.5 \times \alpha <$ height $< 3 \times \alpha$). Any object of height less than $0.5 \times \alpha$ is ignored as noise. A small object which has height less than $0.8 \times \alpha$ and width of between 0.1 and $0.6 \times \alpha$ is marked as a possible dot. This will be subjected to syntactic

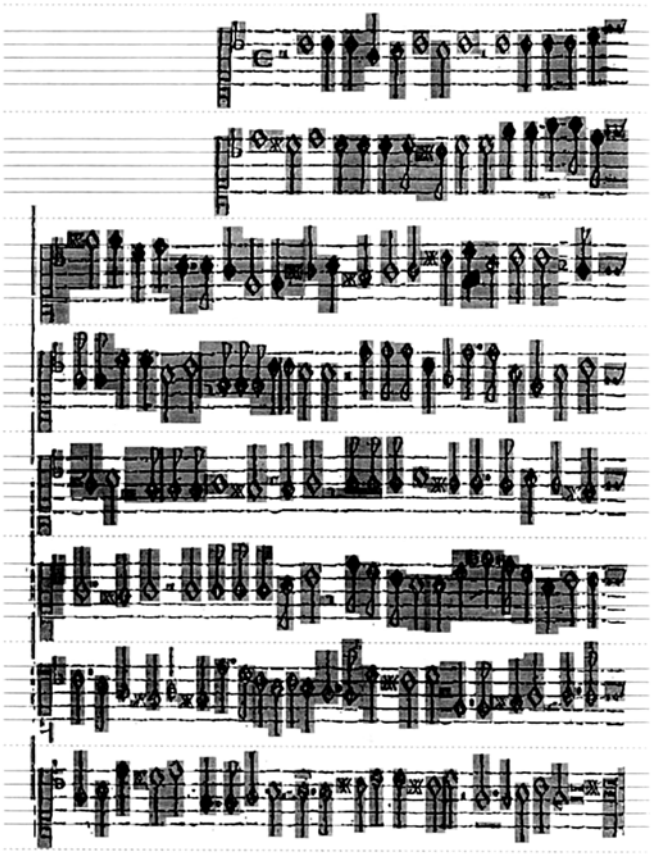

Figure 8. Processed output for Figure 7.

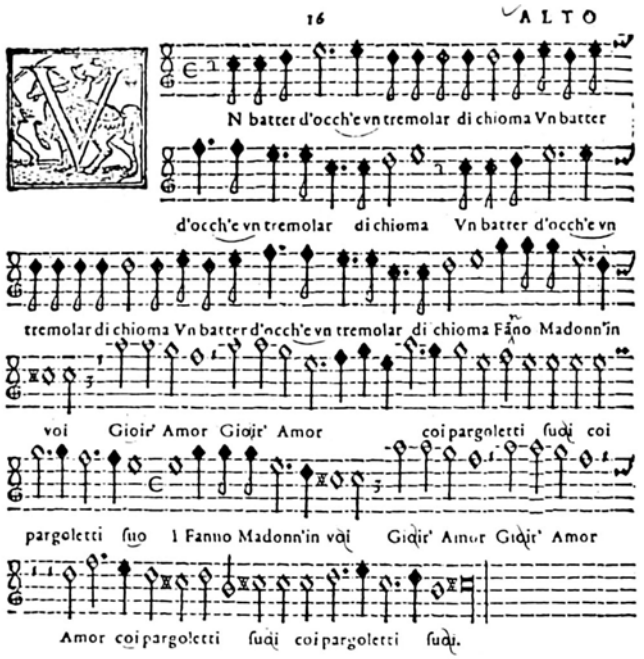

Figure 9. An image of seventeenth century white mensural notation from a different publication—perhaps more typical in terms of image quality.

tests at a later stage to confirm that it is in the appropriate position next to a notehead. All small objects are further tested for horizontal overlap with large objects using an array of horizontal extents compiled during the search for large objects. The use of the horizontal extent of objects is based on the fact that the music on each stave is basically a one-dimensional string of symbols unlike a large proportion of conventional notation which has a far more complex, two-dimensional structure.

Initially, a search is made for a clef at the left hand end of the current stave by testing for an object containing multiple vertical lines and width $> 0.5 \times \alpha$. This is done by producing the transformed line adjacency graph of the object but in this case based on horizontally-oriented run-length encoding. Those sections which have aspect ratio greater than two and length greater than or equal to α are identified as line fragments. All large and small objects which are to the left of the clef are then removed. This will typically include shadow lines at the edge of the page, separated fragments of the clef itself and noise.

The next stage of analysis takes advantage of the guaranteed presence at the end of each line of a subsequent pitch indicator (shaped line a ~) or, in the case of the lowest line on the page, a long (a note with a hollow, square notehead and a stem). The search for the former is undertaken by examining the list of small objects in reverse order, i.e., right to left, for one of width greater than $0.5 \times \alpha$. If this object is less than $1.5 \times \alpha$ in width, a further search is undertaken within the proximity of the object to attempt to locate other fragments which have become detached. Similarly, the long is found on the lowest of the staves by searching backwards through the list of large objects for an object of width greater than or equal to the interstaveline spacing. All large and small objects to the right of the long can then be removed as these can only be barlines or shadow lines. A double barline will be assumed to be present and therefore added by default to the final output data.

All large objects are examined next, with the eventual aim of distinguishing natural signs from stemmed notes and extracting a pitch and rhythm value for each of the latter. At present only some of the necessary tests are implemented, for example, for measuring the "whiteness" of a note head. This particular test will be used to differentiate hollow noteheads. Firstly, vertical line fragments are found (again using horizontally oriented run length encoding) and subjected to a collinearity test in order to locate the potential note stem. Then the white run lengths which cross a vertical line prolonging the note stem are summed to give a figure for "whiteness." This area will be thresholded in order to make the distinction between the two types of note head.

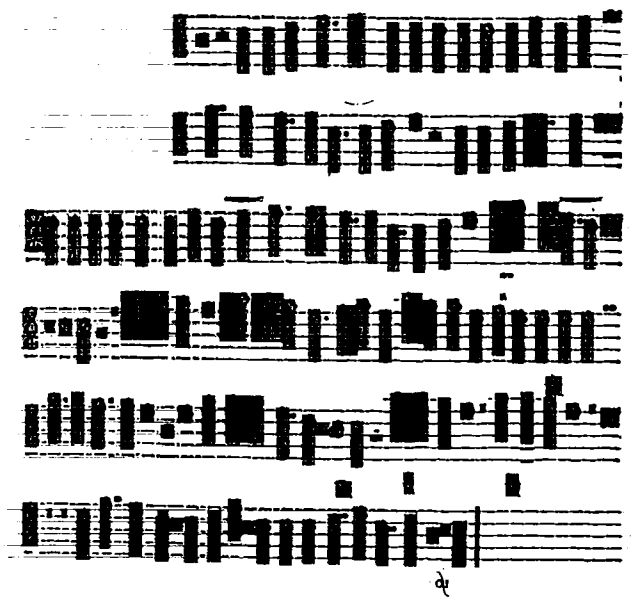

Figure 10. Processed output for Figure 9.

The complete set of small objects on each stave is then examined in adjacent pairs for horizontal overlap. If two small objects are closely spaced vertically and, when combined, would form an object of height greater than $3 \times \alpha$, then they are classified as a fragmented large object. In this case the two small objects would be combined and treated as a single large object. Otherwise, either the two small objects are closely spaced vertically, in which case they are combined into one small object, or one of them must be noise.

A vertical profile is then formed for each small object and where a peak exists which is less than half α in width and greater than half α in height, the object is deemed to be a rest. Additionally, if the height of the peak is between 0.75 and $1.25 \times \alpha$, the rest is categorized as a breve rest (a vertical stroke joining adjacent stavelines).

6 Conclusions

Although the original images for this work were of poor quality, progress has been made in isolating the different categories of symbol present. The word underlay and ornamental graphic have been isolated successfully and some of the necessary tests for categorizing the musical symbols have been implemented (Figures 7–10). Use has been made where possible of syntactic information, such as the guaranteed presence of a clef symbol at the beginning of each stave.

Successful use has been made of the low-level processing operations which have been developed during work on a system for recognition of conventional music notation, including construction of a novel transformation of the line adjacency graph to provide a structural breakdown of the image and object formation for symbol recognition purposes. A variety of thresholds are used which, although their exact values are not crucial, represent domain-specfic knowledge hard-wired into the system. It would be preferable as part of a complete recognition system to separate out this knowledge so that information relating to a different domain, such as conventional music notation or perhaps even circuit diagrams, could be substituted as required.

The main problem revolves around the choice of features for recognition, whether structural or otherwise, in order to produce a reliable system despite the variability of the musical symbols. So far the solution has been found in a mixture of techniques, which also take advantage of the relative simplicity of the musical content of the images.

References

Carter NP, Bacon RA, Messenger T (1988) Acquisition, representation and reconstruction of printed music by computer: A review. Computers and the Humanities 22(2):27–46

Carter NP (1989) Automatic recognition of printed music in the context of electronic publishing. PhD thesis, University of Surrey

Carter NP, Bacon RA (1992) Automatic recognition of printed music. In: Structured document image analysis. Baird HS, Bunke H, Yamamoto K (eds) Springer-Verlag, Heidelberg

Clarke AT (1988) Inexpensive optical character recognition of music notation: A new alternative for publishers. Proceedings of the Computers in Music Conference, Lancaster

Fujinaga I (1988) Optical music recognition using projections. MA dissertation, McGill University, Montreal

Hewlett WB, Selfridge-Field E (1991) Computing in musicology: A directory of research. Center for Computer Assisted Research in the Humanities, Menlo Park, CA

Katayose H, Inokuchi S (1989) The Kansei music system. Computer Music Journal 13(4):72–77

Matsushima T et al. (1985) Automated recognition system for musical score. Bulletin of Science and Engineering Research Laboratory, Waseda University 112:25–52

McGee W, Merkley P (1991) The optical scanning of medieval music. Computers and the Humanities 25:47–53

Prerau DS (1970) Computer pattern recognition of standard engraved music notation. PhD dissertation, Massachusetts Institute of Technology

Pruslin DH (1967) Automatic recognition of sheet music. ScD dissertation, Massachusetts Institute of Technology