Automatic Music Score Recognition

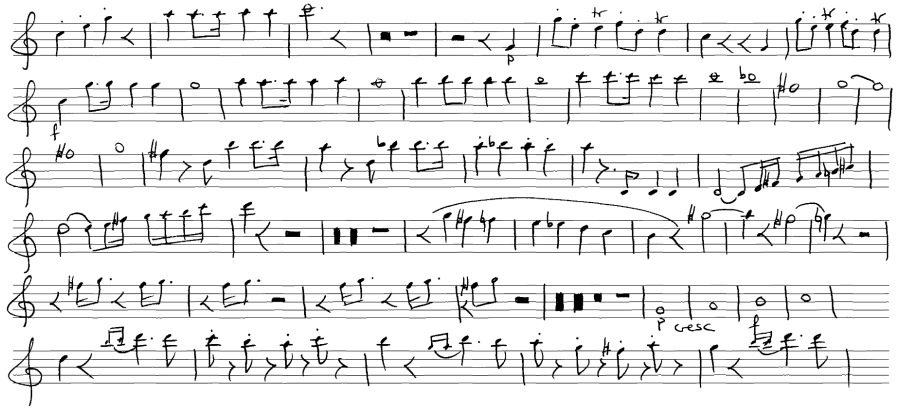
Oral Defense

HW Lee, 07.11.2016

Acoustic and **H**earing **G**roup

Optical Music Recognition

- Score in Western Music
- Image to Symbolic Representation (MusicXML)

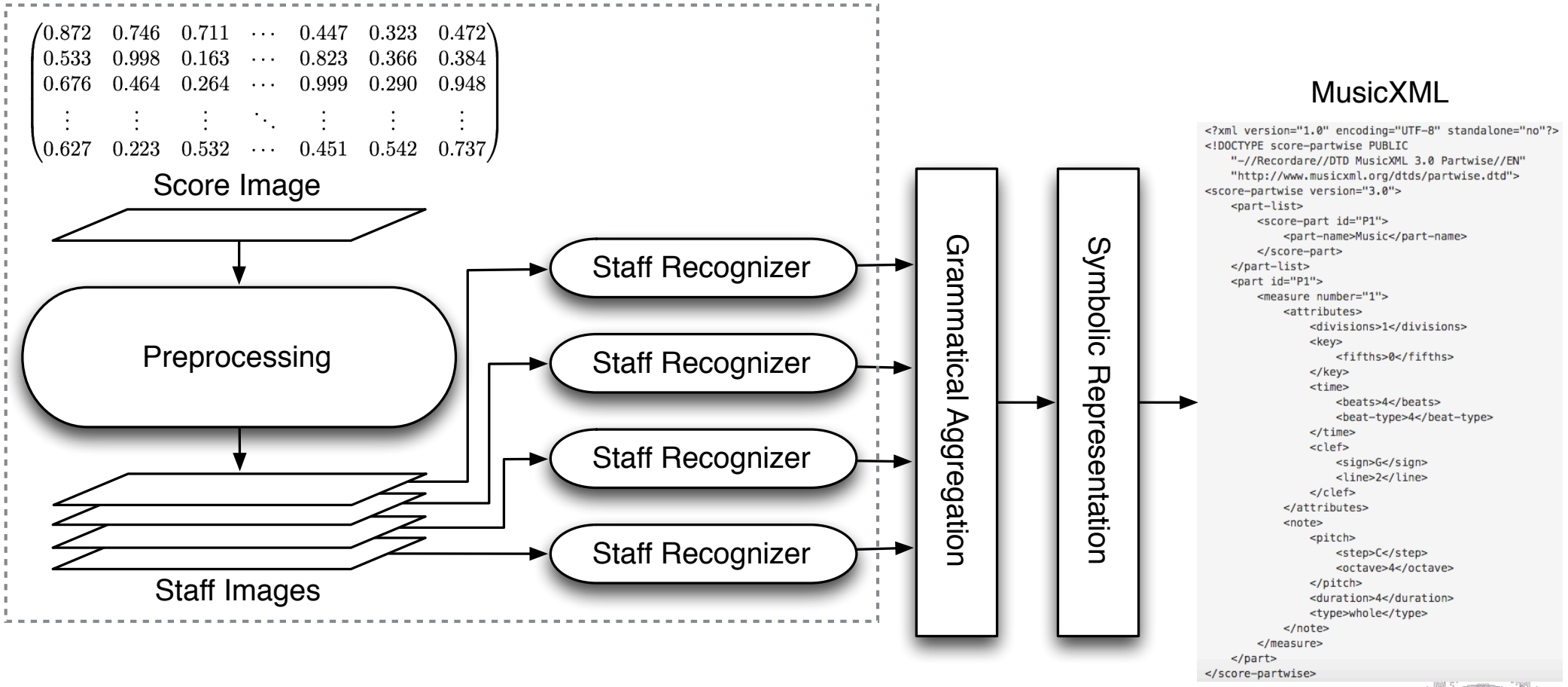


For human

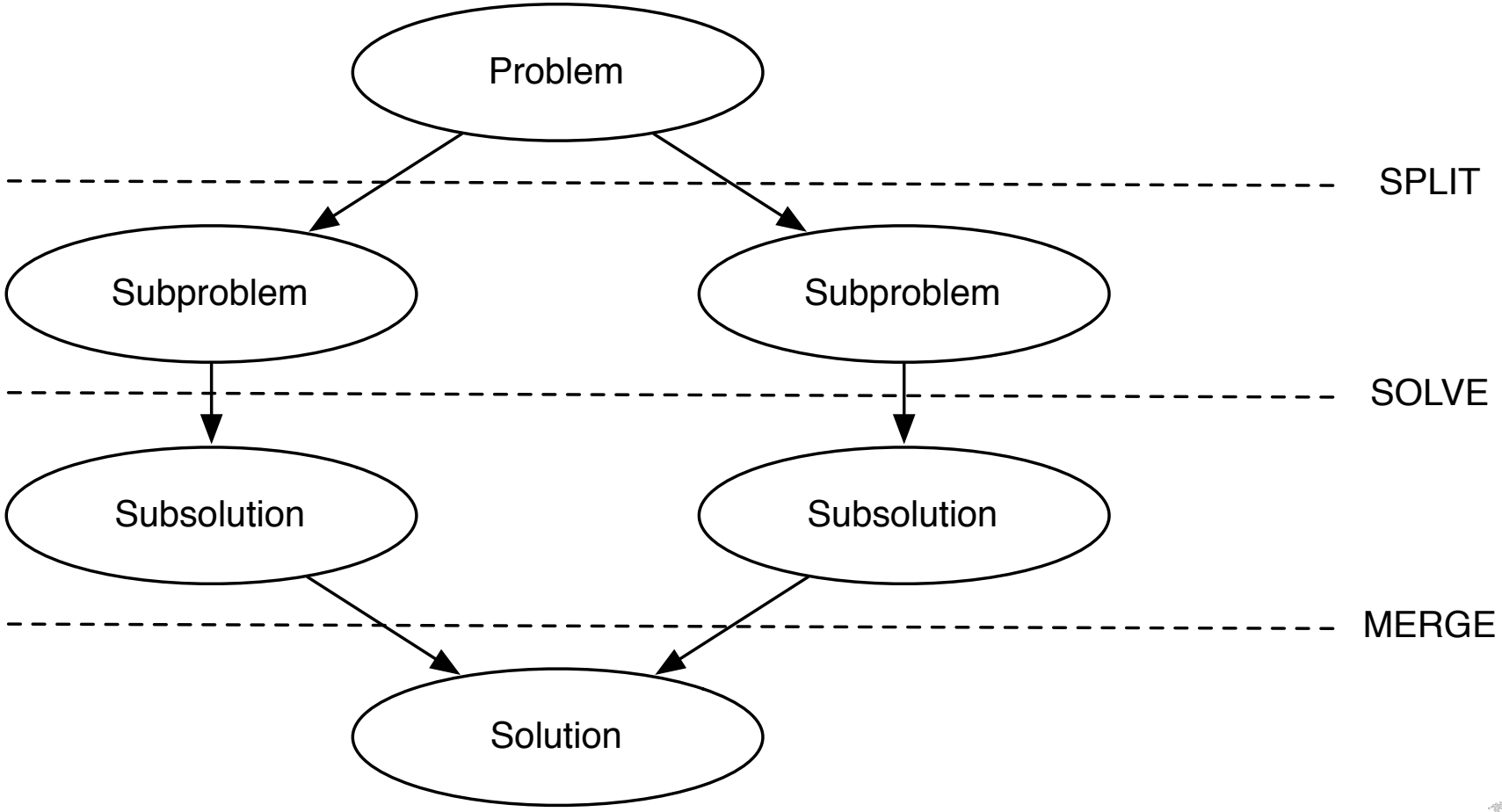
$$\begin{pmatrix} 0.872 & 0.746 & 0.711 & \cdots & 0.447 & 0.323 & 0.472 \\ 0.533 & 0.998 & 0.163 & \cdots & 0.823 & 0.366 & 0.384 \\ 0.676 & 0.464 & 0.264 & \cdots & 0.999 & 0.290 & 0.948 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0.627 & 0.223 & 0.532 & \cdots & 0.451 & 0.542 & 0.737 \end{pmatrix}$$

For computer

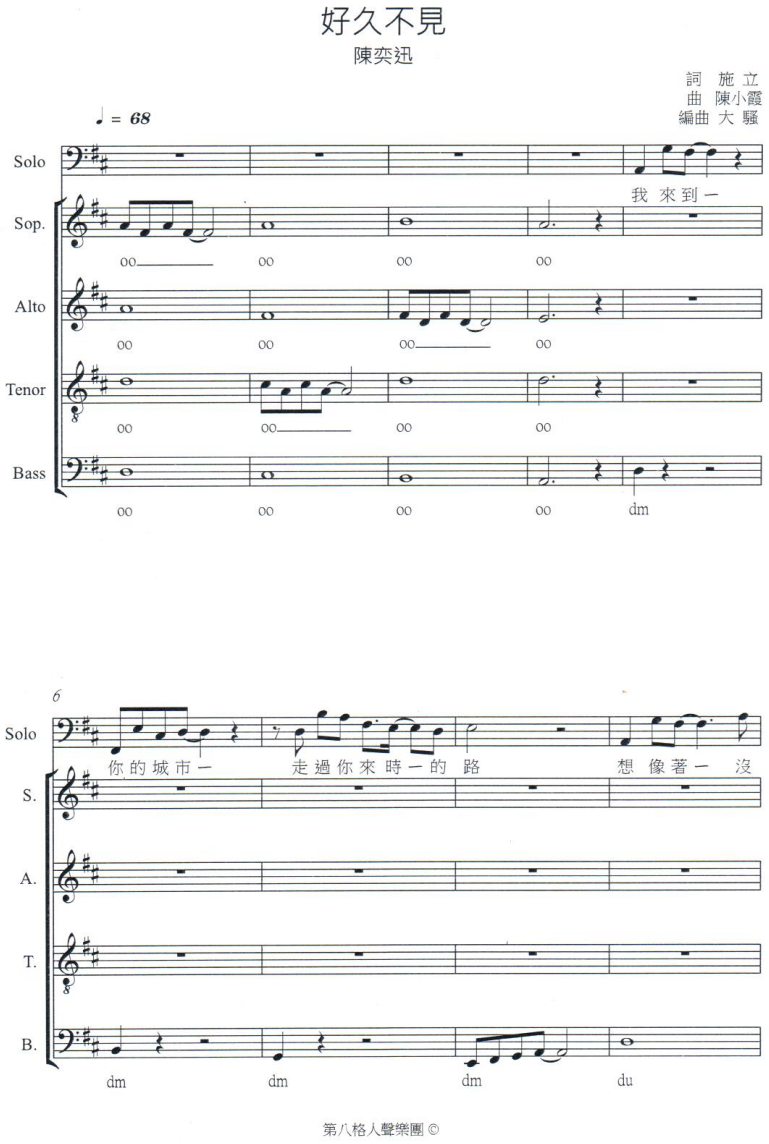
A Typical OMR System

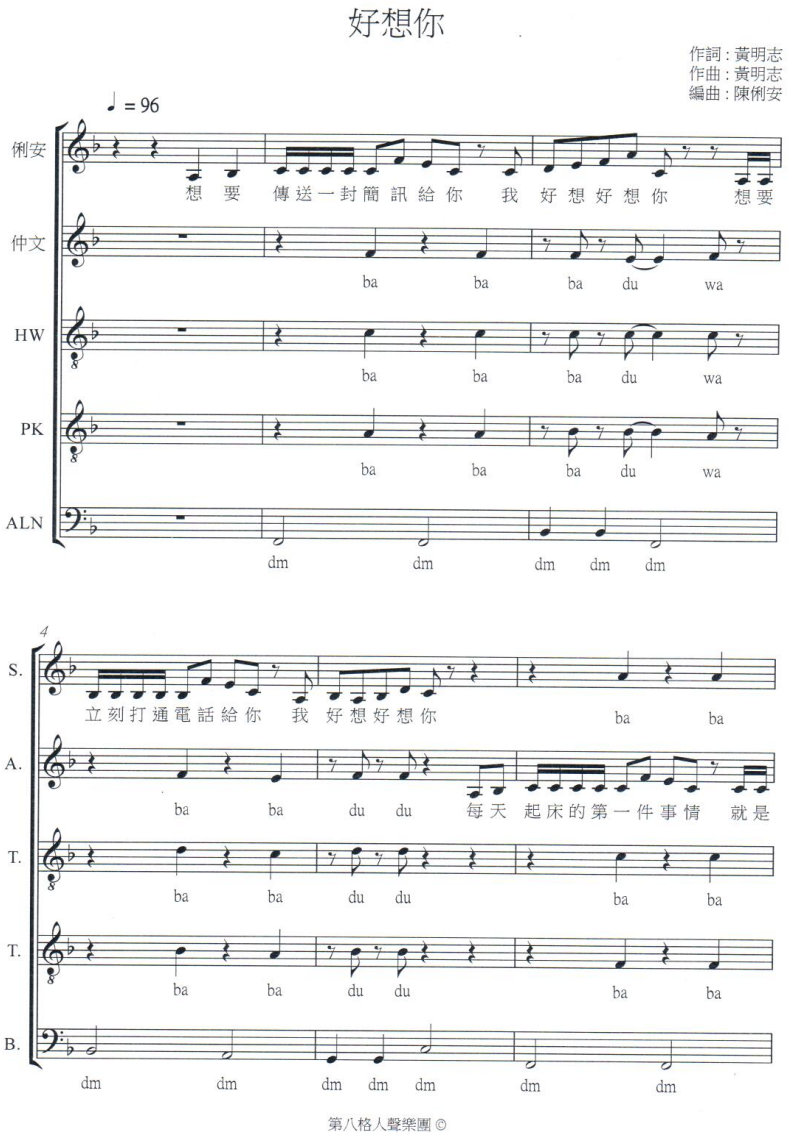


Divide & Conquer



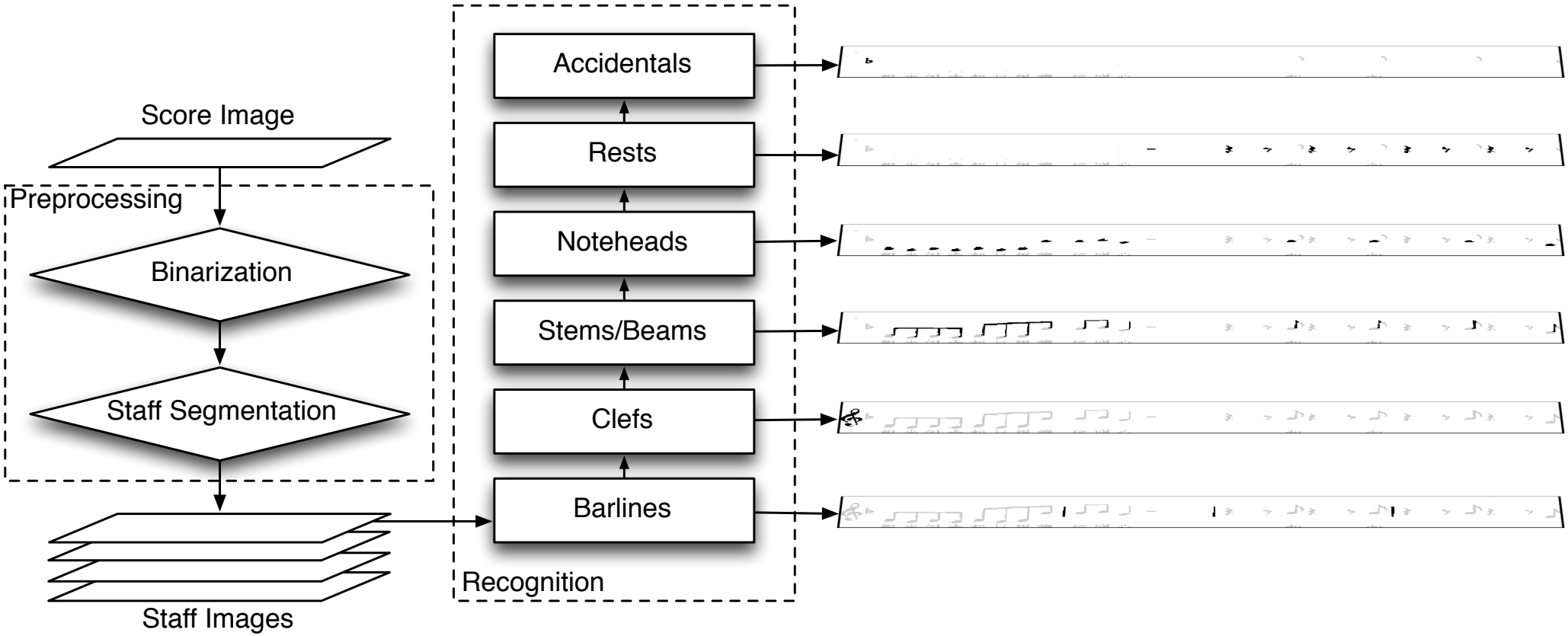
Score Sample





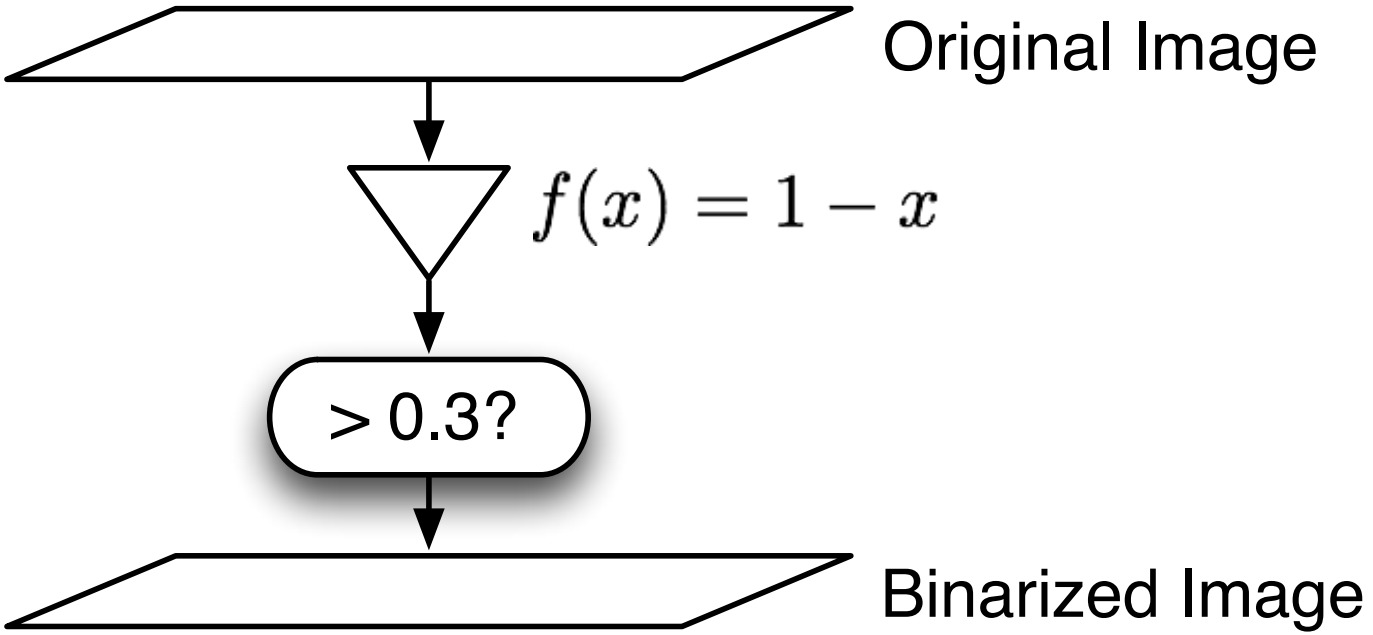
scanned scores
monophonic

System Overview



Preprocessing
Binarization

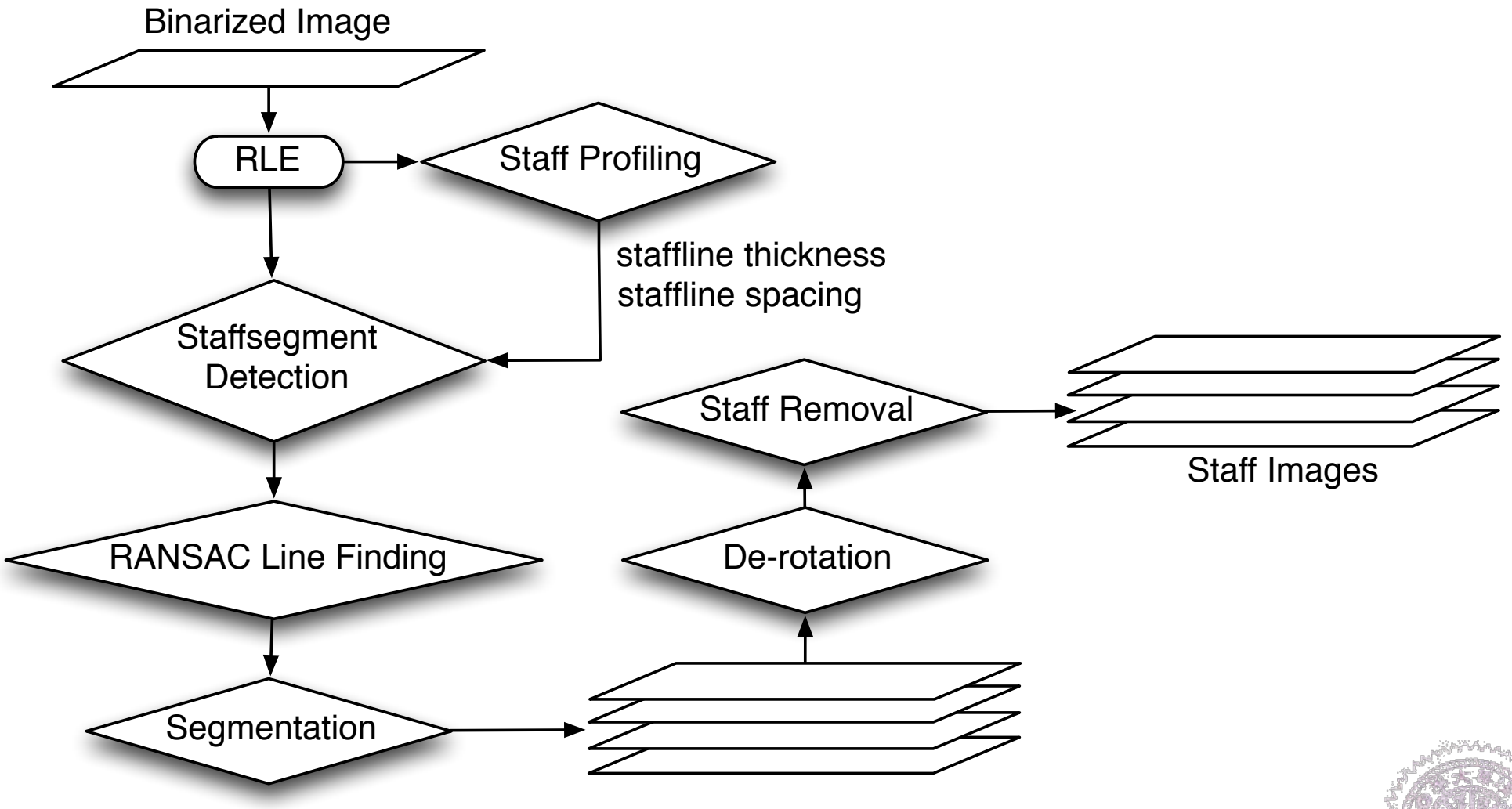
Binarization



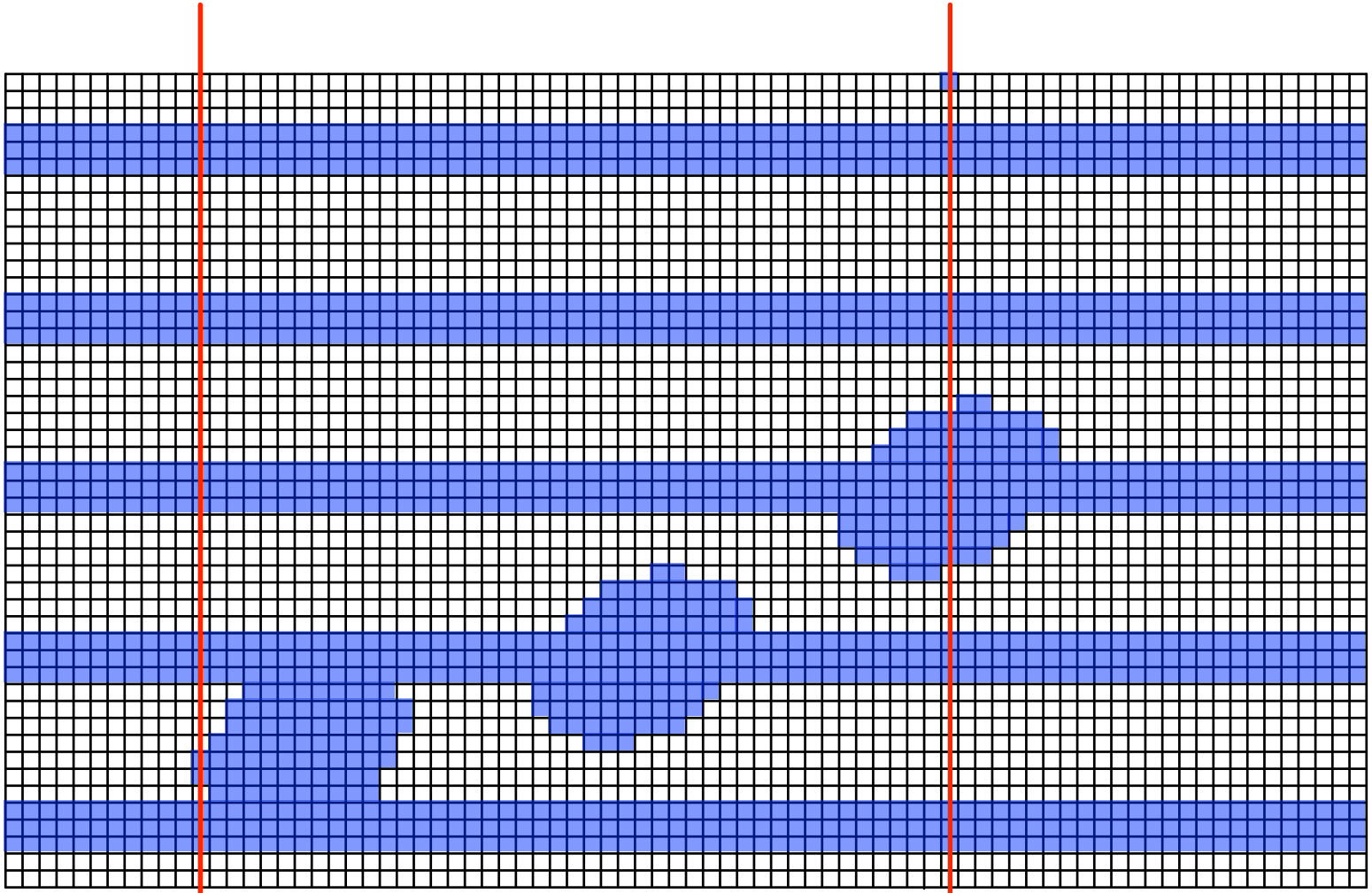
Global Thresholding

Preprocessing
Staff Segmentation

Staff Segmentation



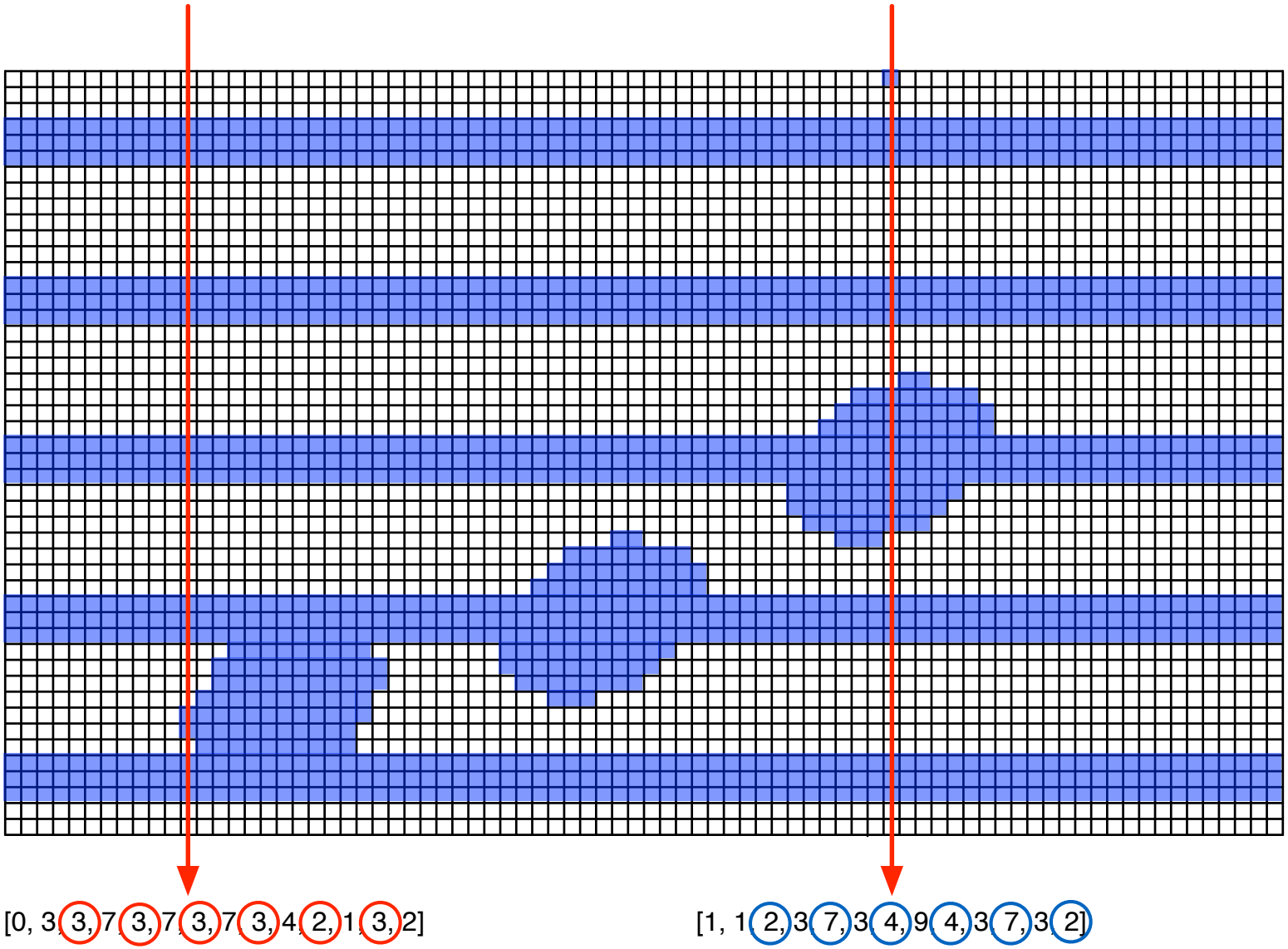
RLE (Run-length Encoding)



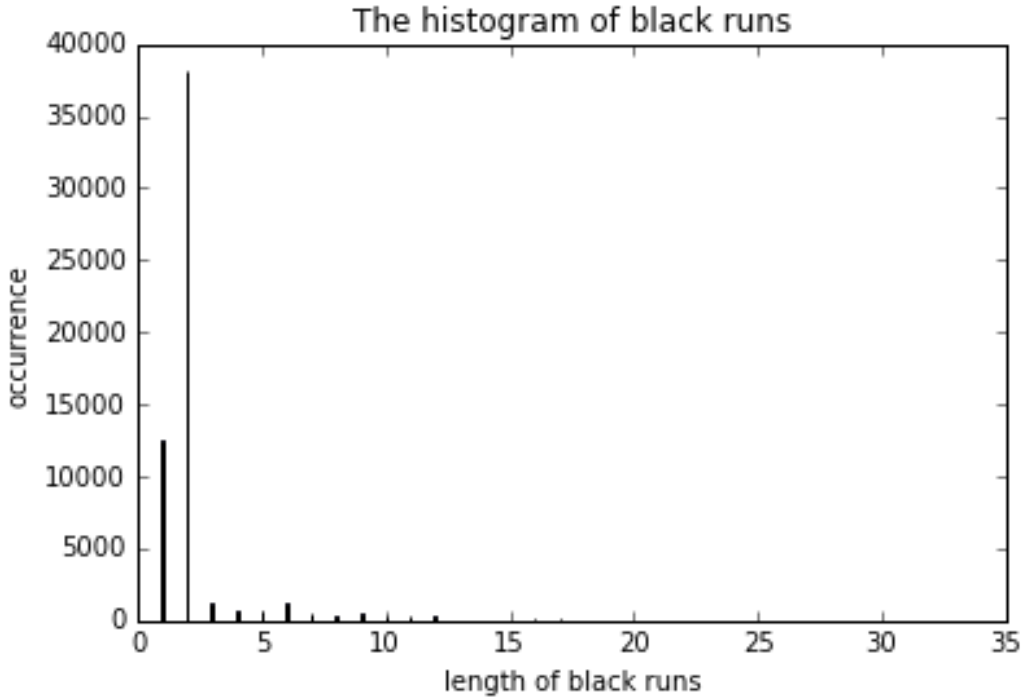
[0, 3, 3, 7, 3, 7, 3, 7, 3, 4, 2, 1, 3, 2]

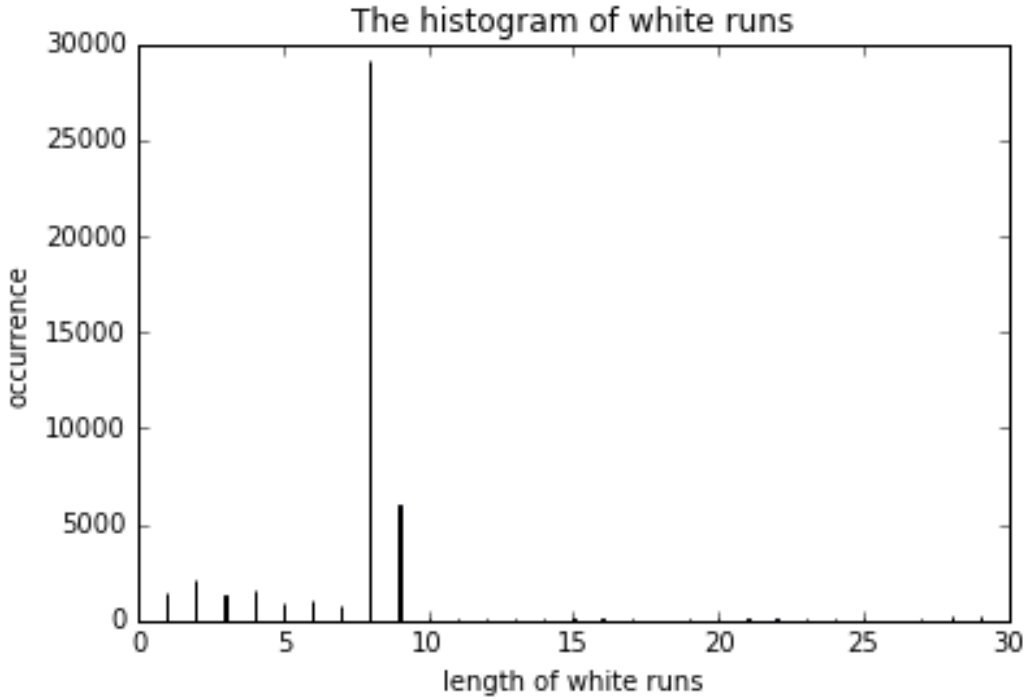
[1, 1, 2, 3, 7, 3, 4, 9, 4, 3, 7, 3, 2]

Staff Profile

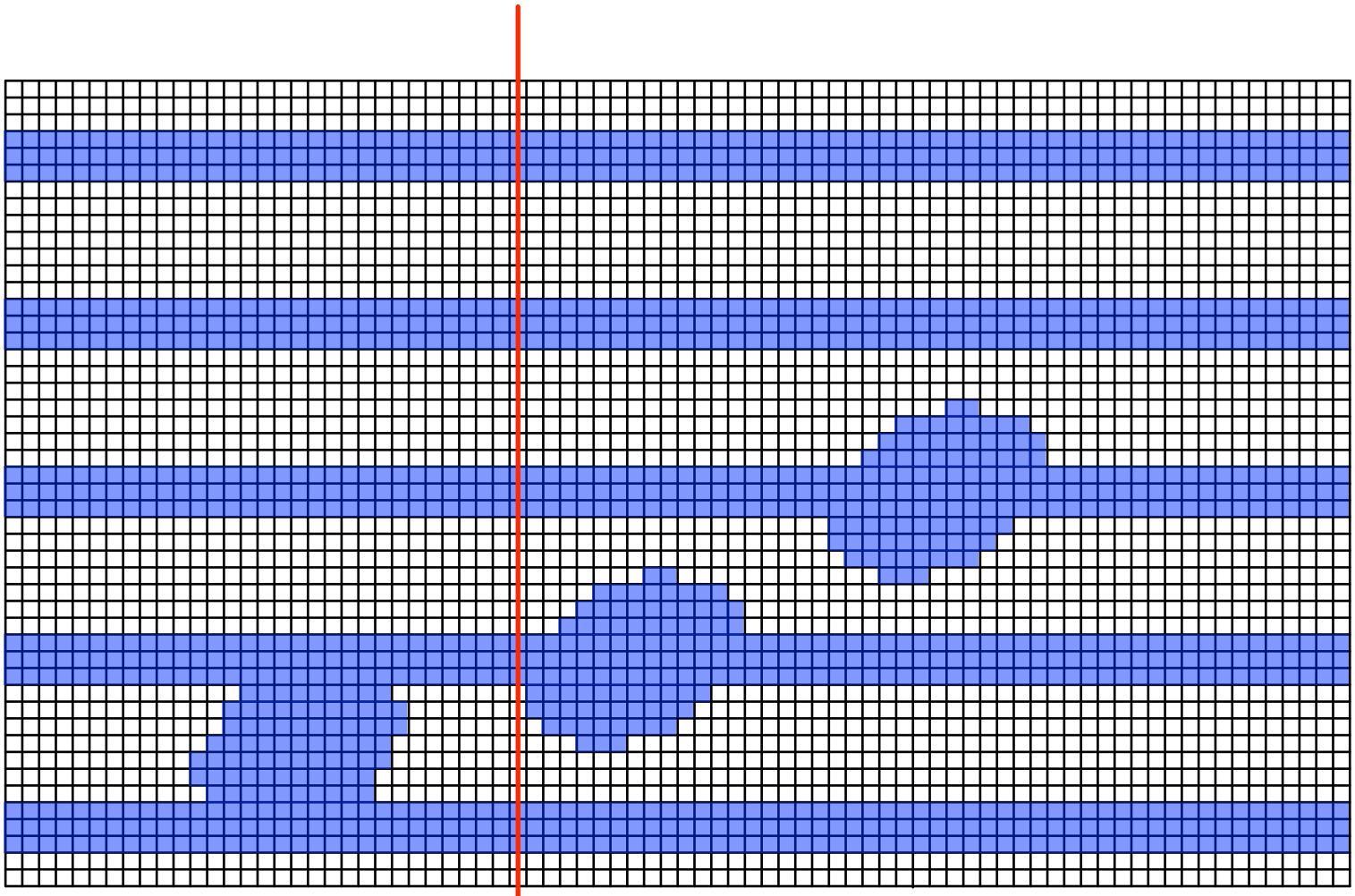


Staff Profile





Staffsegment Detection



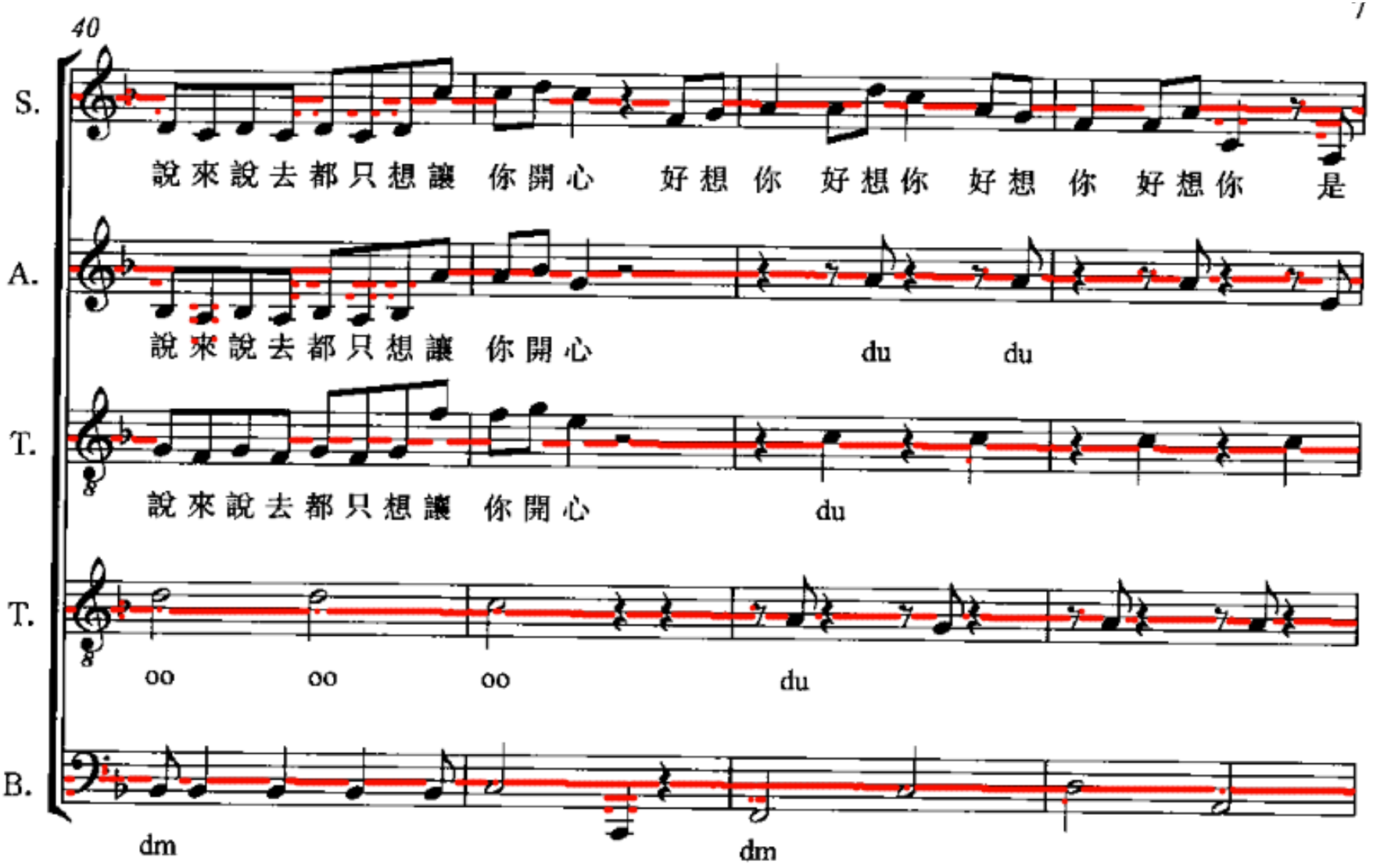
3, 3, 7, 3, 7, 3, 7, 3, 7, 3, 2

3, 3, 7, 3, 7, 3, 7, 3, 7	NO
3, 7, 3, 7, 3, 7, 3, 7, 3	YES
7, 3, 7, 3, 7, 3, 7, 3, 2	NO

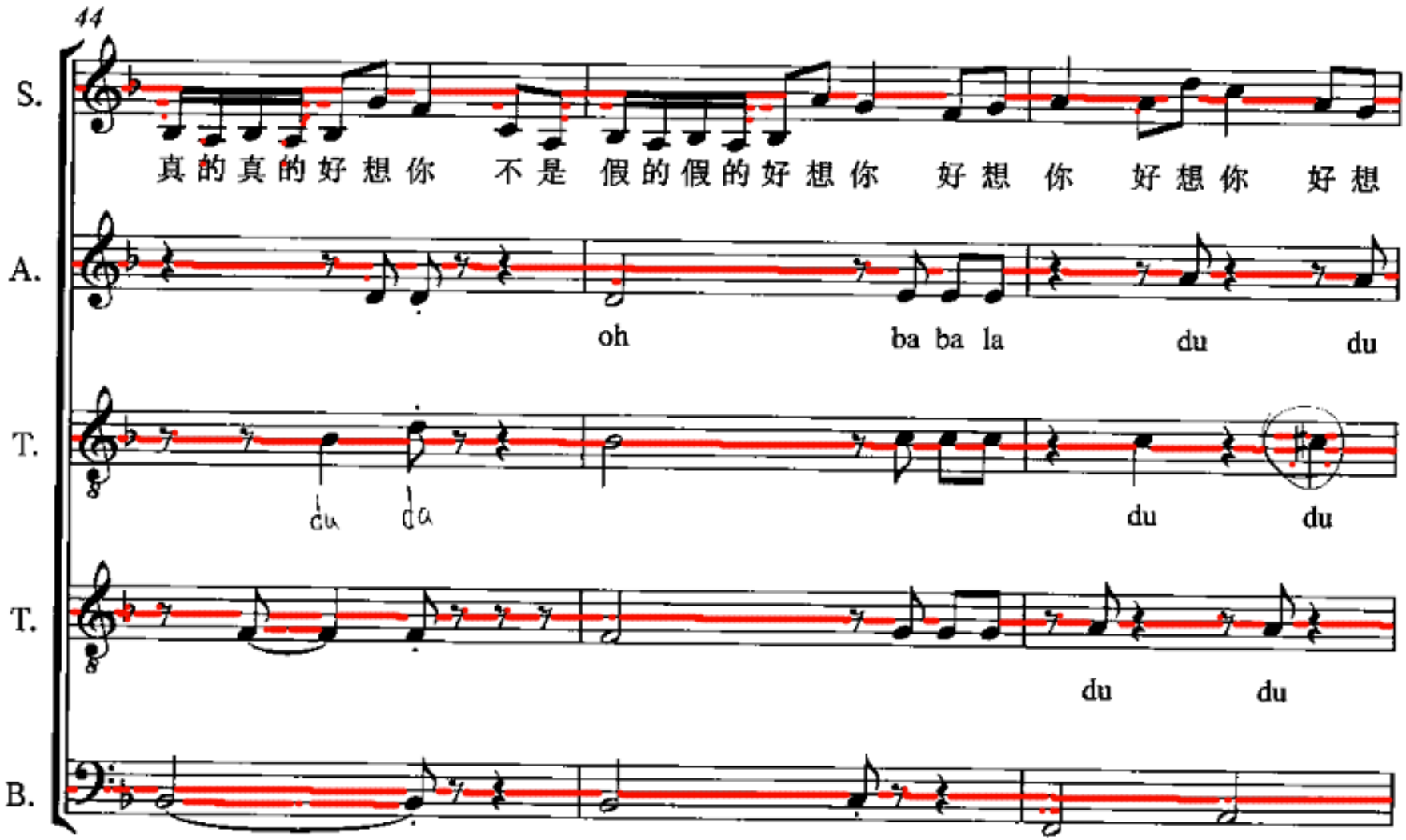
$$e(v') = \frac{\|v' - S\|_1}{\|S\|_1},$$

$$S = [w, s, w, s, w, s, w, s, w]$$

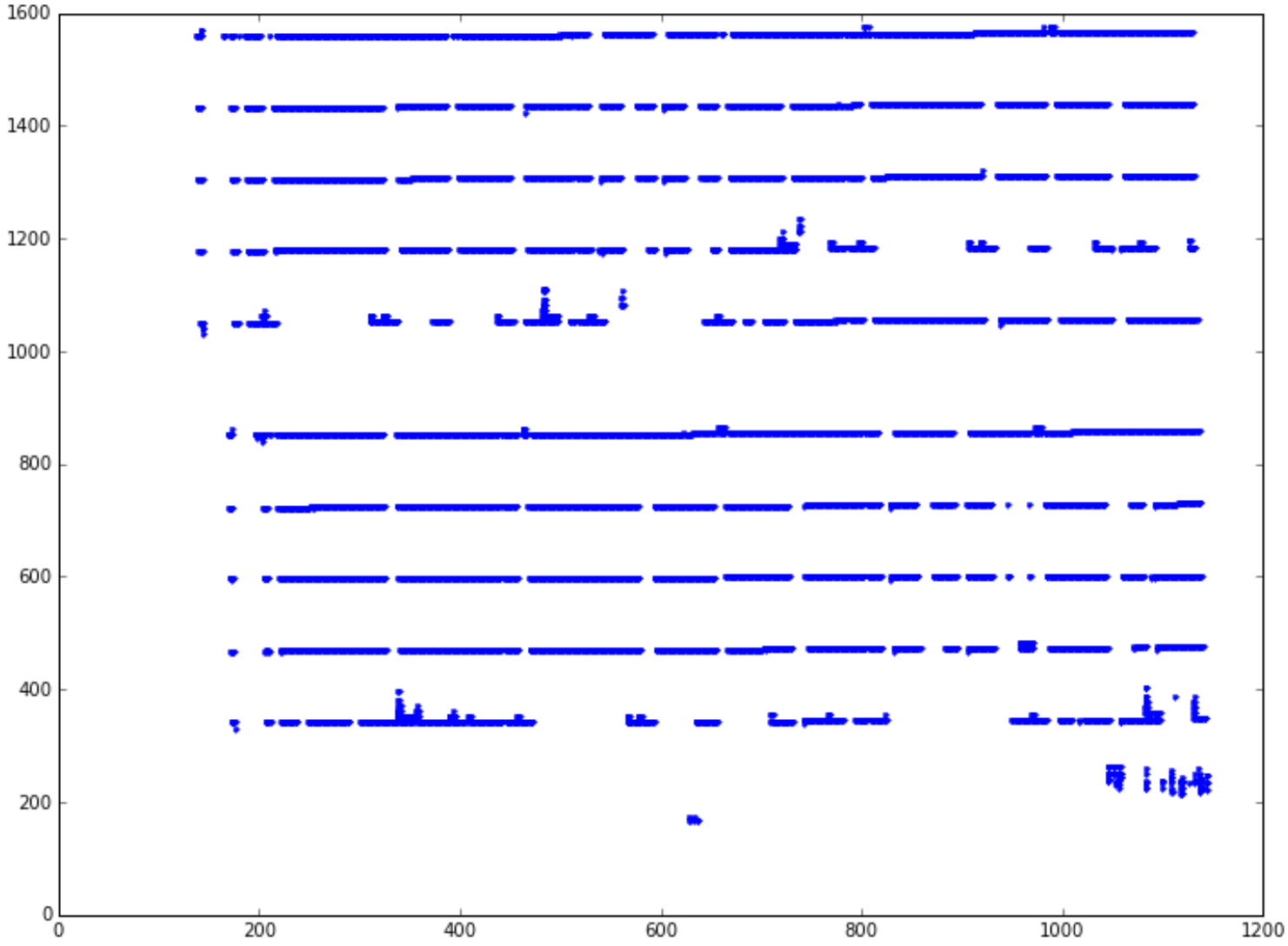
Staffsegment



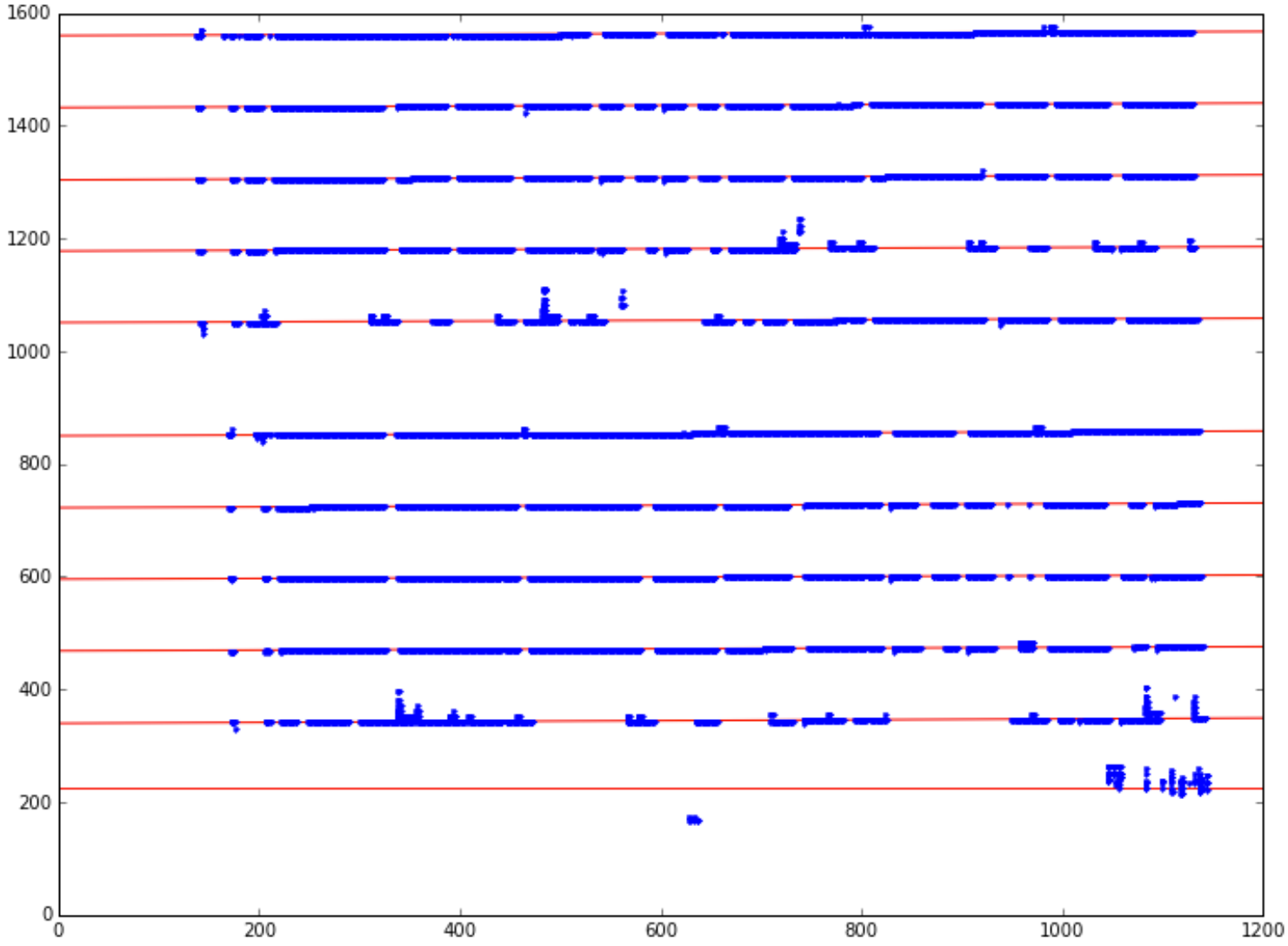
Staffsegment



Multiple Lines Fitting



Multiple Lines Fitting



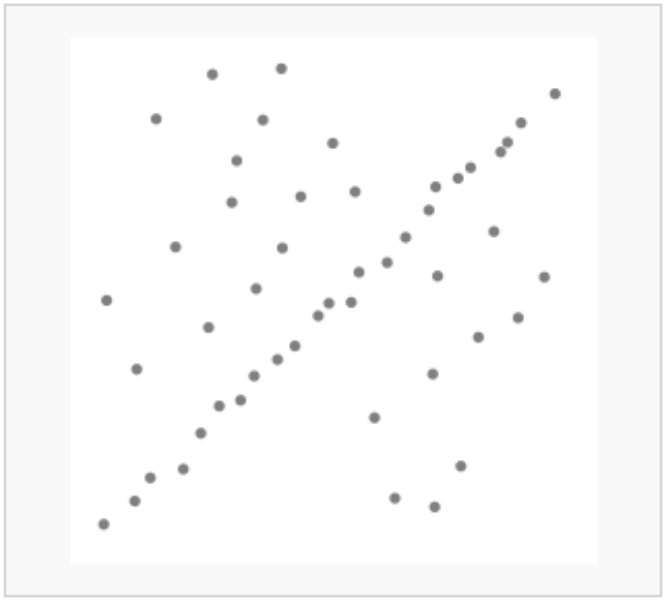
Multiple Lines Fitting

- Dynamic Programming: $O((\frac{N}{W})^2 \times W)$
- Hough Transform: $O(L \times N)$
- **RANSAC**: $O(N_{\text{iter}} \times N)$

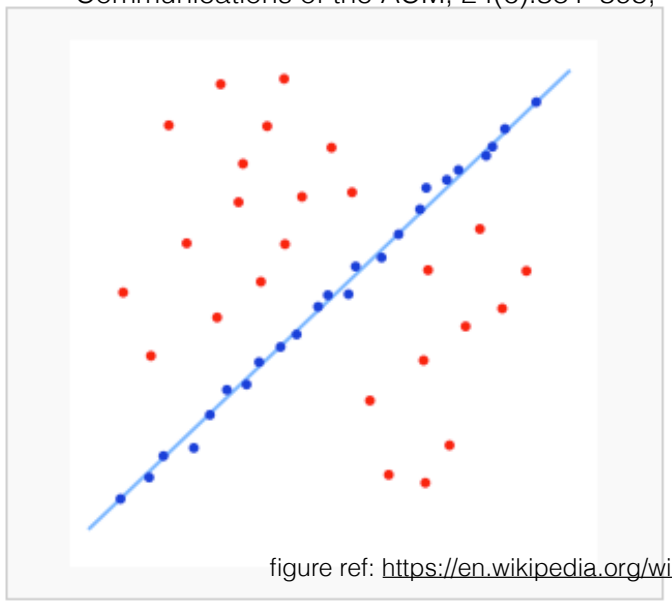
Multiple Lines Fitting

- RANSAC (RANdom SAmple Consensus)
  - single model fitting method
  - robust to outliers

M.A. Fischler and R.C. Bolles. **Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography**. Communications of the ACM, 24(6):381–395, 1981.



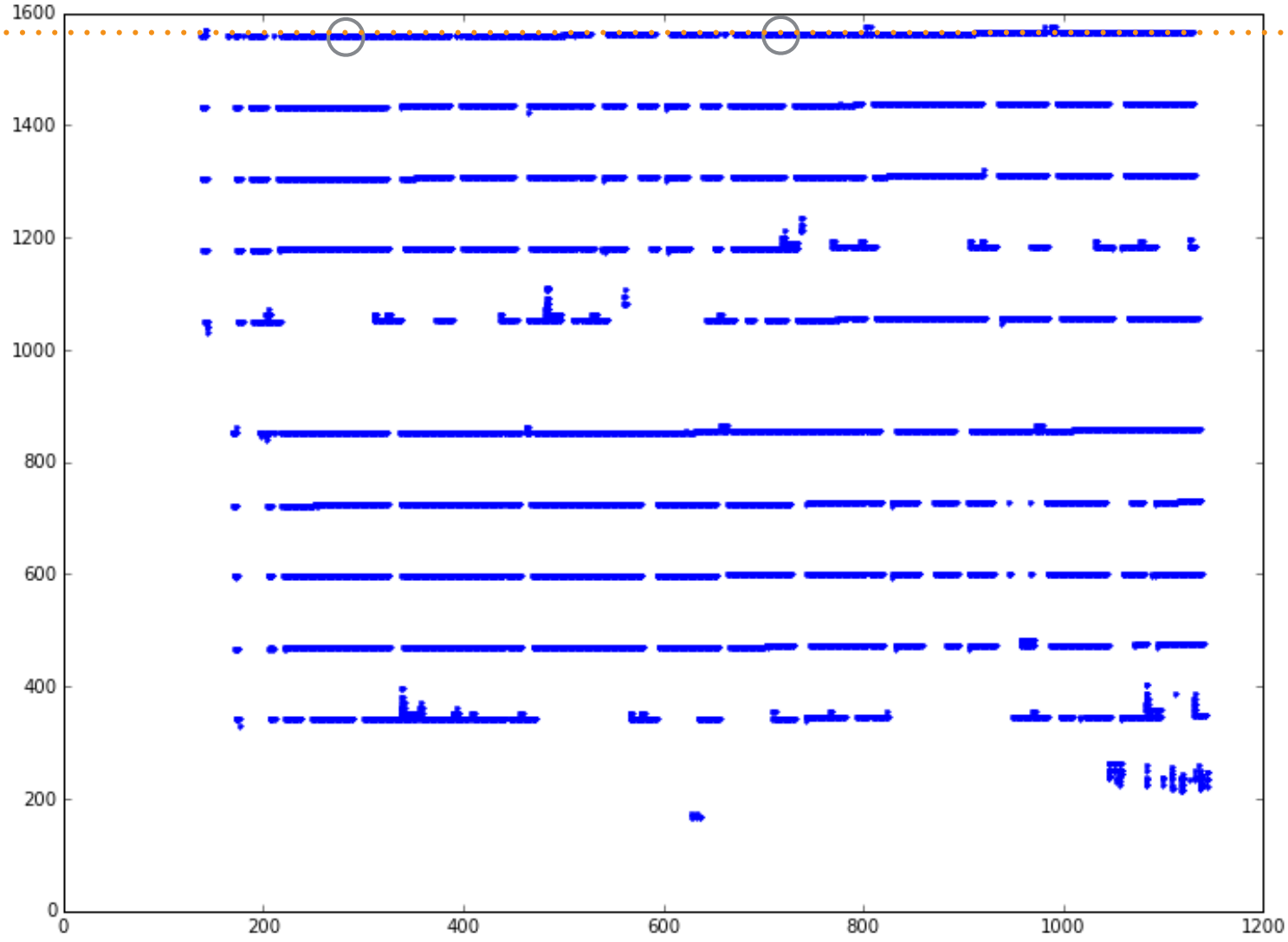
A data set with many outliers for which a line has to be fitted.



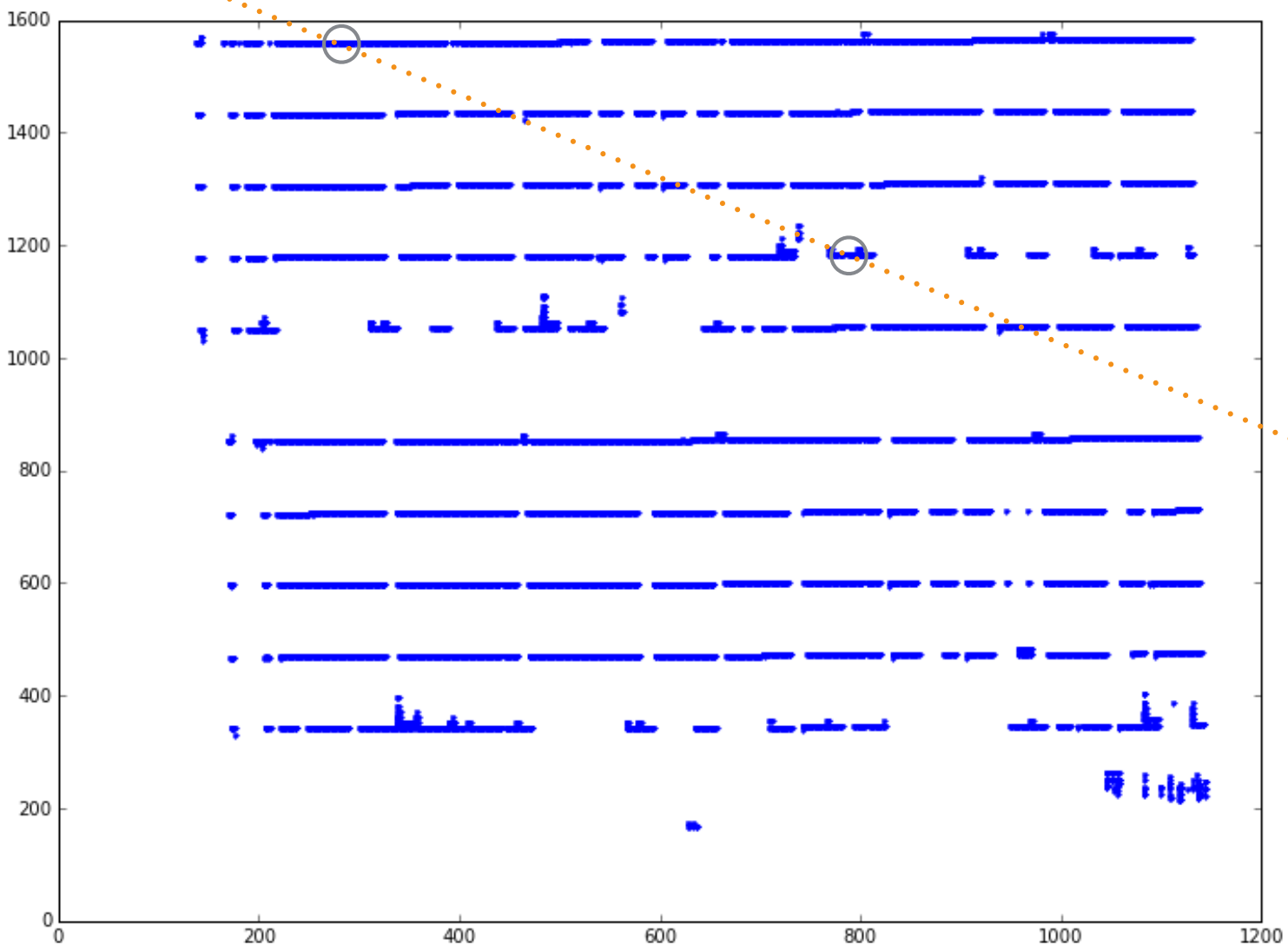
Fitted line with RANSAC; outliers have no influence on the result.

figure ref: https://en.wikipedia.org/wiki/RANSAC

RANSAC: Example

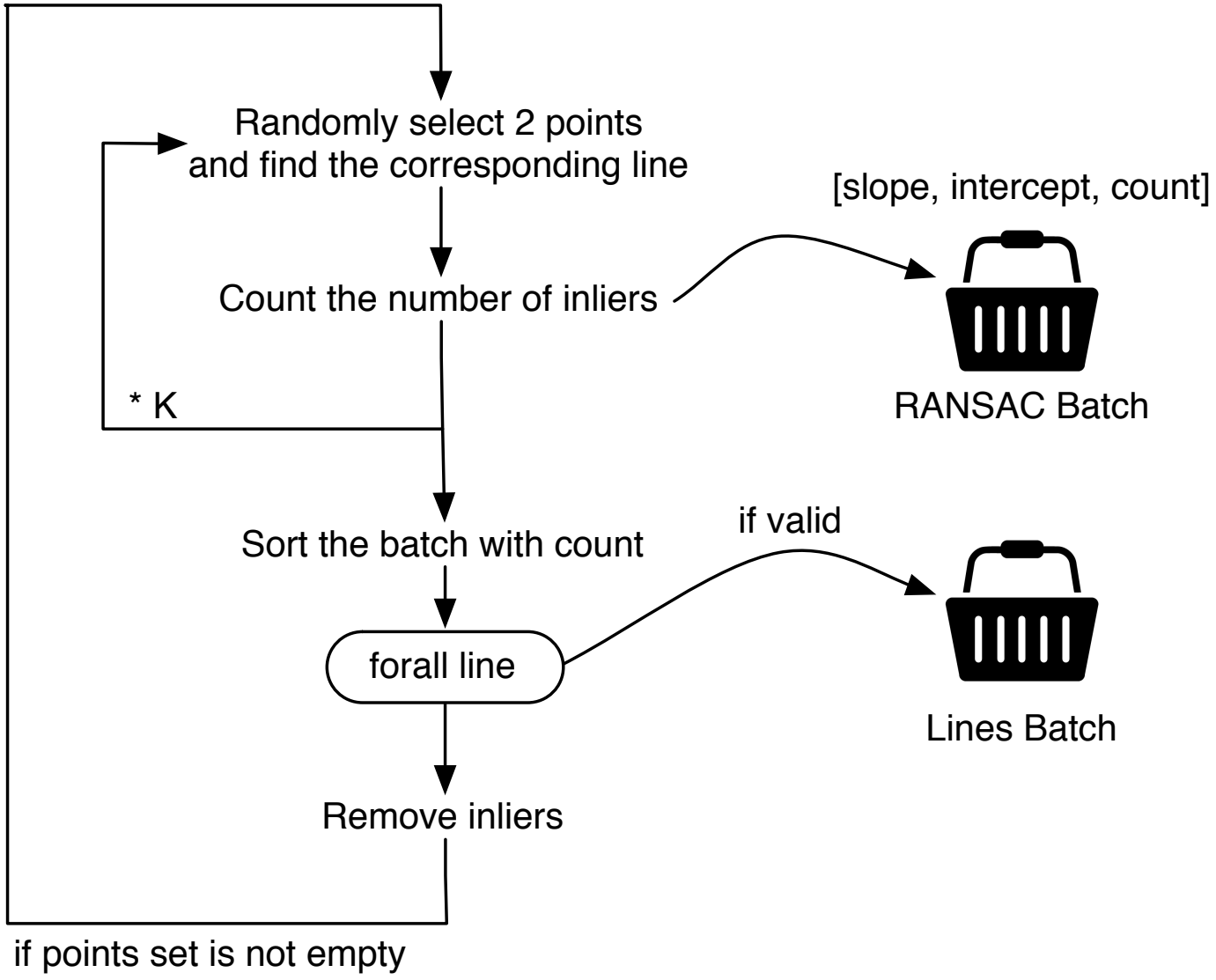


RANSAC: Example

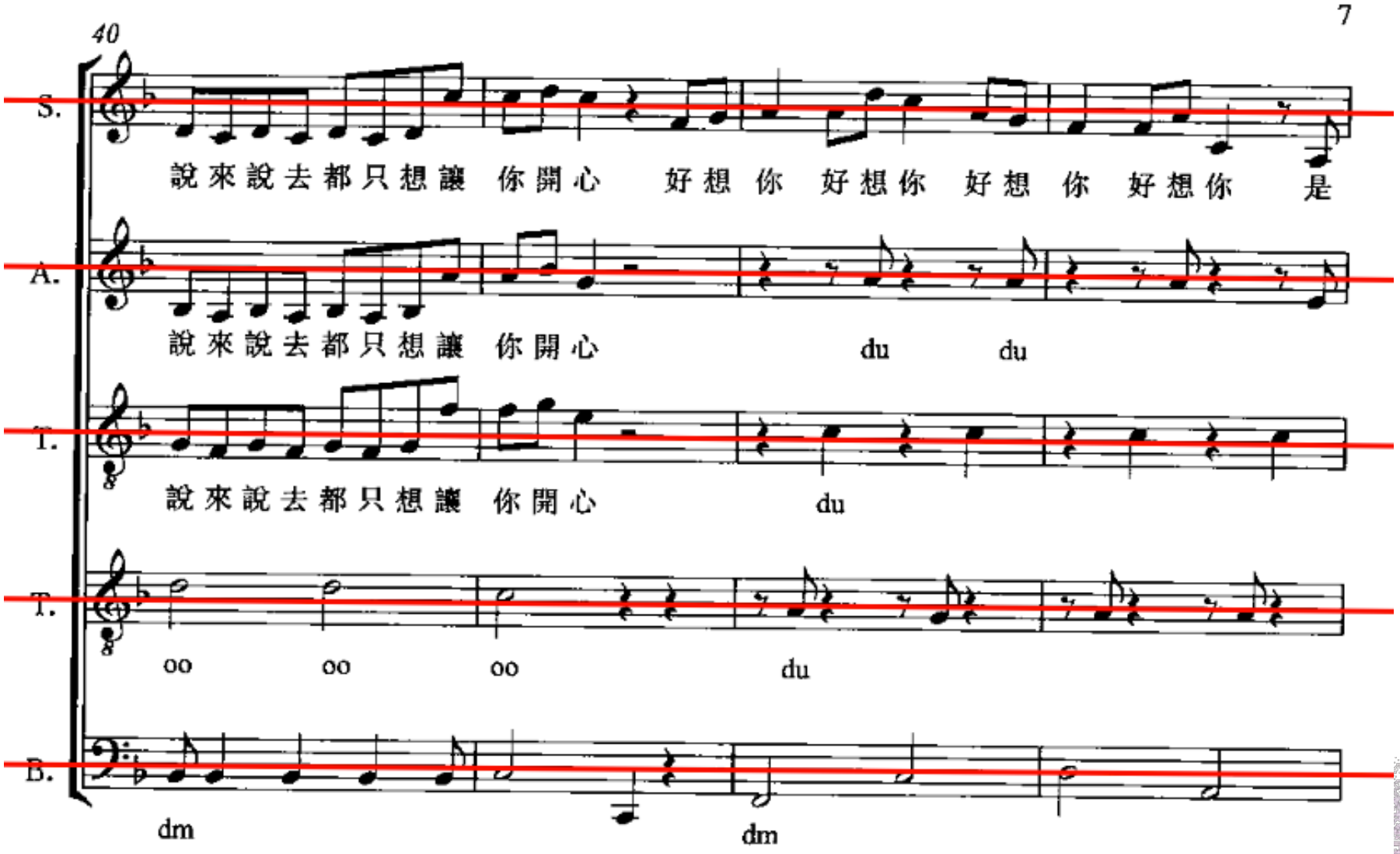


7

Iterative RANSAC

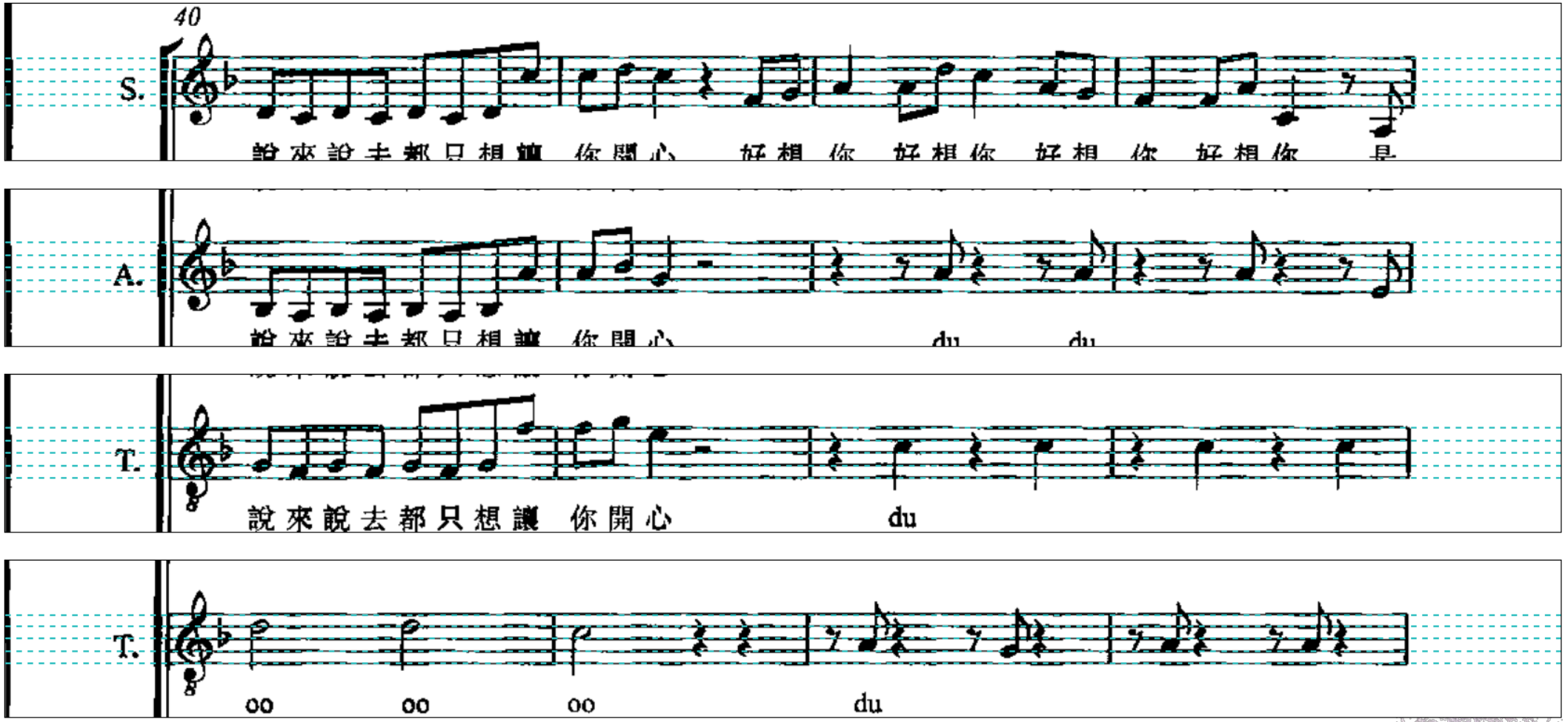


Detected Staves



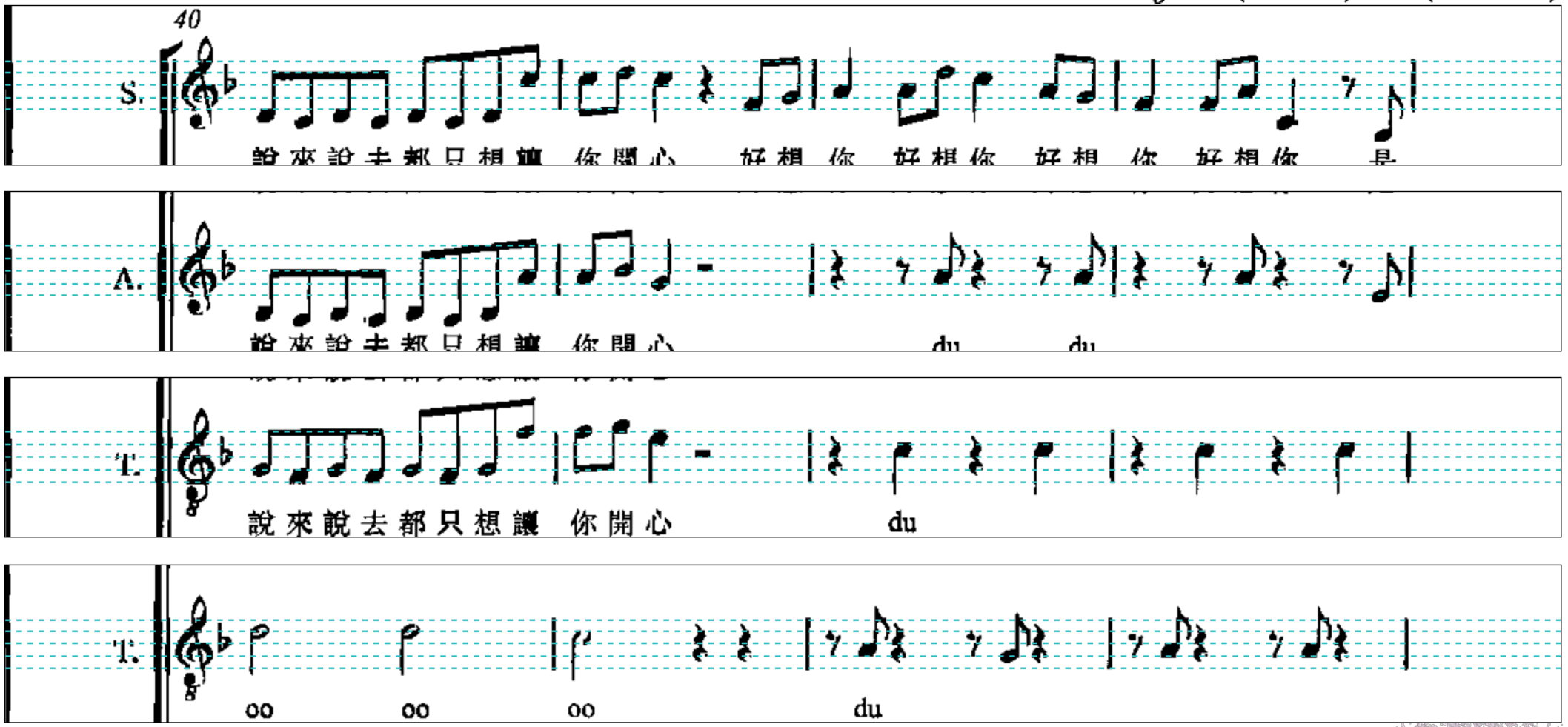
Staff Images (Before Staff Removal)

$$l_i = c_y + (i-3) \times (w+s)$$



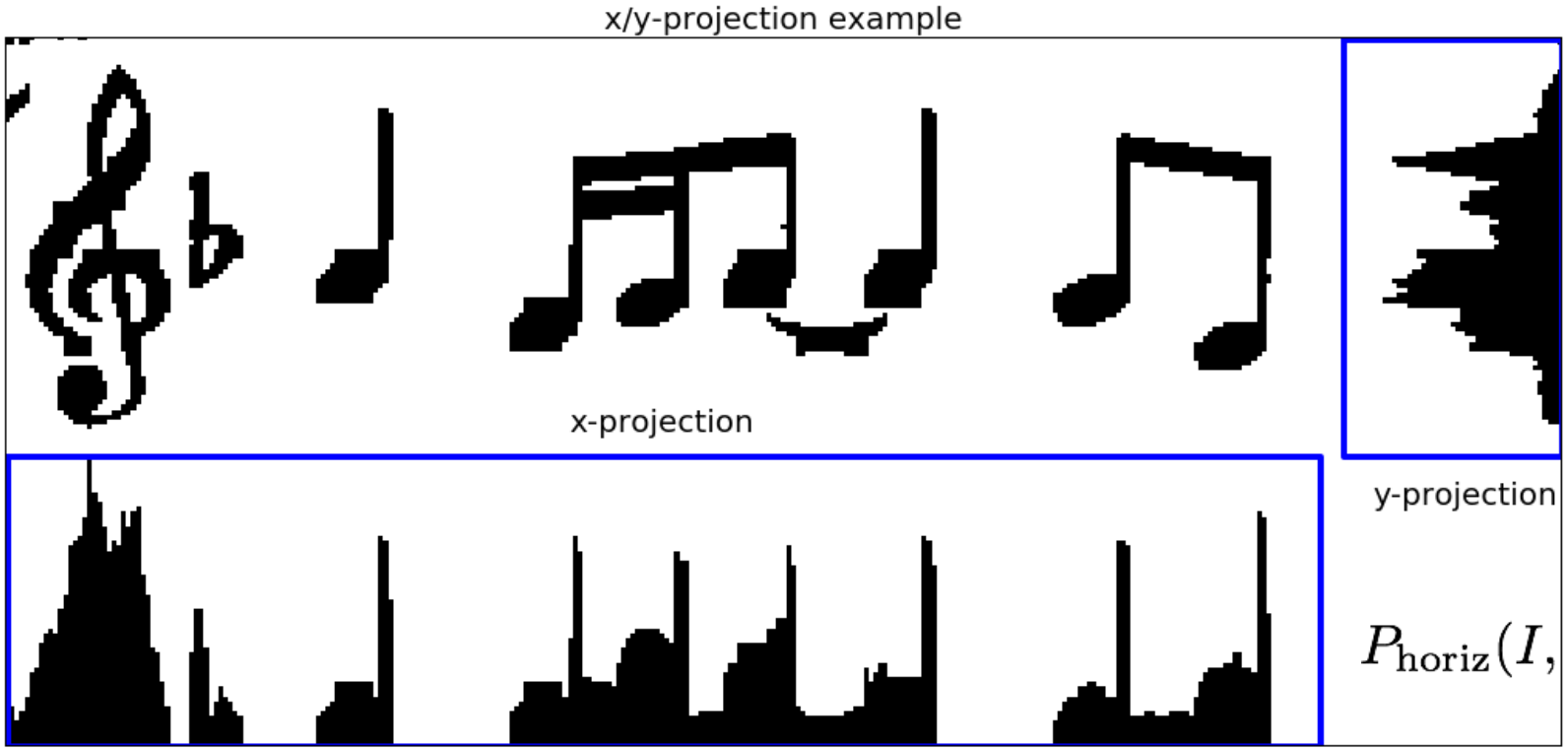
Staff Images (After Staff Removal)

$$l_i = c_y + (i-3) \times (w+s)$$



Recognition

Projection

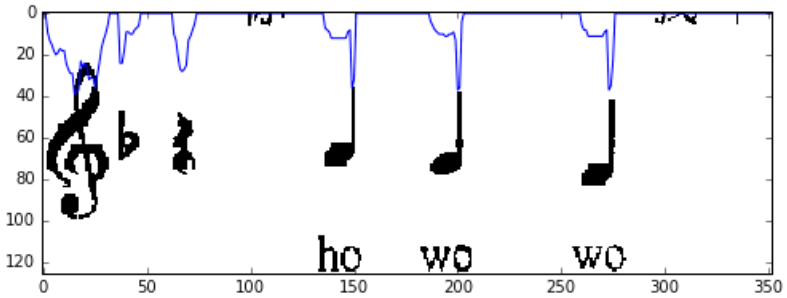


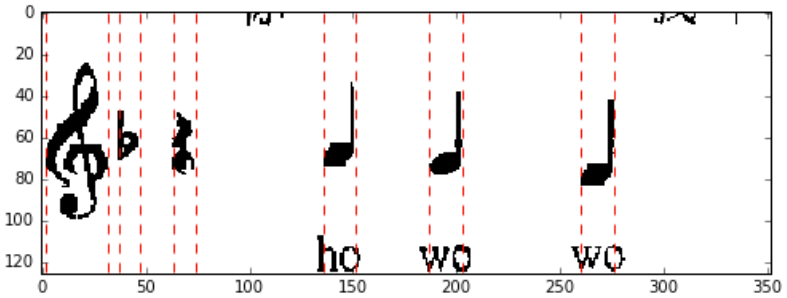
$$P_{\text{horiz}}(I, x_0) = \sum_y I(x_0, y)$$

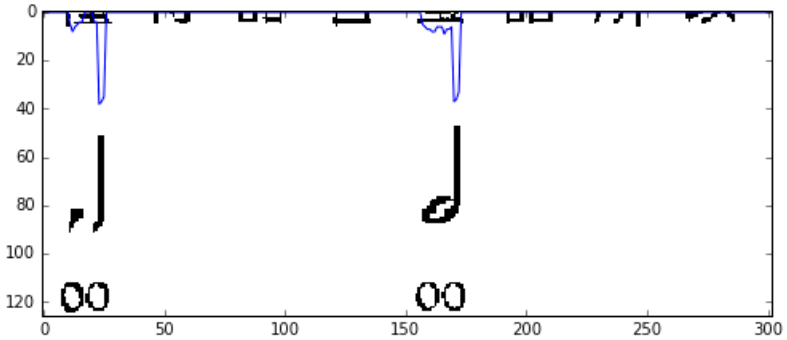
$$P_{\text{vert}}(I, y_0) = \sum_x I(x, y_0)$$

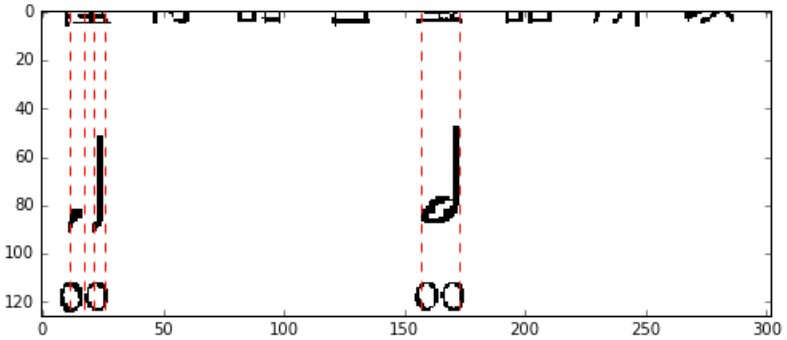
Onset Detection

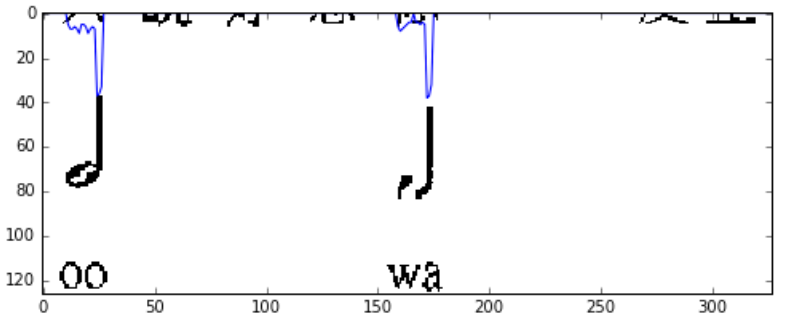
merge_thresh=0

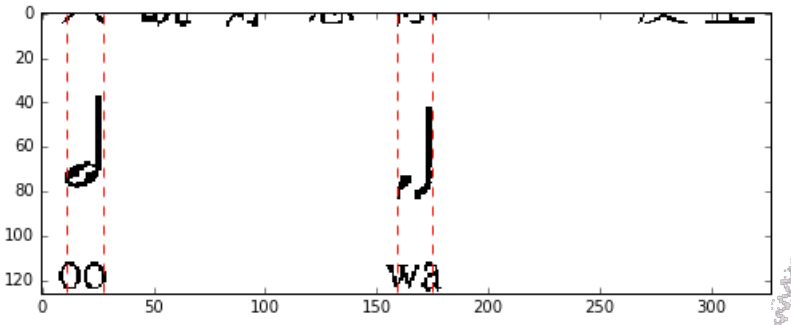






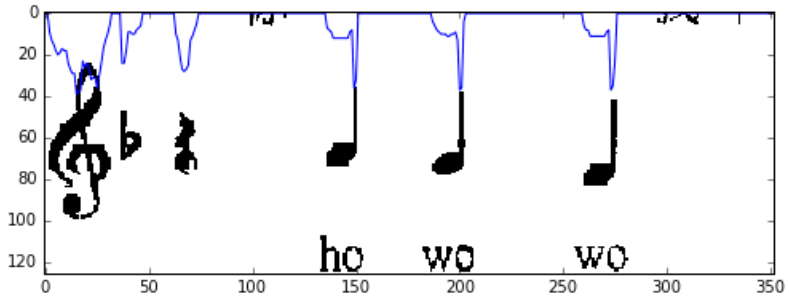


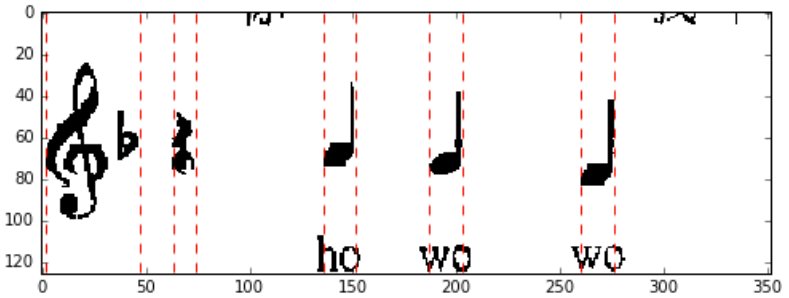


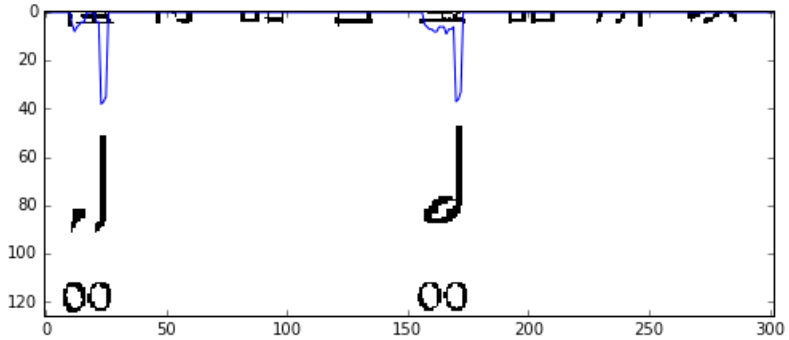


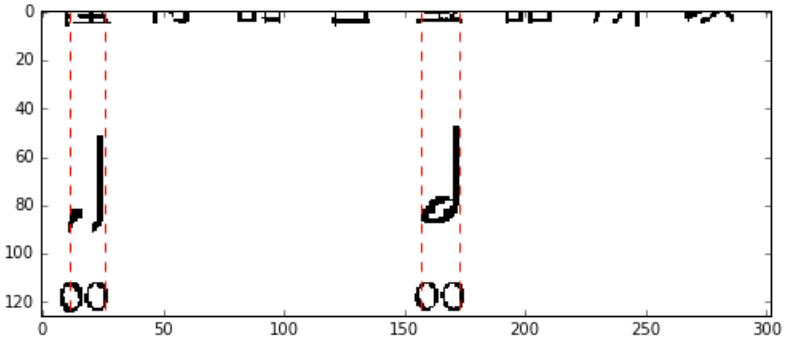
Onset Detection

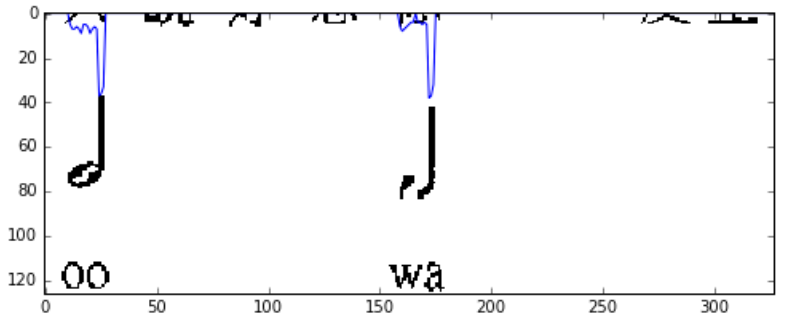
`merge_thresh=10`

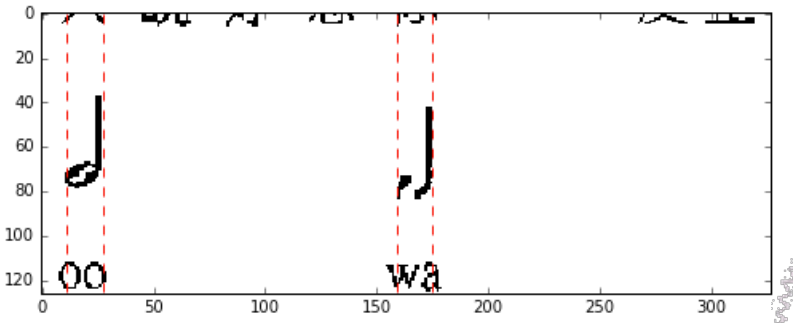






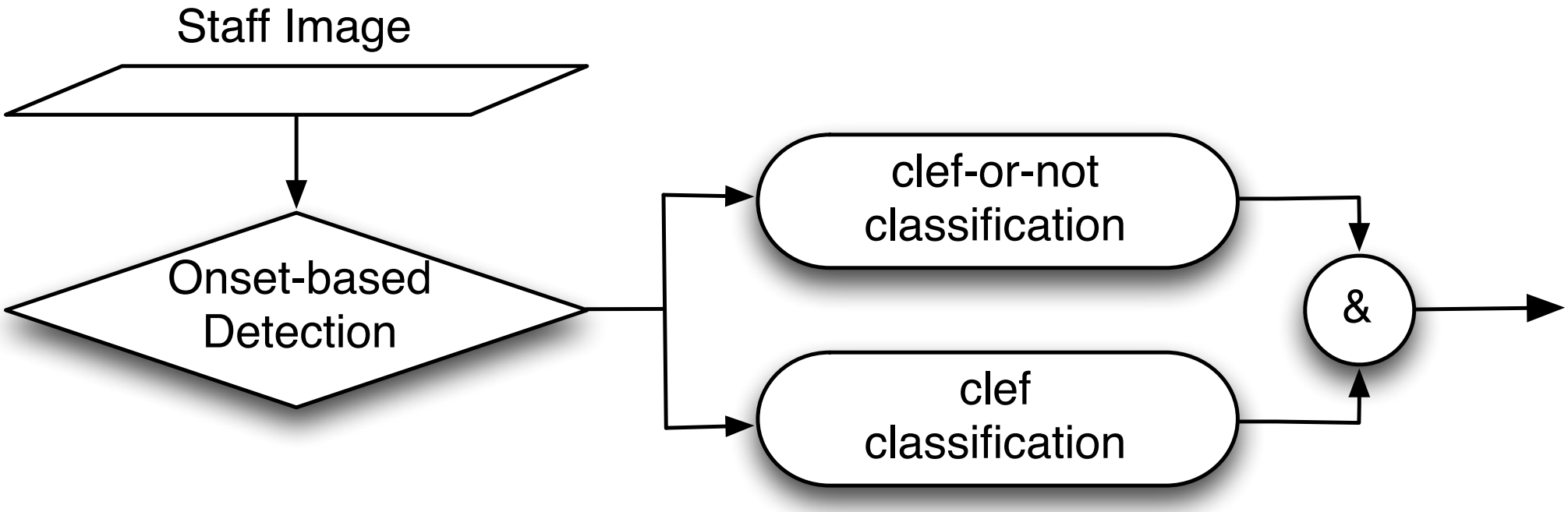




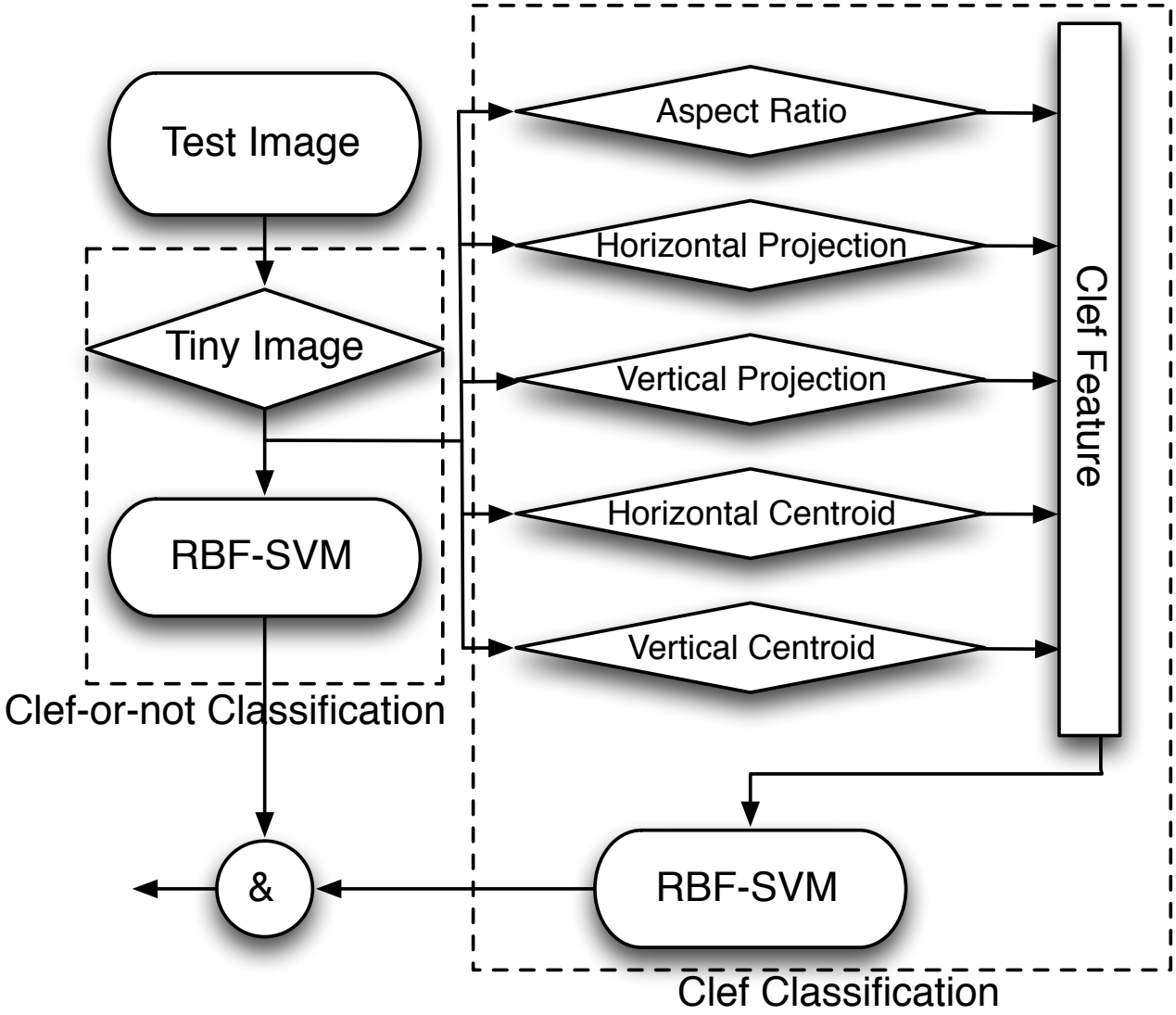


Recognition
Pitch-unrelated Symbols

Clef Detection



Clef Detection: Feature



Centroid

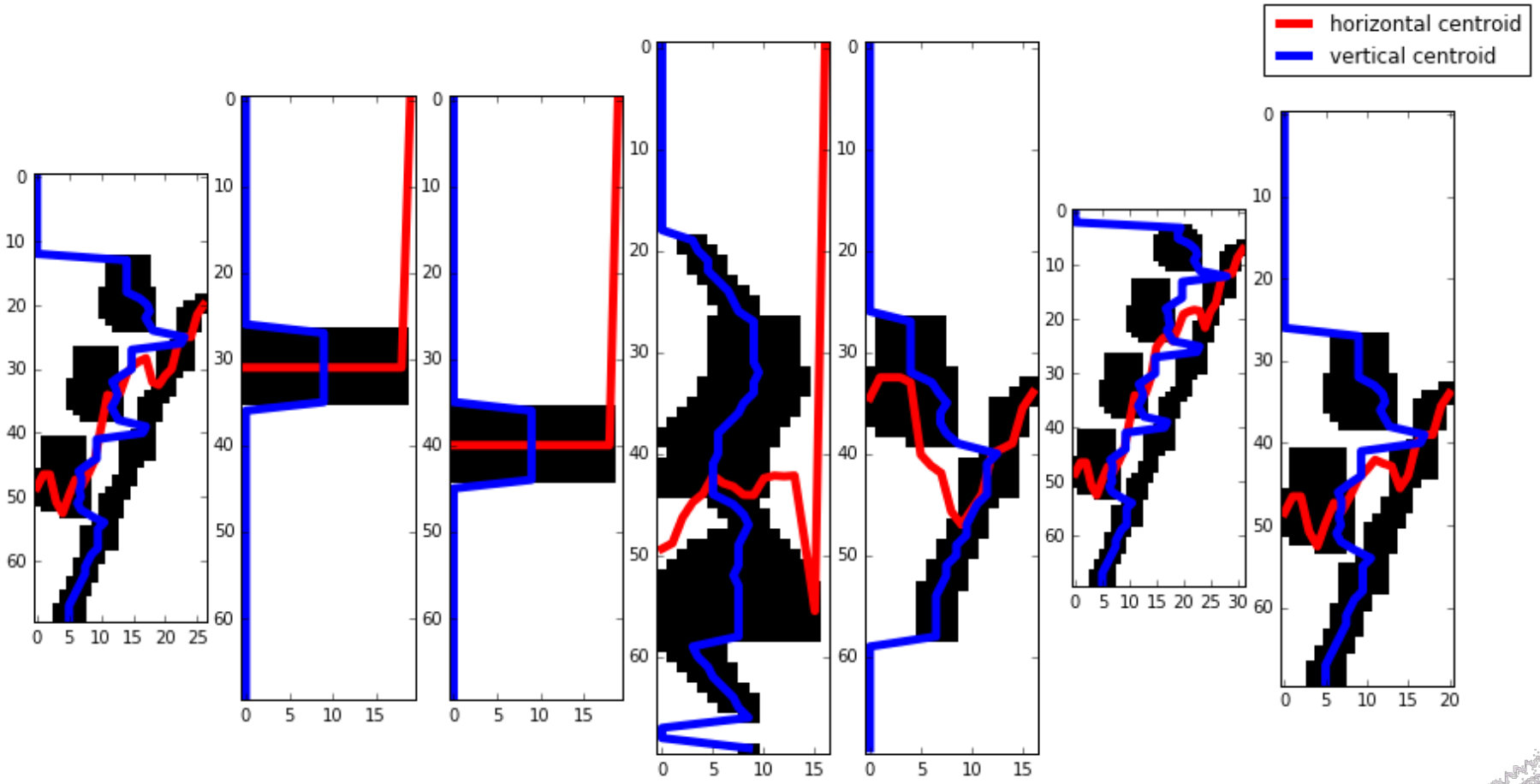
$$C_{\text{horiz}}(x_0) = \frac{\sum_y I(x_0, y) \times y}{\sum_y I(x_0, y)}$$

$$C_{\text{vert}}(y_0) = \frac{\sum_x I(x, y_0) \times x}{\sum_x I(x, y_0)}$$

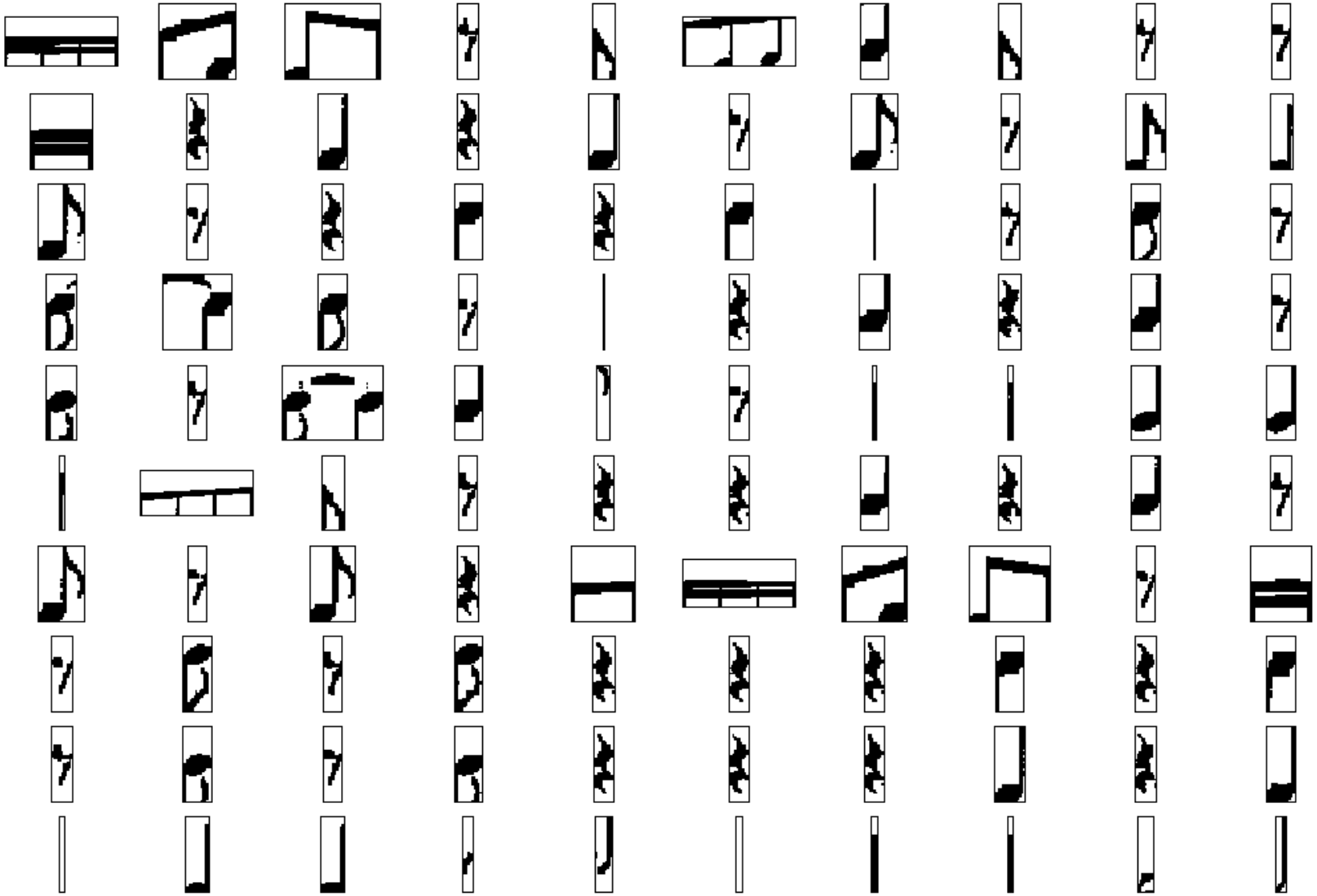
Centroid

$$C_{\text{horiz}}(x_0) = \frac{\sum_y I(x_0, y) \times y}{\sum_y I(x_0, y)}$$

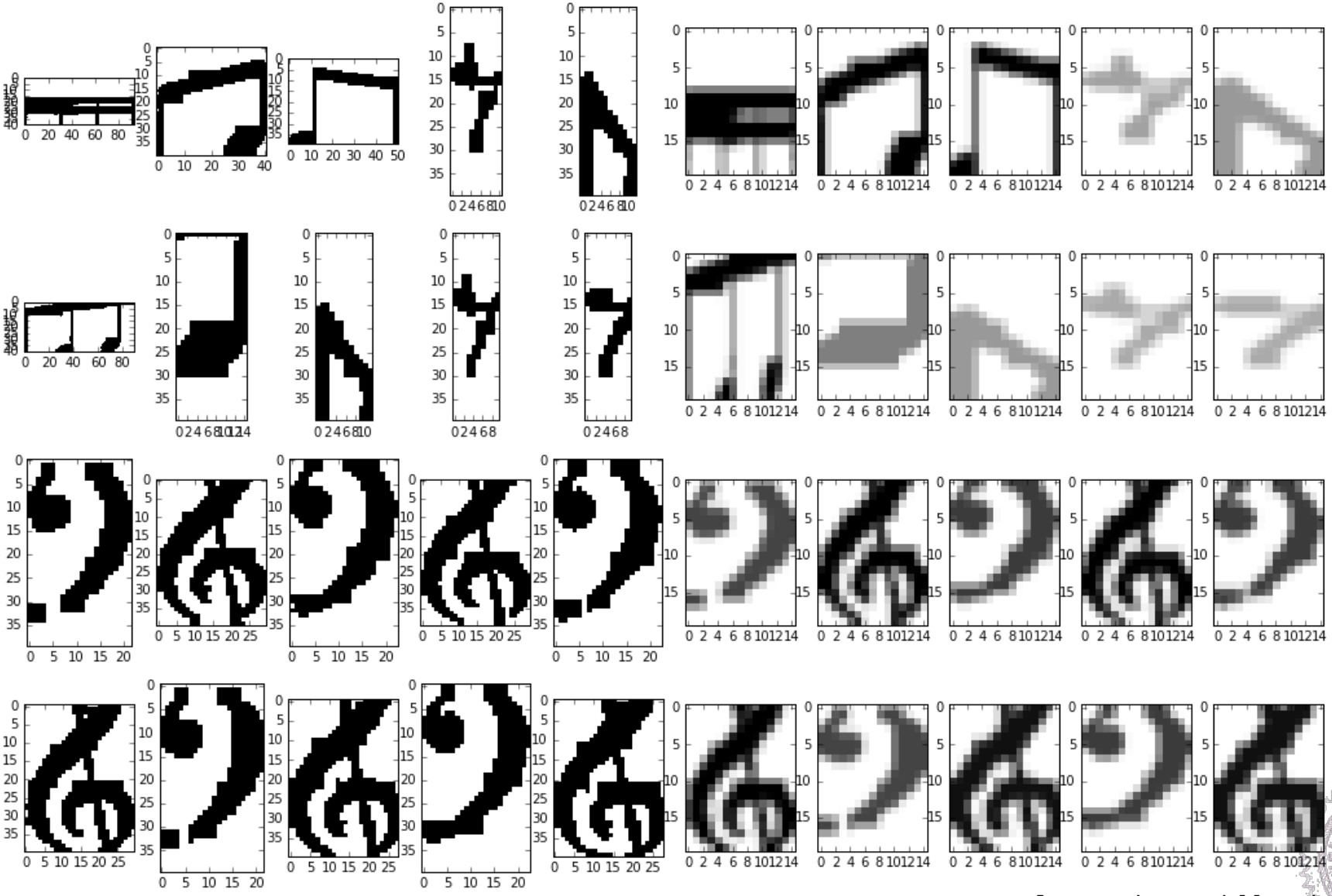
$$C_{\text{vert}}(y_0) = \frac{\sum_x I(x, y_0) \times x}{\sum_x I(x, y_0)}$$



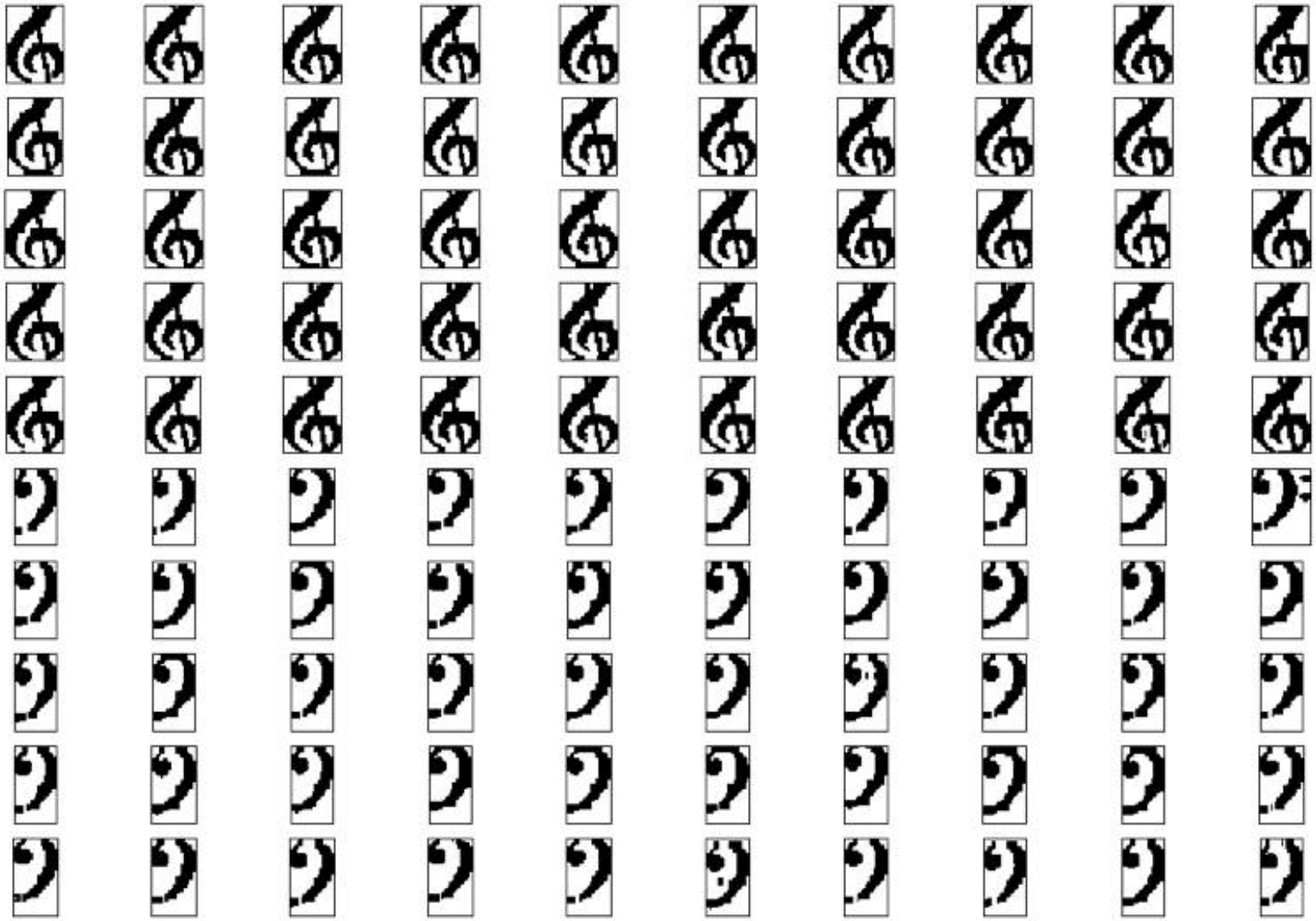
Clef-or-not Samples



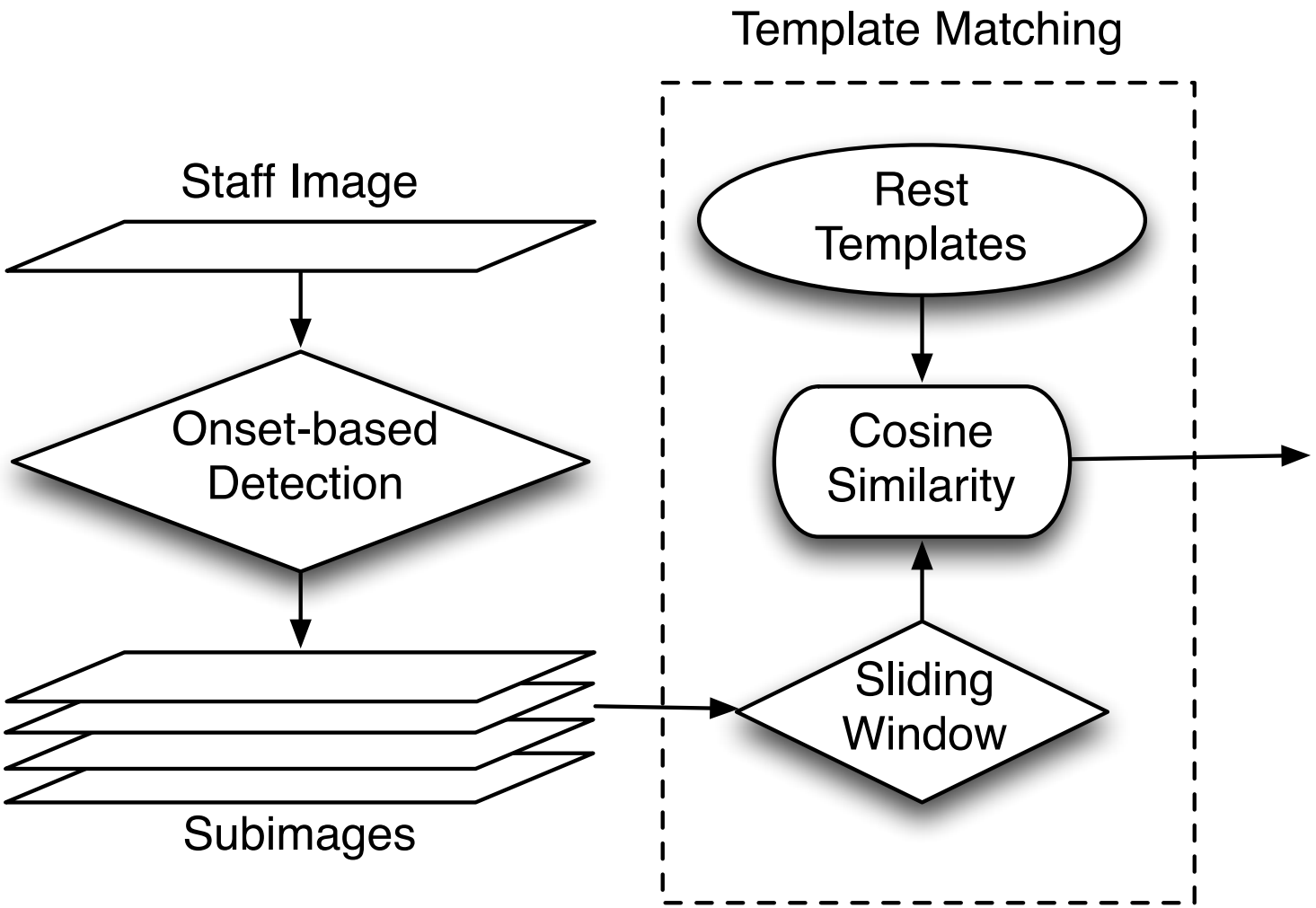
Clef-or-not Feature



Clef Samples

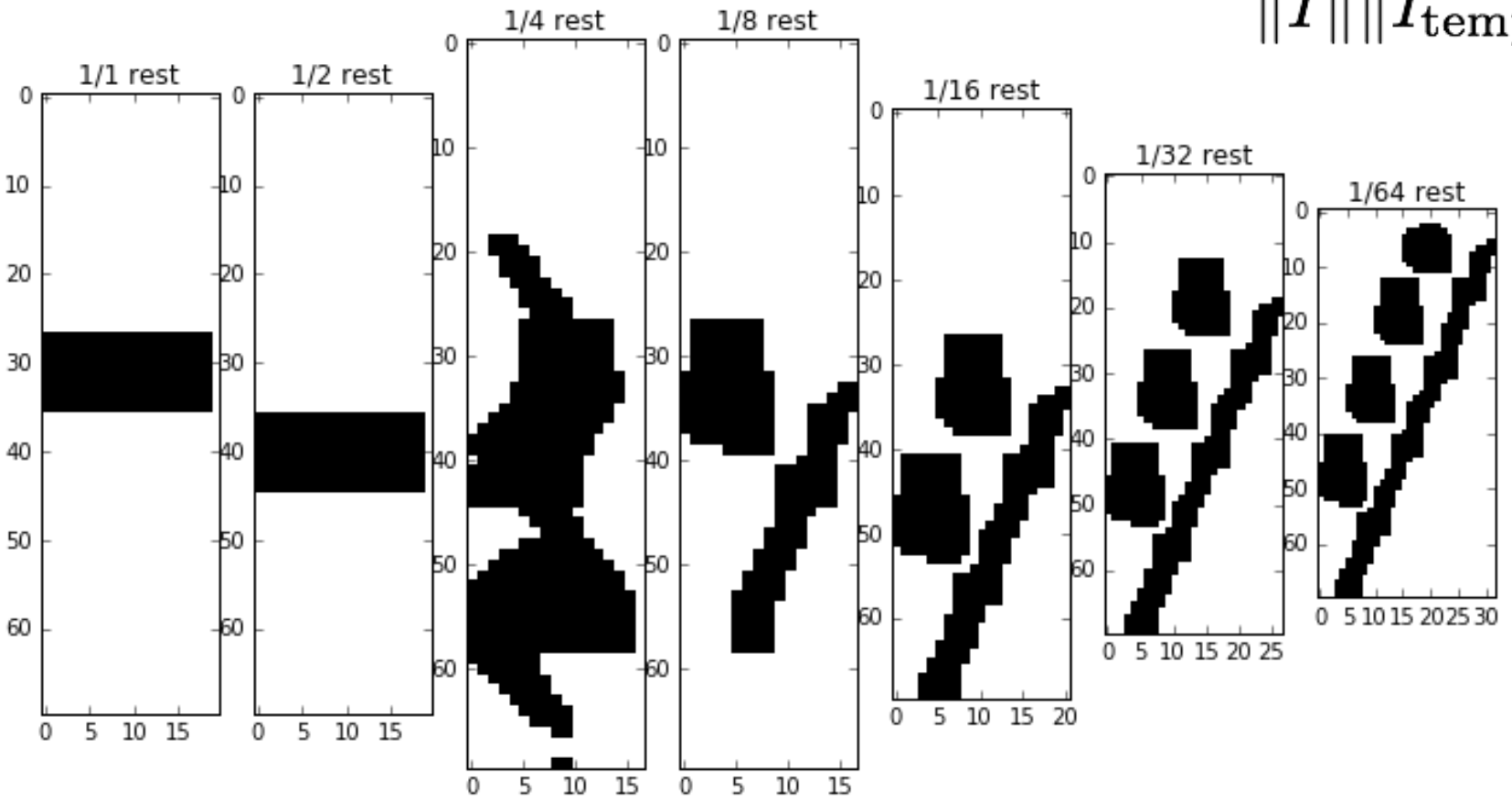


Rest Detection

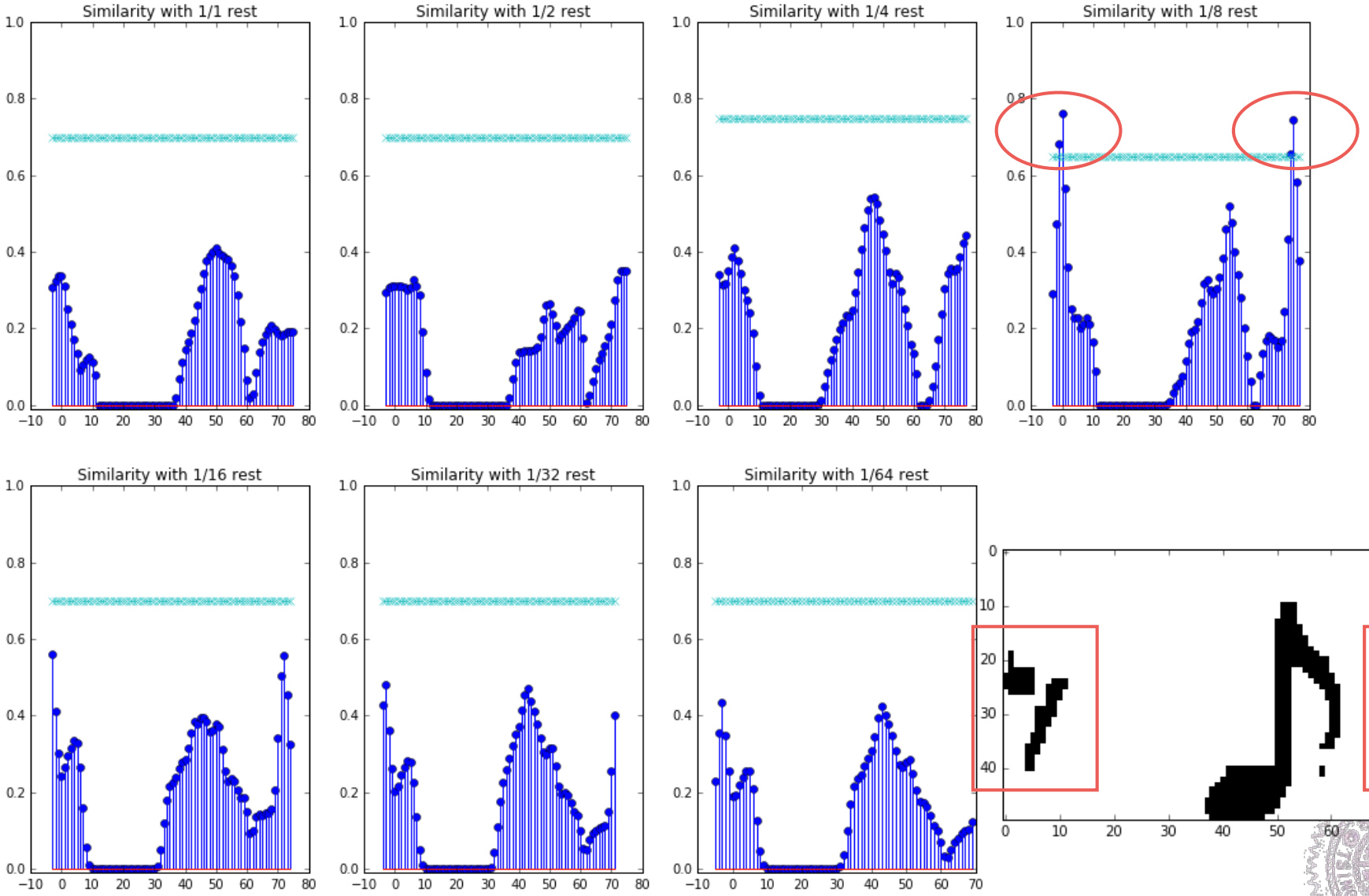


Rest Templates

$$\text{score} = \frac{\langle I,\ I_{\text{template}} \rangle}{\|I\| \|I_{\text{template}}\|}$$



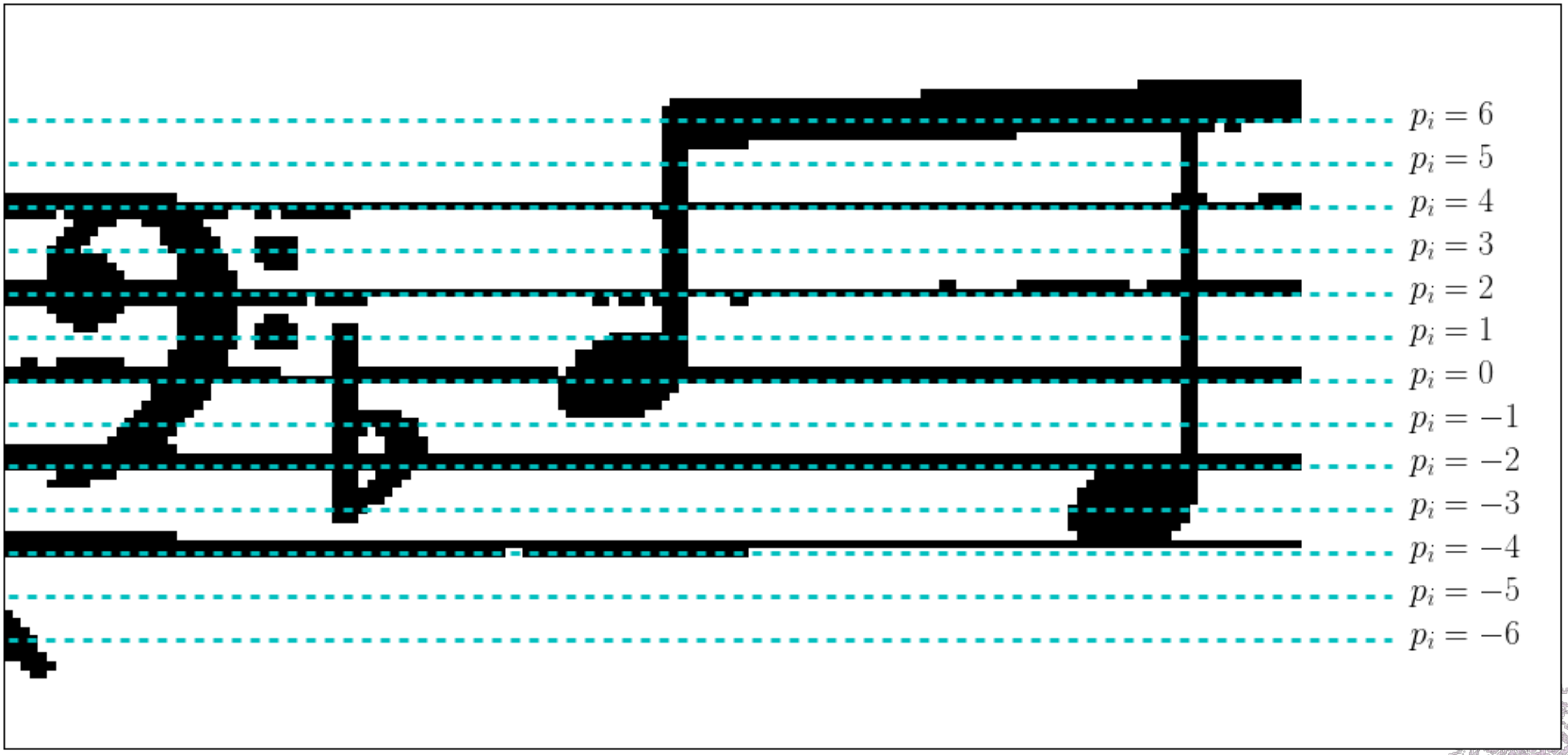
Rest Detection
Template Matching



Recognition
Pitch-related Symbols

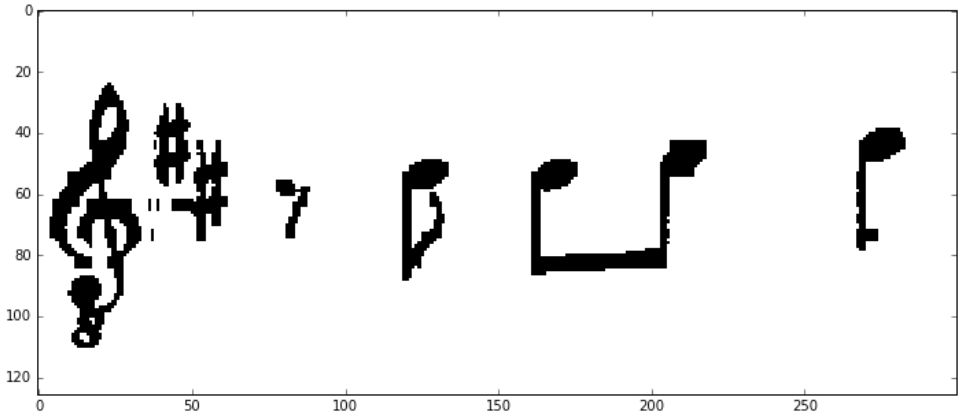
Pitch Cropping

$$y_{p_i} = l_3 - p_i \times 0.5(w + s)$$

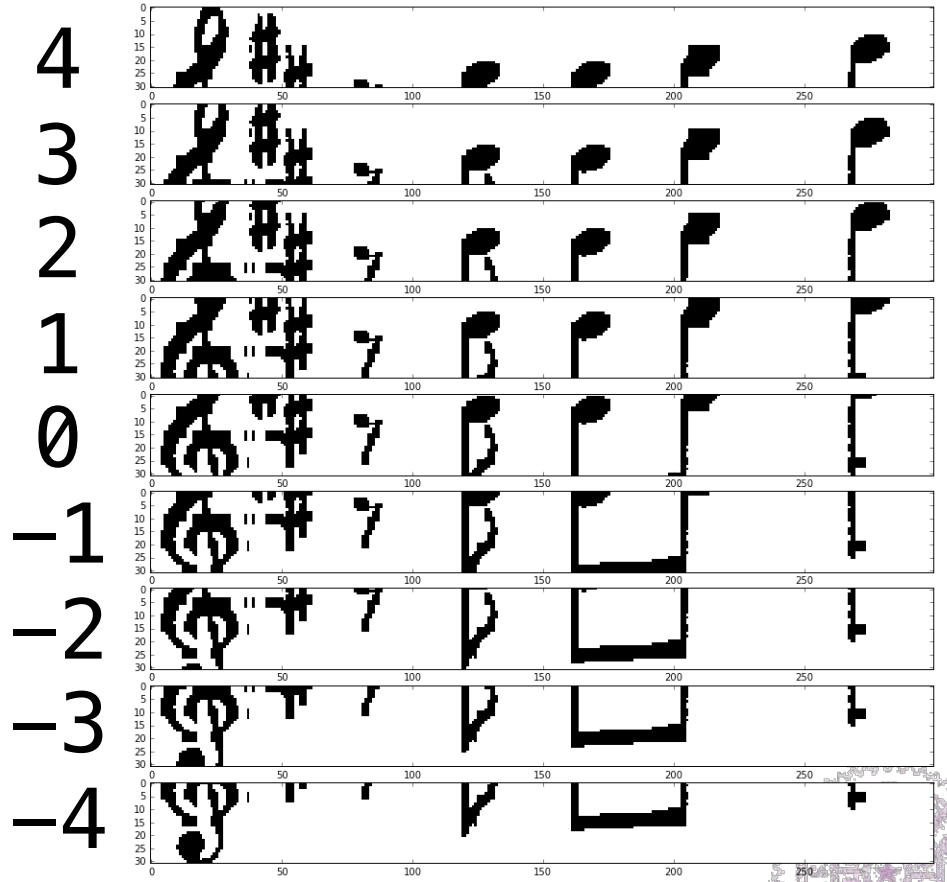


Pitch Cropping

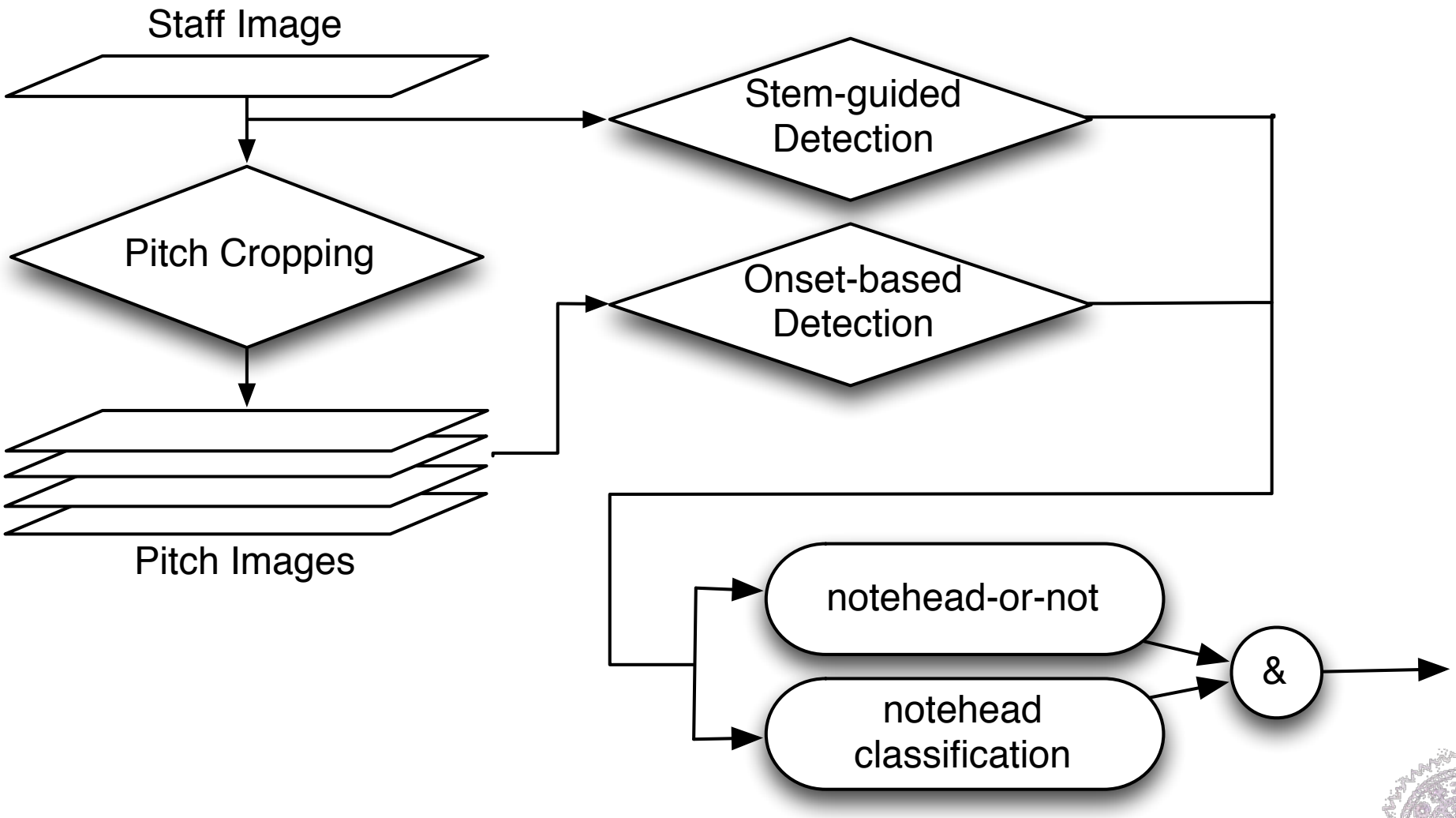
$$\rho_{\text{seg}}(I,\, p_{\text{range}},\, r) = \{I(:,\, y_{p_i} - r : y_{p_i} + r) \mid p_i = -p_{\text{range}},\, \cdots,\, p_{\text{range}}\}$$



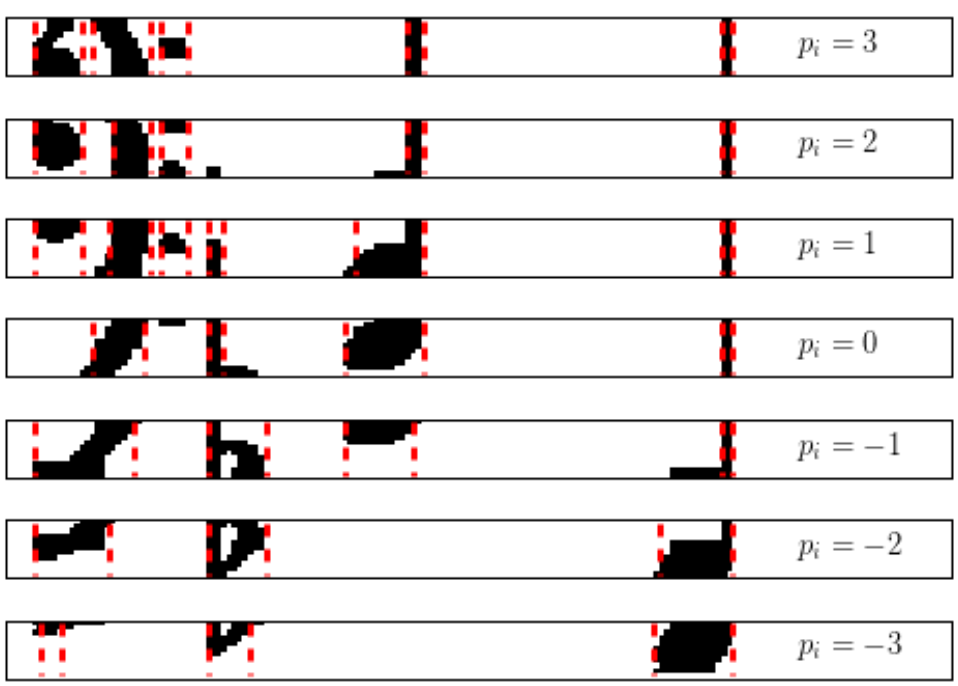
$p_{\text{range}} = 4,\ r = 1.5(w + s)$

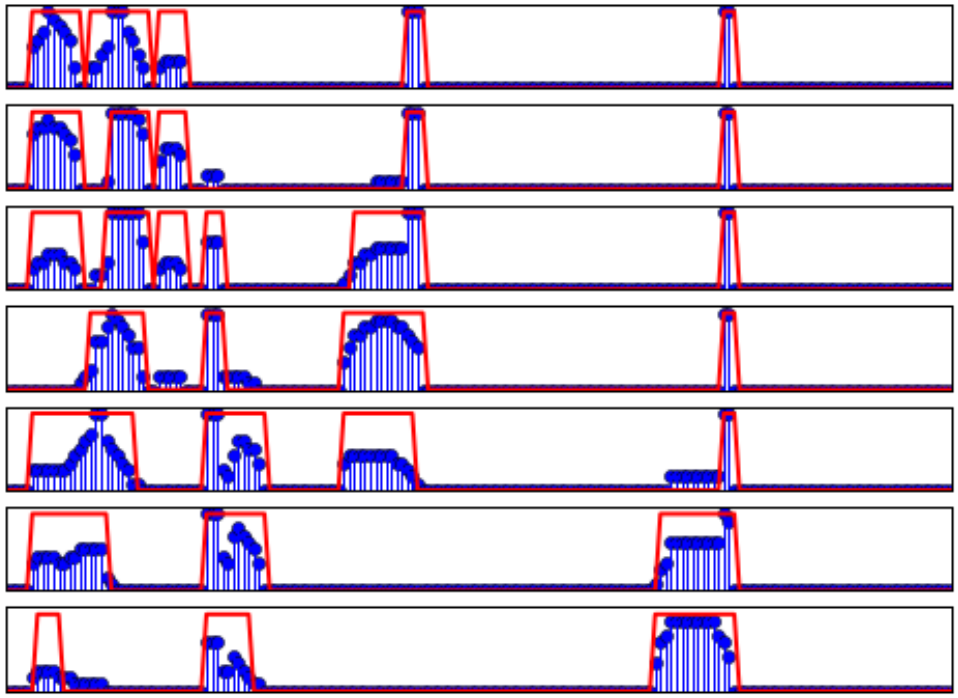


Notehead Detection

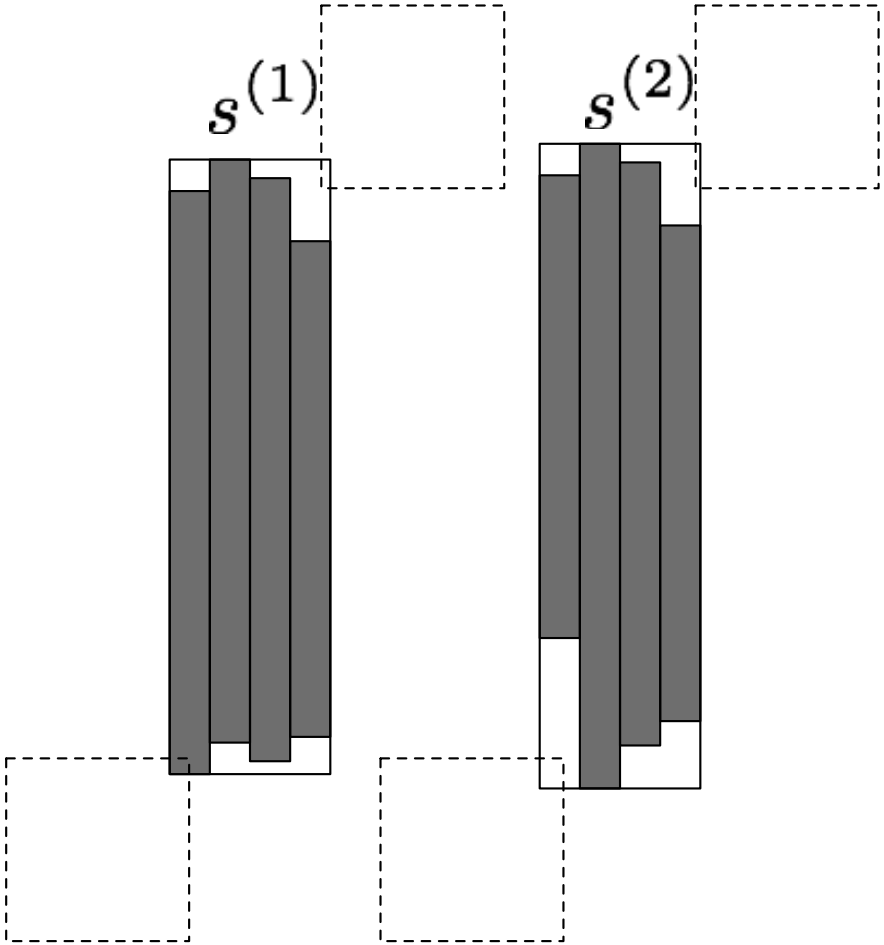


Notehead Detection
Onset-based





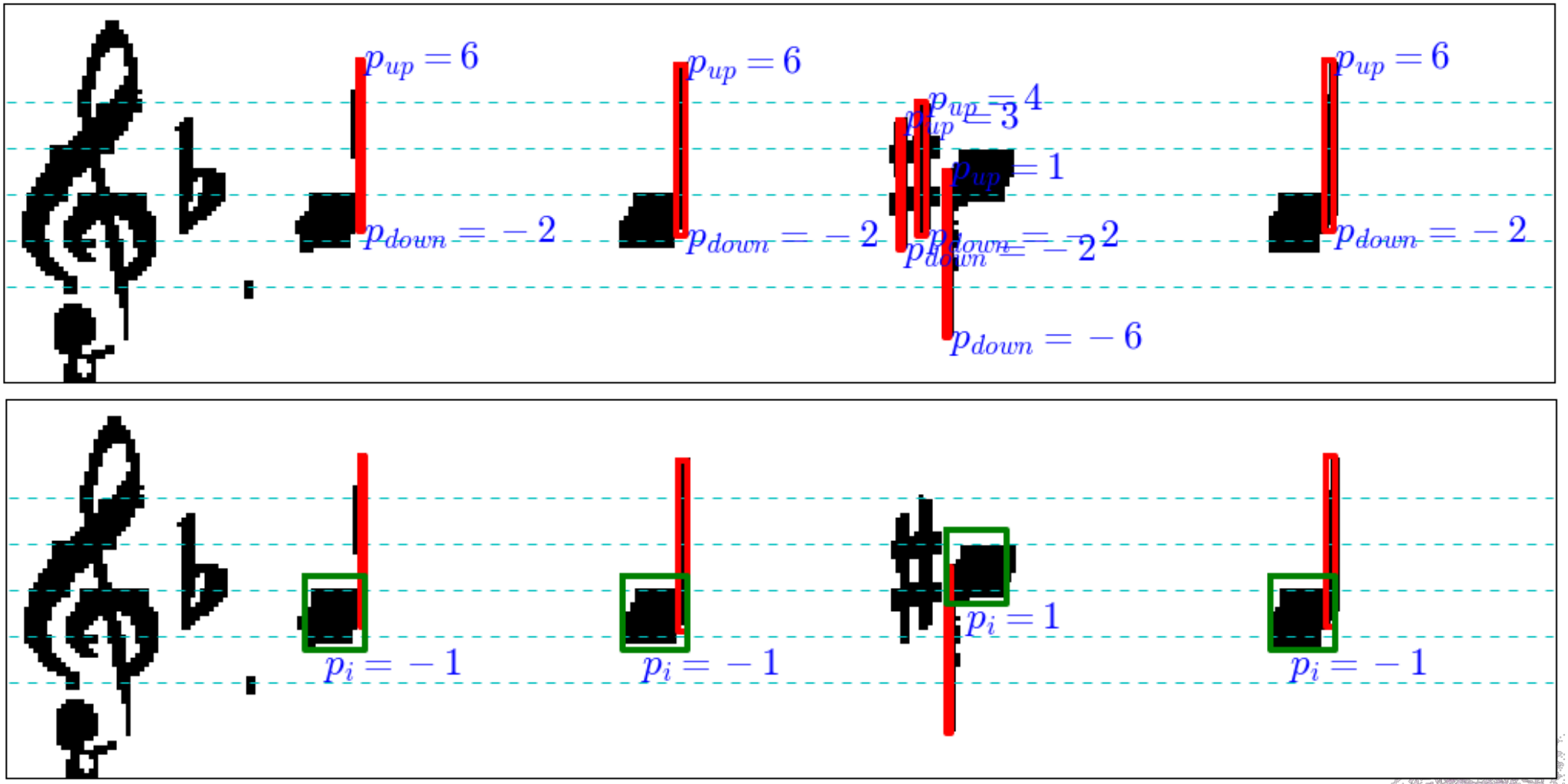
Notehead Detection
Stem-guided



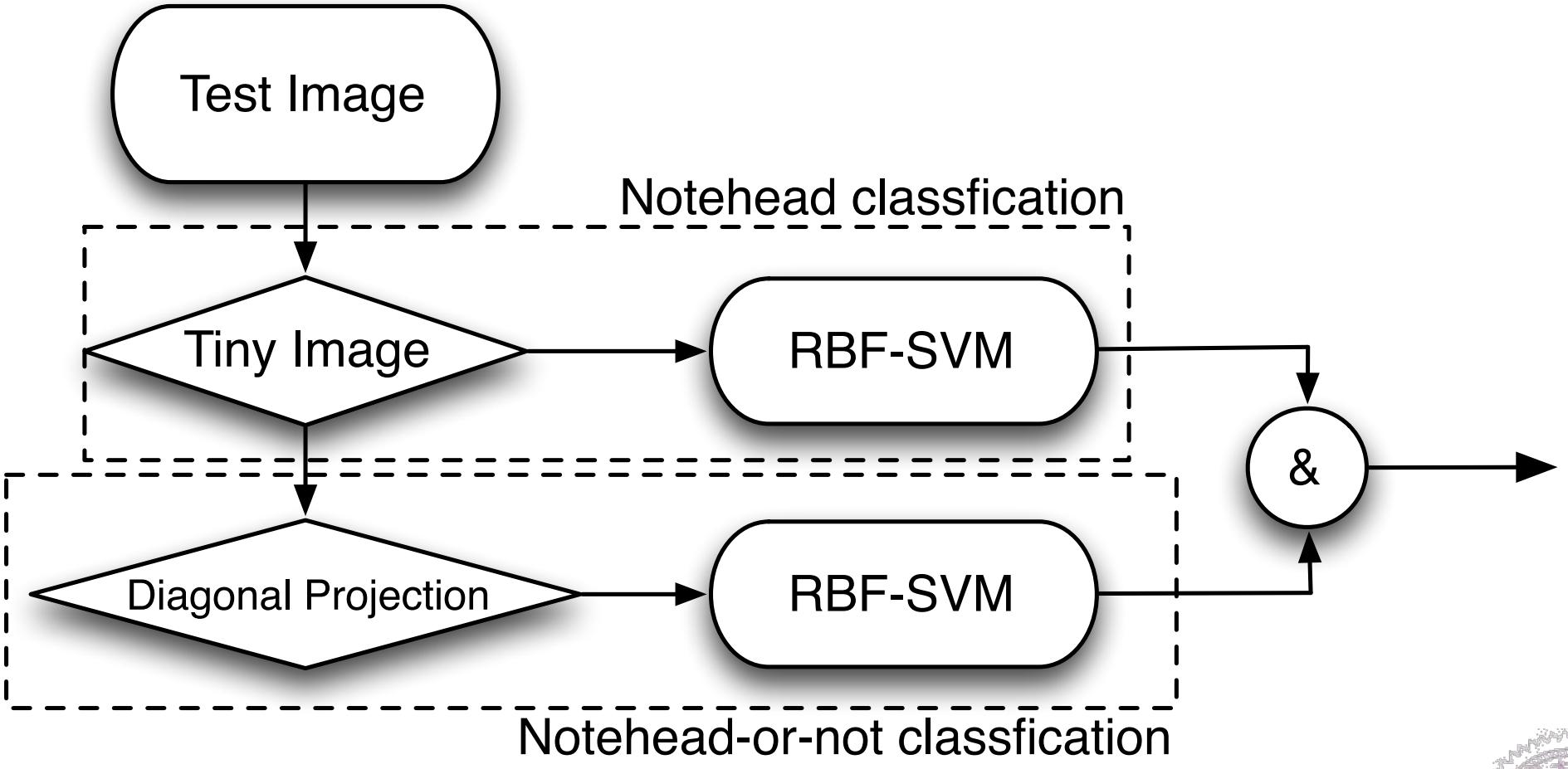
$$p_{\mathrm{u}} = \mathrm{round}\left(\frac{l_3 - y_{\mathrm{min}}}{0.5(w+s)}\right)$$

$$p_{\mathrm{d}} = \mathrm{round}\left(\frac{l_3 - y_{\mathrm{max}}}{0.5(w+s)}\right)$$

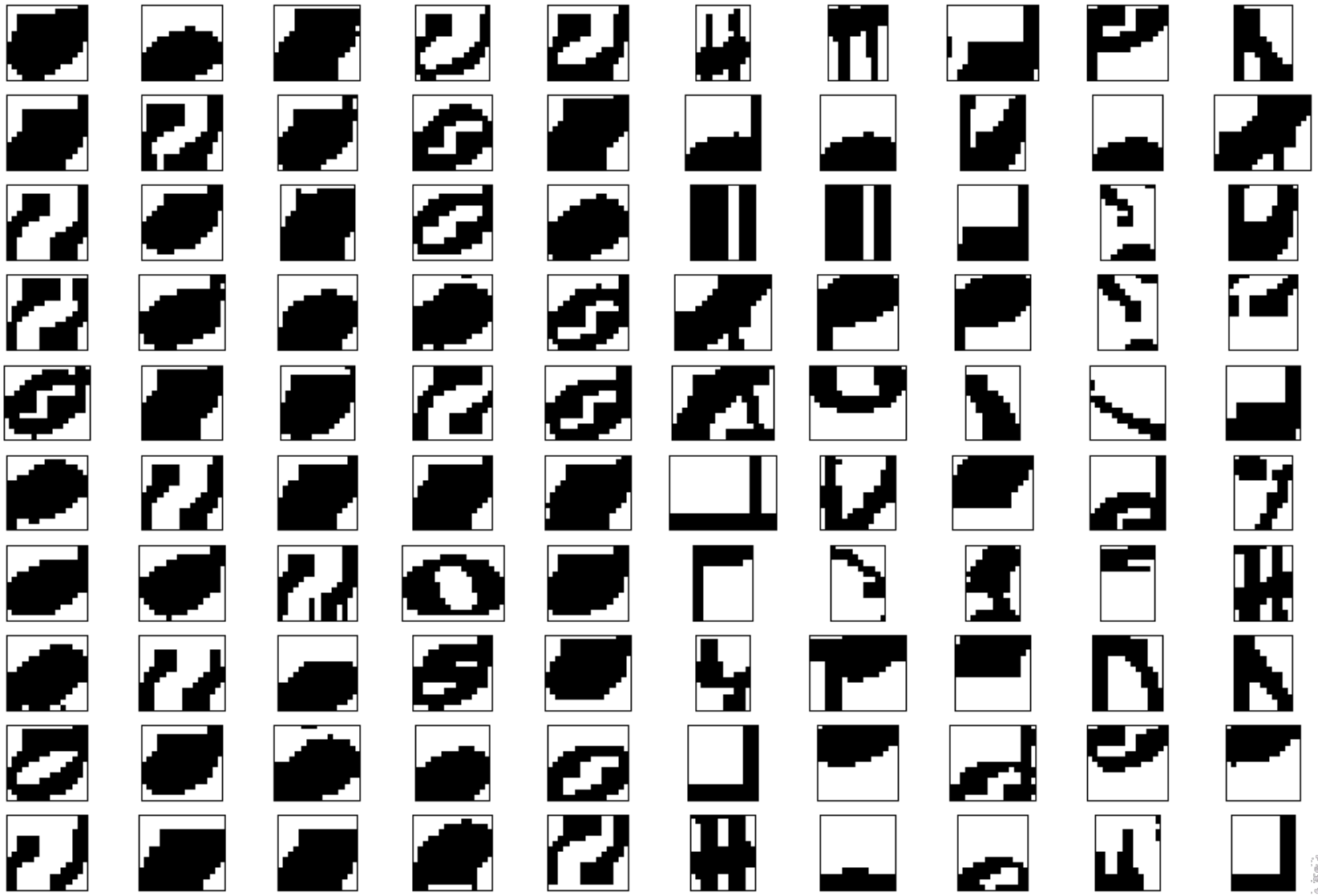
Notehead Detection
Stem-guided



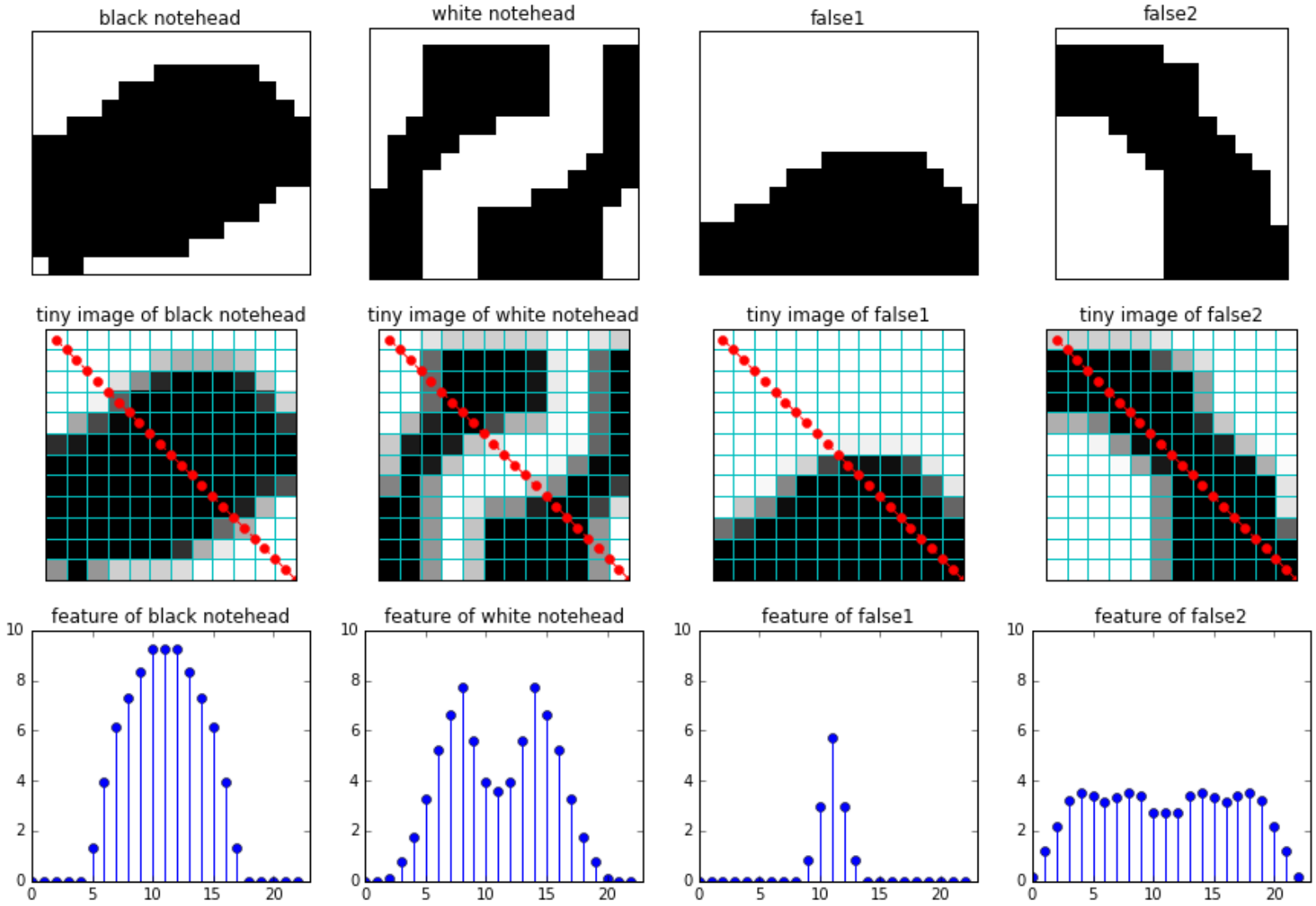
Notehead Detection: Feature



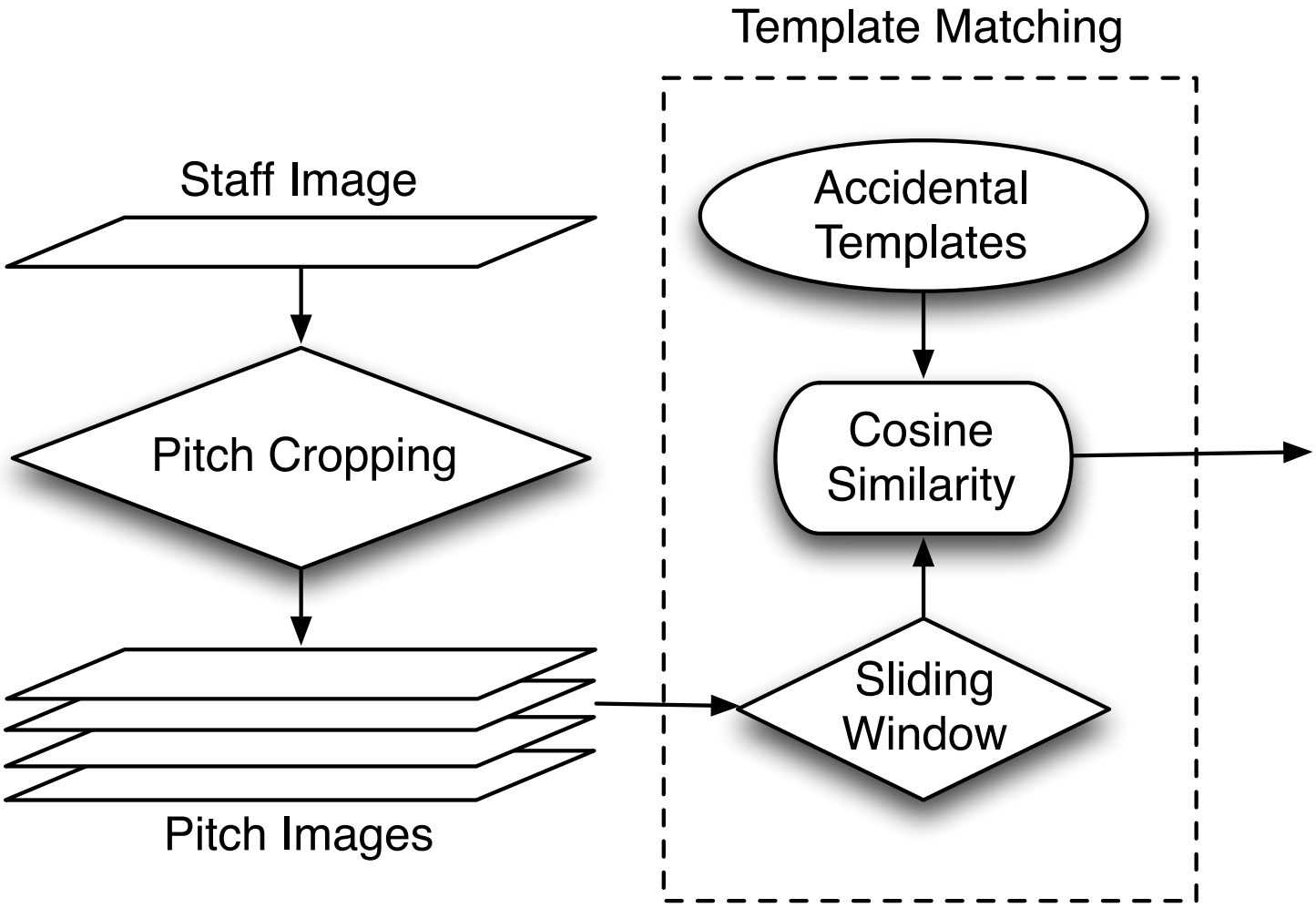
Notehead-or-not Samples



Notehead-or-not Feature

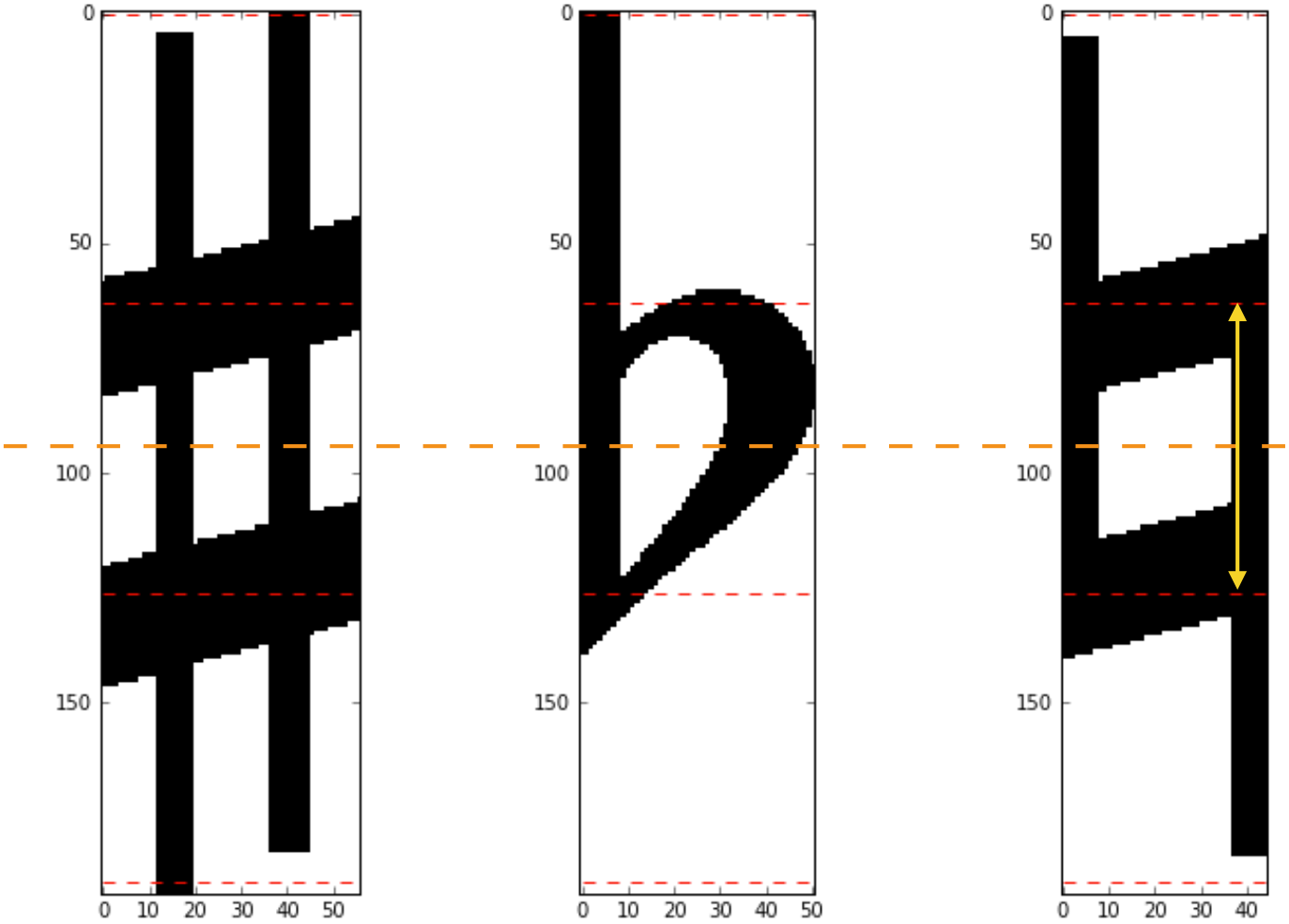


Accidental Detection



Accidental Templates

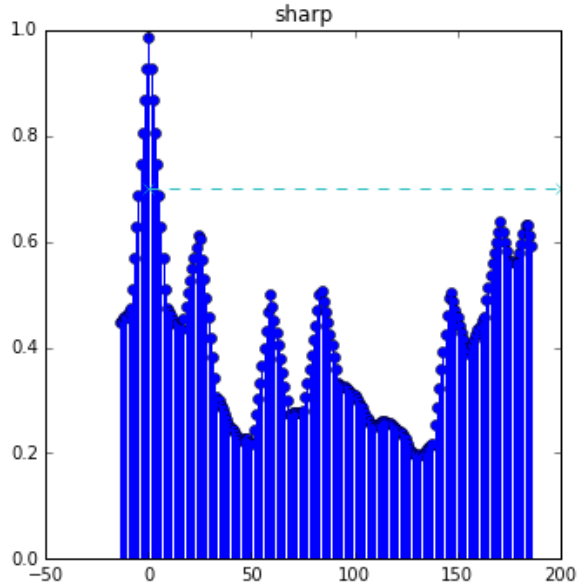
$$\text{score} = \frac{\langle I,\ I_{\text{template}} \rangle}{\|I\| \|I_{\text{template}}\|}$$

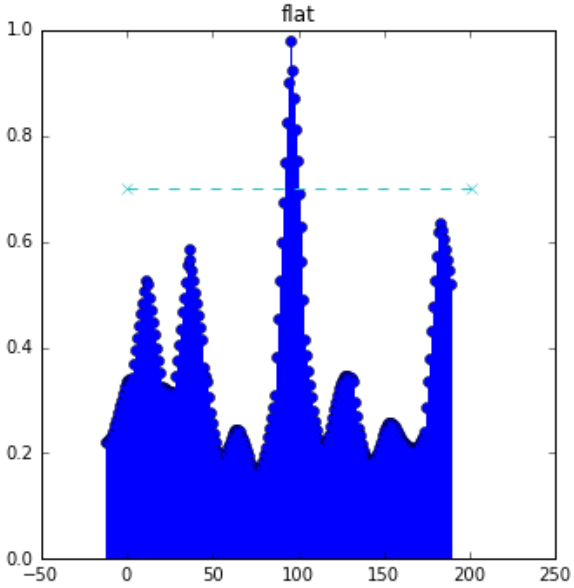


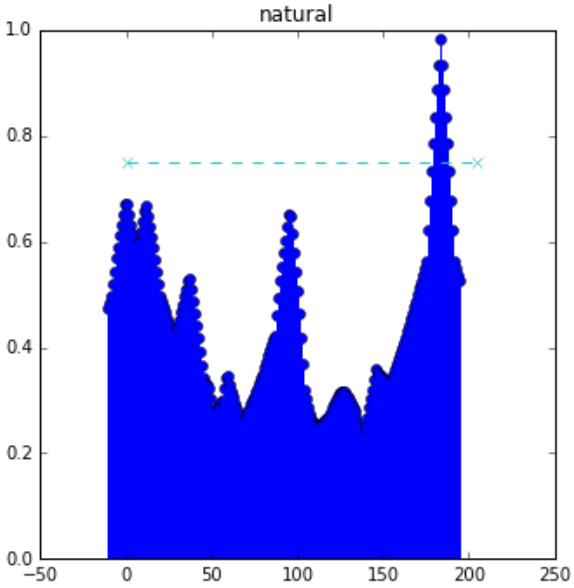
the margin between adjacent stafflines

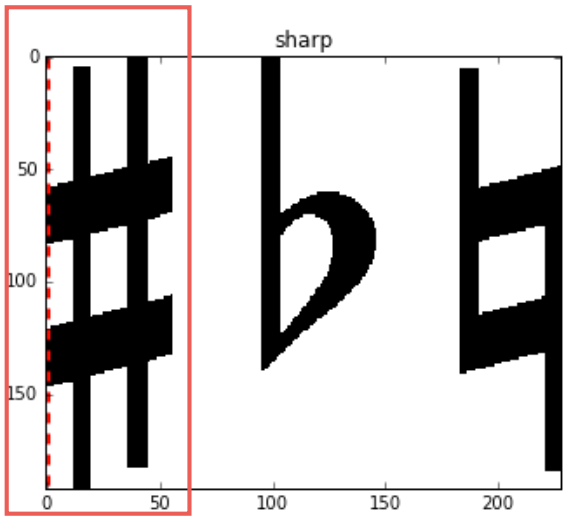
pitch

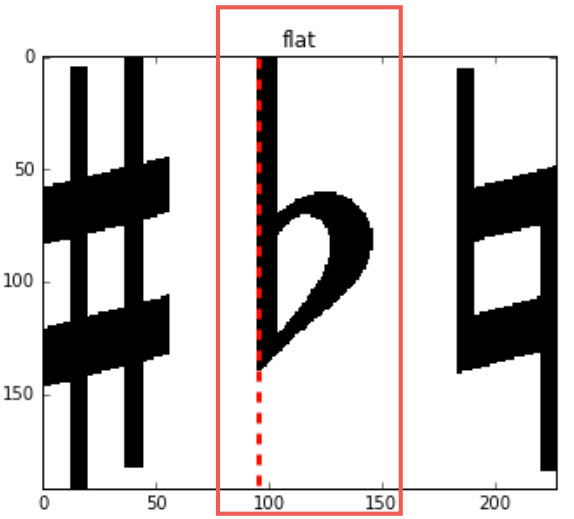
Accidental Detection
Template Matching

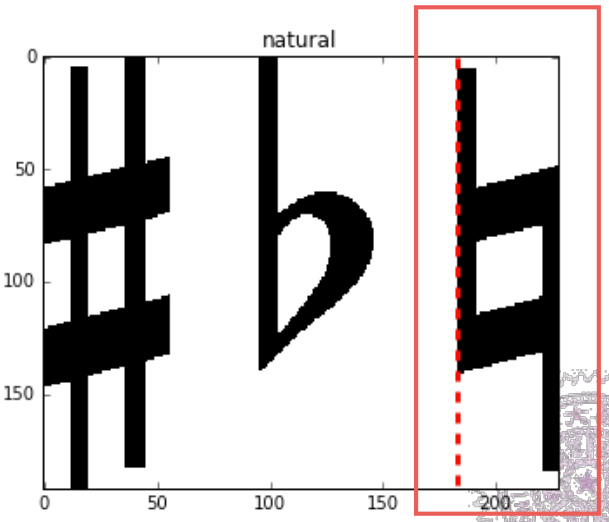






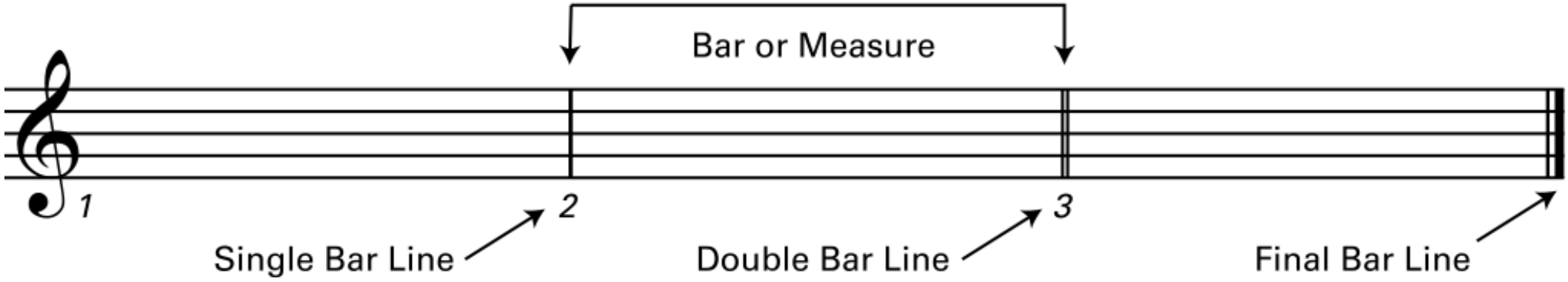


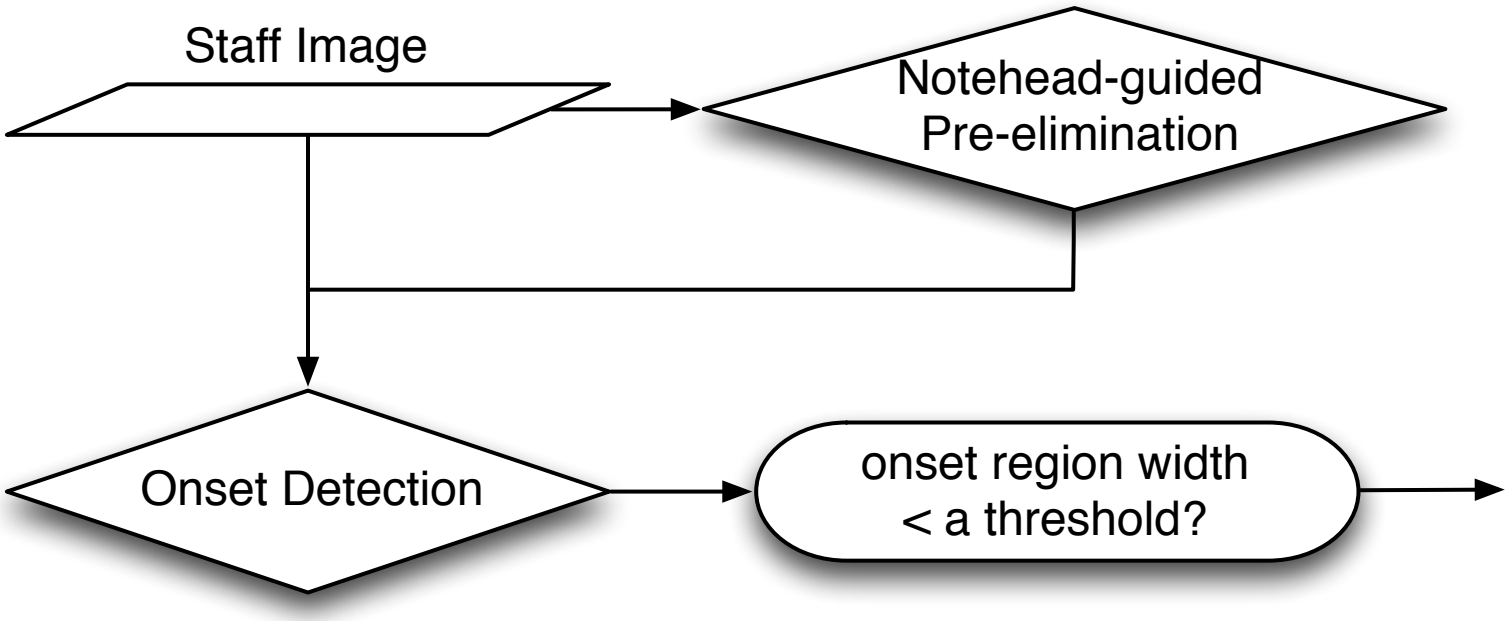




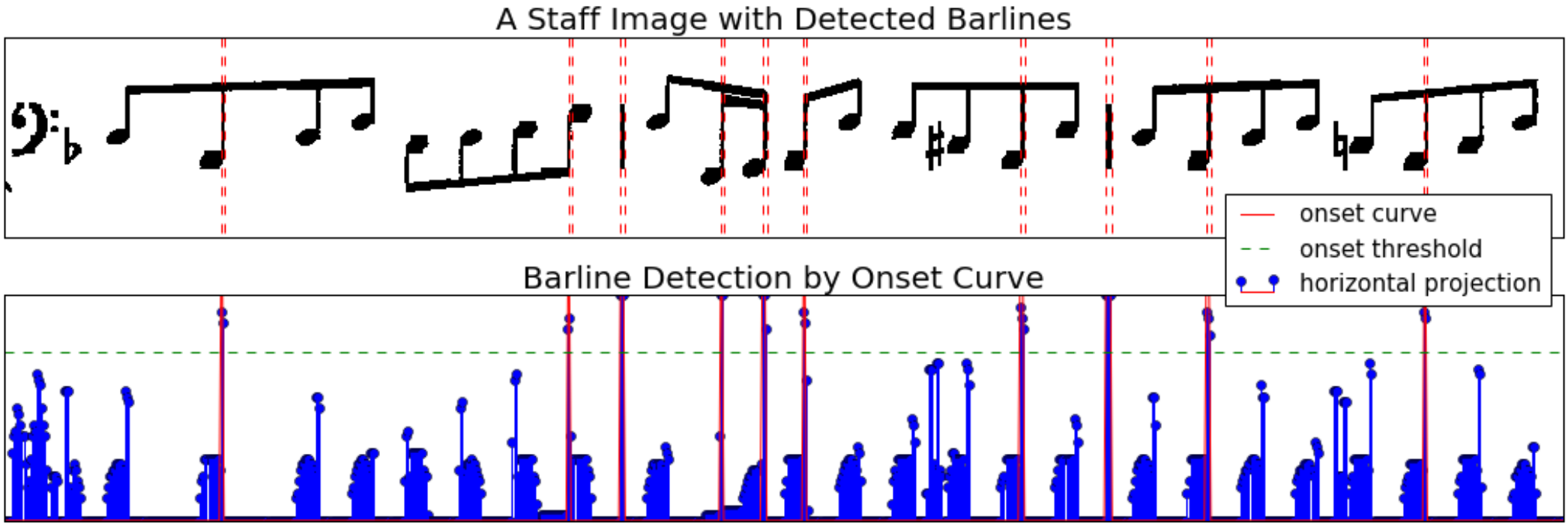
Recognition
Other Symbols

Barline Detection

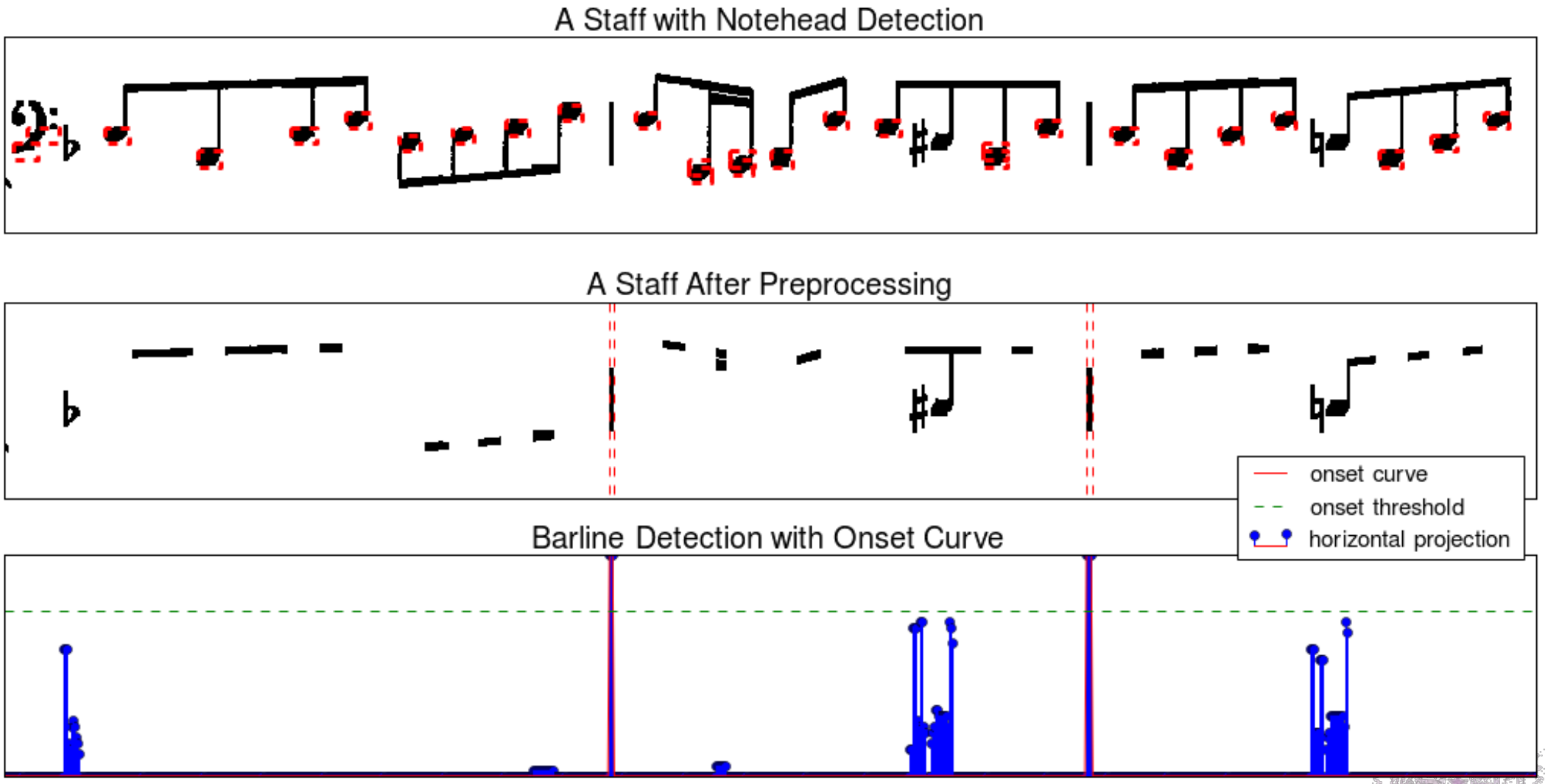




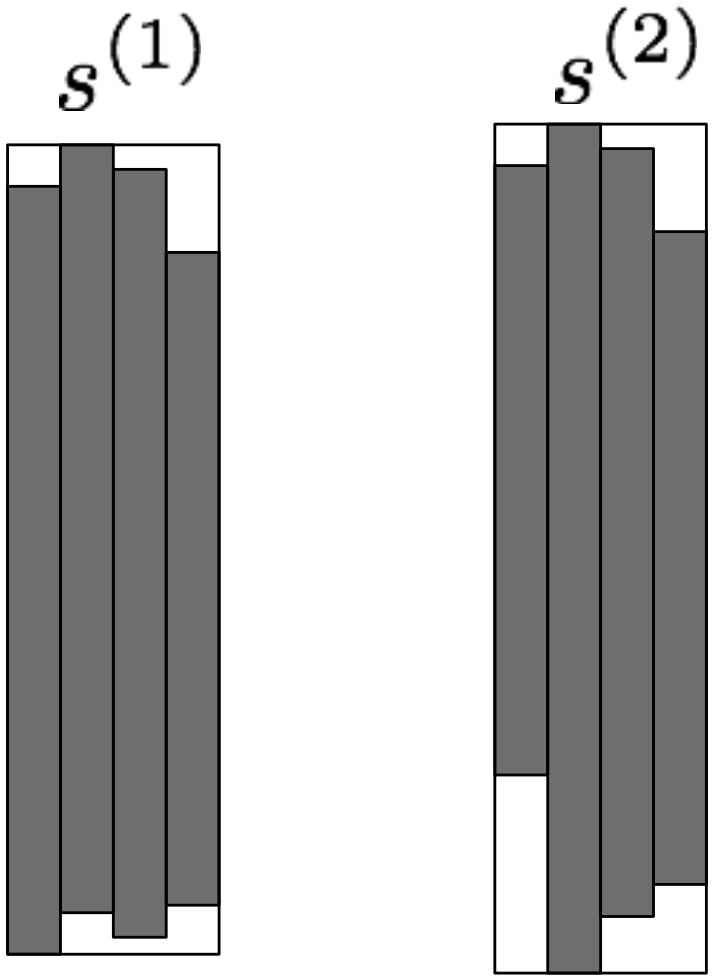
Barline Detection



Barline Detection with Preprocessing



Stem Detection



Algorithm 1 Procedure of finding vertical lines

Input: A gray-scaled image I
Output: A set of 4-tuples

$\mathscr{V} \leftarrow []$
for each $c \leftarrow$ the x^{th} column of I **do**
  if $c(y) = 1, \forall y = y_{\min}, y_{\min}+1, \cdots, y_{\max}$ and $y_{\max} - y_{\min} \geq 2(w+s)$ **then**
    Append $(x, y_{\min}, y_{\max})$ to $\mathscr{V}$
  end if
end for
$\mathscr{R} \leftarrow []$
for each $(x, y_{\min}, y_{\max}) \in \mathscr{V}$ **do**
  if $\mathscr{R}$ is empty **then**
    Append $(x, x, y_{\min}, y_{\max})$ to $\mathscr{R}$
    Continue
  end if
  $r \leftarrow$ the reference of the last element of $\mathscr{R}$
  if $r_2 = x - 1$ **then**
    Update $r_2 \leftarrow x$
    Update $r_3 \leftarrow \min(r_3, y_{\min})$
    Update $r_4 \leftarrow \max(r_4, y_{\max})$
  else
    Append $(x, x, y_{\min}, y_{\max})$ to $\mathscr{R}$
  end if
end for
Return $\mathscr{R}$

Beam Grouping

Algorithm 2 Procedure of beam grouping

Input: the image I and the matrix M_{vl}

Output: A set of $(\mathscr{S}, \delta)$'s where $\mathscr{S}$ is the collection of stems that belong to the same beam and $\delta = \begin{cases} 0, & \text{the beam runs through } y_{\min}\text{'s} \\ 1, & \text{the beam runs through } y_{\max}\text{'s} \end{cases}$ is a binary symbol referring to the beam is upper or lower.

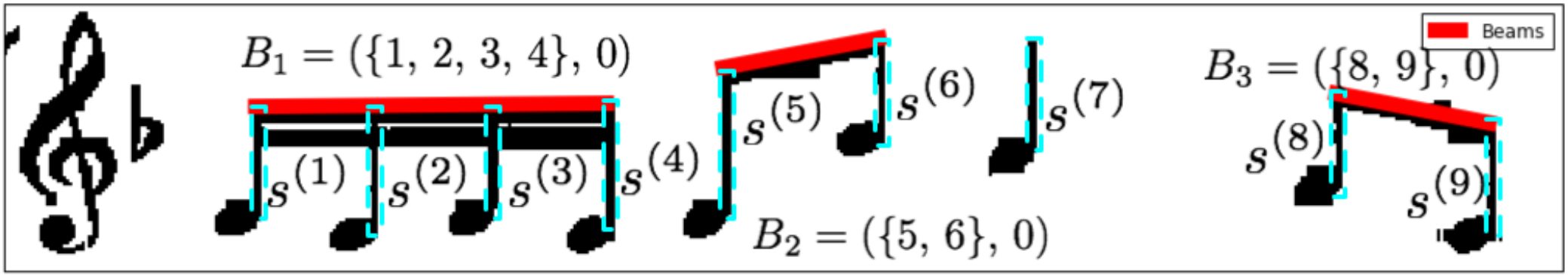
$\mathscr{G} \leftarrow []$
for $i \leftarrow 1, \cdots, |\mathscr{R}_{\text{vl}}|$ **do**
  $s^{(i)} \leftarrow$ the i^{th} column of M_{vl}
  for $j \leftarrow i+1, \cdots, |\mathscr{R}_{\text{vl}}|$ **do**
    $s^{(j)} \leftarrow$ the j^{th} column of M_{vl}
    $p_1 \leftarrow (s_1^{(i)}, s_3^{(i)}), p_2 \leftarrow (s_1^{(j)}, s_3^{(j)})$
    if $L_I(p_1, p_2) \geq 0.9(p_{2x} - p_{1x})$ **then**
      Append $(i, j, 0)$ to $\mathscr{G}$
      Break
    end if
    $p_1 \leftarrow (s_1^{(i)}, s_4^{(i)}), p_2 \leftarrow (s_1^{(j)}, s_4^{(j)})$
    if $L_I(p_1, p_2) \geq 0.9(p_{2x} - p_{1x})$ **then**
      Append $(i, j, 1)$ to $\mathscr{G}$
      Break
    end if
  end for
end for

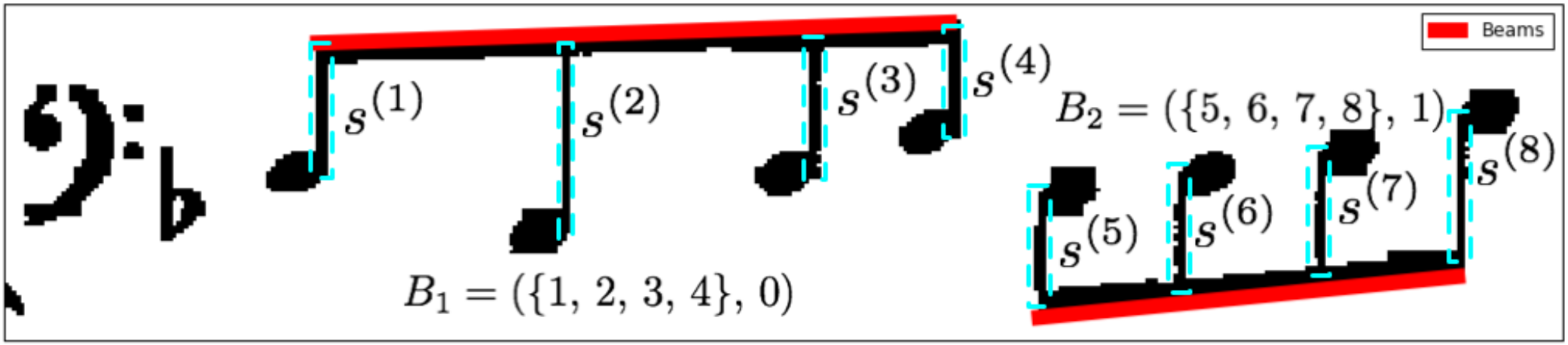
$\mathscr{S}_{\text{beam}} \leftarrow []$
for each $(i, j, \delta) \in \mathscr{G}$ **do**
  if $\mathscr{S}_{\text{beam}}$ is empty **then**
    Append $(\{i, j\}, \delta)$ to $\mathscr{S}_{\text{beam}}$
    Continue
  end if
  $(\mathscr{S}_{\text{last}}, \delta_{\text{last}}) \leftarrow$ the last element of $\mathscr{S}_{\text{beam}}$
  $j_{\text{last}} \leftarrow \max(\mathscr{S}_{\text{last}})$
  if $j_{\text{last}} = i$ **then**
    Add j to $\mathscr{S}_{\text{last}}$
    Update the last element of $\mathscr{S}_{\text{beam}}$
  else
    Append $(\{i, j\}, \delta)$ to $\mathscr{S}_{\text{beam}}$
  end if
end for
Return $\mathscr{S}_{\text{beam}}$

$$l_{p_1,p_2}(x) = y,$$

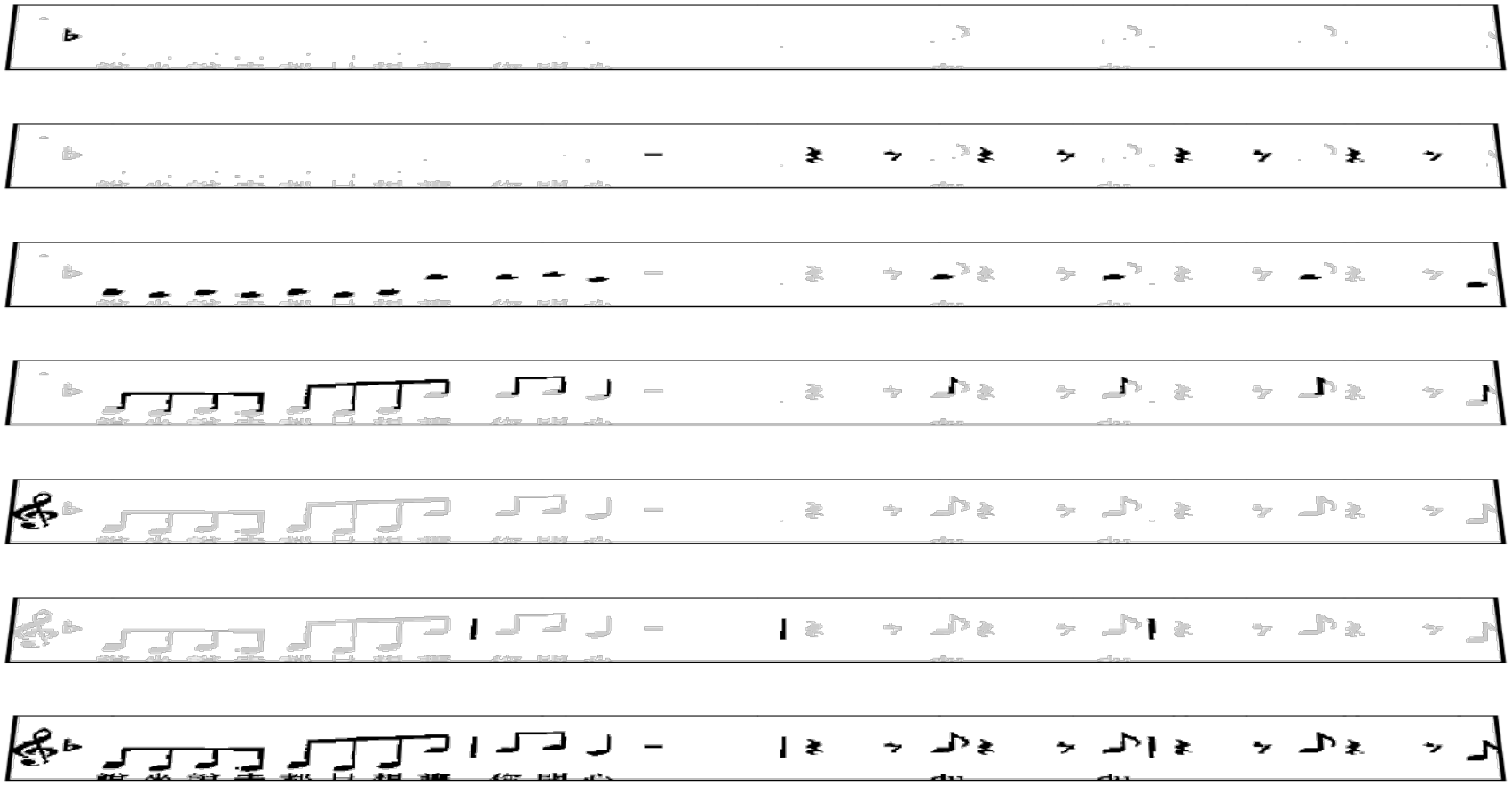
$$L_I(p_1, p_2) = \sum_{x=p_{1x}}^{p_{2x}} I(x, l_{p_1,p_2}(x)),\ p_{1x} \leq p_{2x}$$

Beam Grouping





Hierarchical Detection



Summary of Detection on Different Symbols

	Detection Approach	Locating Symbols	Parameters	Working Images
Barlines	Horizontal Projection	Onset Curve	width, onset threshold	Staff
Clefs	Double-stage SVM	Onset Curve	merge, onset threshold	Measure
Rests	Template Matching	Onset Curve	similarity, onset threshold	Measure
Accidentals	Template Matching	Sliding Window	similarity threshold	Pitch
Noteheads	Double-stage SVM	Onset Curve	onset threshold	Pitch
		Stem-guided	N/A	Measure
Stems	RLE	Sliding Window	length threshold	Measure
Beams	Line Integral	Stem-guided	line integral threshold	Measure

Summary of Features on Different Symbols

	Features
Barlines	RLE of Horizontal Projection
Clefs	Tiny Image, Horizontal/Vertical Projection and Centroid
Rests	Cosine Similarity
Accidentals	Cosine Similarity
Noteheads	Tiny Image, Diagonal Projection
Stems	RLE
Beams	Rectangles of Stems

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] R. Randriamahefa, J. P. Cocquerez, F. Fluhr, C. Pepin, and S. Philipp,"Printed music recognition," in *Proc. the Second Int. Conf. on Document Analysis and Recognition*, pp. 898–901, 1993.

[8] K. T. Reed and J. Parker, "Automatic computer recognition of printed music," in *Proc. the 13th Int. Conf. on Pattern Recognition*, vol. 3, p. 803–807, 1996.

[9] P. Bellini, I. Bruno, and P. Nesi, "Optical music sheet segmentation," in *Proc. the First Int. Conf. on Web Delivering of Music*, pp. 183–190, 2001.

[10] H. Miyao, "Stave extraction for printed music scores," *Intelligent Data Engineering and Automated Learning-IDEAL 2002, H. Yin, N. Allinson, R. Freeman, J. Keane, and S. Hubbard, Eds. Springer*, pp. 621–634, 2002.

[11] F. Rossant and I. Bloch, "Robust and adaptive OMR system including fuzzy modeling, fusion of musical rules, and possible error detection," *EURASIP J. on Advances in Signal Processing*, vol. 1, 2007.

[12] 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.

References

[13] J. d. S. Cardoso, A. Capela, A. Rebelo, C. Guedes, and J. Pinto da Costa,"Staff detection with stable paths," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 31, p. 1134–1139, 2009.

[14] A. Dutta, U. Pal, A. Fornés, and J. Lladós, "An efficient staff removal approach from printed musical documents," in *Proc. the 20th Int. Conf. on Pattern Recognition*, p. 1965–1968, IEEE Computer Society, 2010.

[15] A. Rebelo, G. Capela, and J. S. Cardoso, "Optical recognition of music symbols: a comparative study," *Int. J. Document Analysis Recognition*, vol. 13, pp. 19–31, 2010.

[16] I. Leplumey, J. Camillerapp, and G. Lorette, "A robust detector for music staves," in *Proc. the Int. Conf. on Document Analysis and Recognition*, p. 203–210, 1993.

[17] Q.-A. Arshad, W. Z. Khan, and Z. Ihsan, "Overview of algorithms and techniques for optical music recognition."

[18] I. Fujinaga, *Adaptive Optical Music Recognition*. PhD thesis, Faculty of Music, McGill University, Montréal, Canada, 1996.

[19] 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.

[20] T. Kanungo, R. M. Haralick, H. S. Baird, W. Stuezle, and D. Madigan, "A statistical, nonparametric methodology for document degradation model validation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, pp. 1209–1223, 2000.

[21] L. Pugin, "Optical music recognition of early typographic prints using hidden Markov models," in *Proc. the Int. Society for Music Information Retrieval*, pp. 53– 56, 2006.

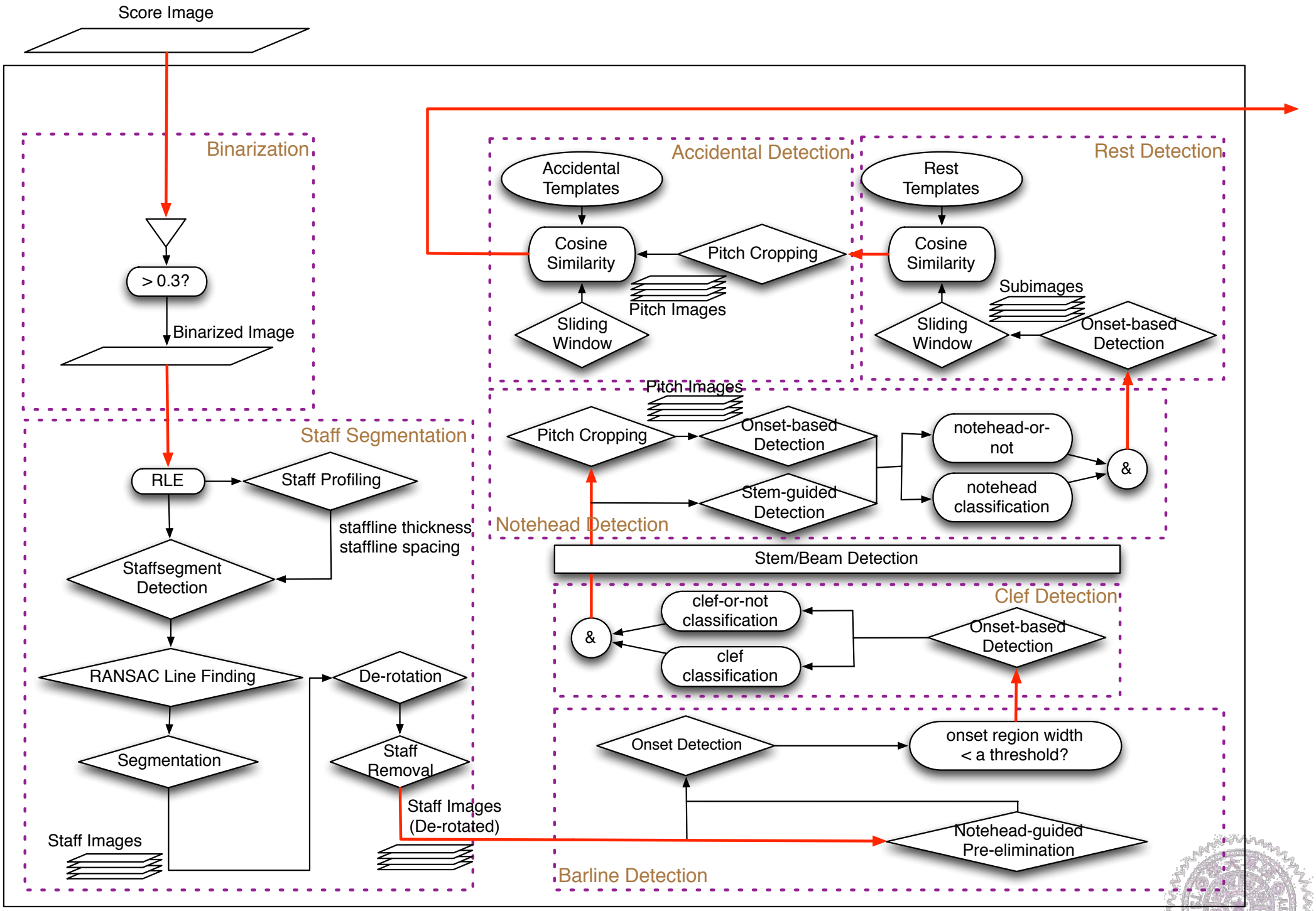
[22] A. Fornés, S. Escalera, J. L Ladòs, G.Sànchez, P.Radeva, and O.Pujol,"Handwritten symbol recognition by a boosted blurred shape model with error correction," in *Proc. the 3rd Iberian Conf. on Pattern Recognition and Image Analysis, Part I. Springer, Berlin*, pp. 13–21, 2007.

[23] H. Miyao and Y. Nakano, "Head and stem extraction from printed music scores using a neural network approach," in *Proc. Third Int. Conf. on Document Analysis and Recognition*, pp. 1074–1079, IEEE Computer Society, 1995.

[24] L. Chen and C. Raphael, "Human-directed optical music recognition," in *Electronic Imaging, Document Recognition and Retrieval XXIII*, pp. 1–9, Society for Imaging Science and Technology, 2016.

[25] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting applications to image analysis and automated cartography," in *Proc. Image Understanding Workshop*, pp. 71–88, 1980.

[26] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, *A Practical Guide to Support Vector Classification*. National Taiwan University, Taipei 106, Taiwan, 2016.


Score Image
Binarization
> 0.3?
Binarized Image
Staff Segmentation
RLE
Staff Profiling
staffline thickness
staffline spacing
Staffsegment Detection
RANSAC Line Finding
De-rotation
Segmentation
Staff Removal
Staff Images
Staff Images (De-rotated)
Accidental Detection
Accidental Templates
Cosine Similarity
Pitch Cropping
Pitch Images
Sliding Window
Rest Detection
Rest Templates
Cosine Similarity
Subimages
Sliding Window
Onset-based Detection
Pitch Images
Pitch Cropping
Onset-based Detection
Stem-guided Detection
notehead-or-not
notehead classification
&
Notehead Detection
Stem/Beam Detection
Clef Detection
&
clef-or-not classification
clef classification
Onset-based Detection
Barline Detection
Onset Detection
onset region width < a threshold?
Notehead-guided Pre-elimination