[DOCUMENT AND PATTERN RECOGNITION LAB]

R·I·T

An Overview of Current Research into Recognition and Retrieval of Mathematical Notation

Richard Zanibbi, Department of Computer Science,
Rochester Institute of Technology, Rochester, NY, USA *rlaz@cs.rit.edu*

Nov. 10, 2011, IEEE Computer Society Meeting

Quick Notes

Document Recognition and Retrieval XIX

Being held at SPIE Electronic Imaging this January (in San Francisco). Invited Speakers:

- Samy Bengio (Google) - Machine Learning
- Christopher Manning (Stanford) - Information Retrieval

Survey on Math Recognition and Retrieval

R. Zanibbi and D. Blostein (2011) Recognition and Retrieval of Mathematical Notation, *Int'l. Journal of Document Analysis and Recognition*, to appear, available online from www.springerlink.com, 28 pp.

Motivation

Find math in documents (.pdf, scanned images, html, etc.)

If we want to look up a term we check the index.

What if an undergraduate student wants to find the sum of squared error function in their textbook, and does *not* know the name of the function?

Find *similar* expressions

...and related information in technical documents, using a combination of math and keywords ('conventional retrieval')

- e.g. try to locate research papers employing a particular cost metric in an image processing/computer vision paper database

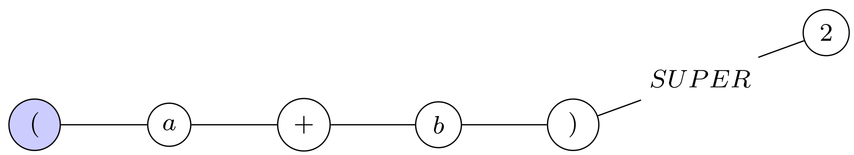
Target Users: initially, *non-experts*

What is Math?

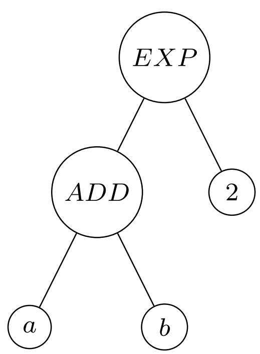
Mathematical Notation

- A *natural* visual language: frequently adapted by authors for their own purposes
  - (consider the overloading of 'f' or 'x') - *dialects*
- Often used to represent quantities (e.g. real numbers) or structures (e.g. matrices, graphs, logical statements), and operations upon them
- History of Math Notation: see Cajori, "History of Mathematical Notations" (2 Vols.), 1929
- To evaluate an expression, need to know assumptions/conventions regarding symbols (e.g. variable bindings), as well as the primitives and operations represented in an expression.

Math Encodings, Example: $(a+b)^2$



(a) Symbol layout tree. The tree is rooted at left ('('). Horizontally adjacency relationship edges are unlabeled



(b) Operator tree. The tree represents the addition of a and b, squared.

Symbol Layout Tree

Defines symbols and their relative positions (e.g. LaTeX, Presentation MathML)

Operator Tree

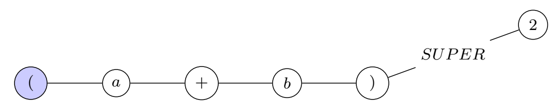
Defines (partial) application of operations to operands (e.g. Content MathML, Maple, Matlab, Mathematica)

Math Encodings, Cont'd

$$(a+b)^2$$

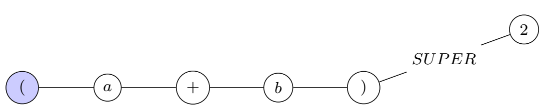
(a) Expression Image

```
<msup>
    <mfenced>
            <mi>a</mi>
            <mo>+</mo>
            <mi>b</mi>
    </mfenced>
    <mn>2</mn>
</msup>
```



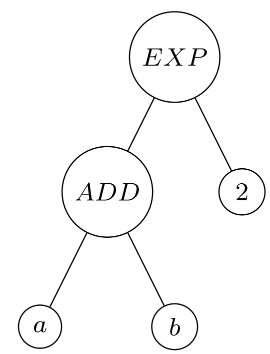
(b) Presentation MathML [10]
(Symbol Layout Tree)

```
(a+b)^2
```



(c) LaTeX
(Symbol Layout Tree)

```
<apply>
    <power/>
    <apply>
        <plus/>
        <ci>a</ci>
        <ci>b</ci>
    </apply>
    <cn>2</cn>
</apply>
```



(d) Content MathML
(Operator Tree)

Encoding Semantics

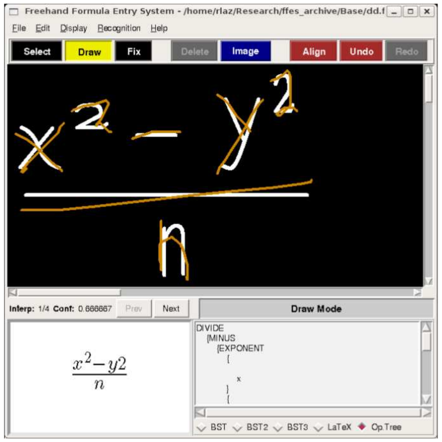
- OpenMath: uses *content dictionaries* to represent computational 'meaning' of notational primitives and operations
- Currently not widely used in practice; some integration with MathML 3

Key Problems

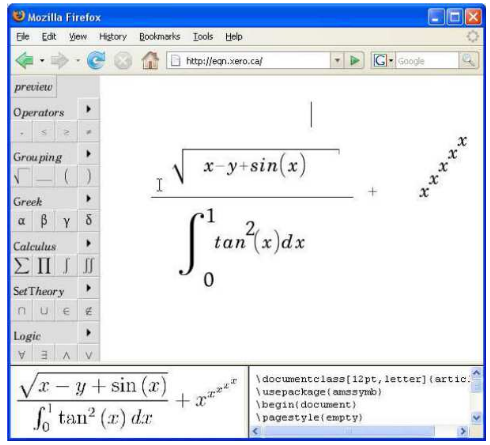
1. Interfaces for Easier Math Entry

2. Recognition of Handwritten and Typeset Math

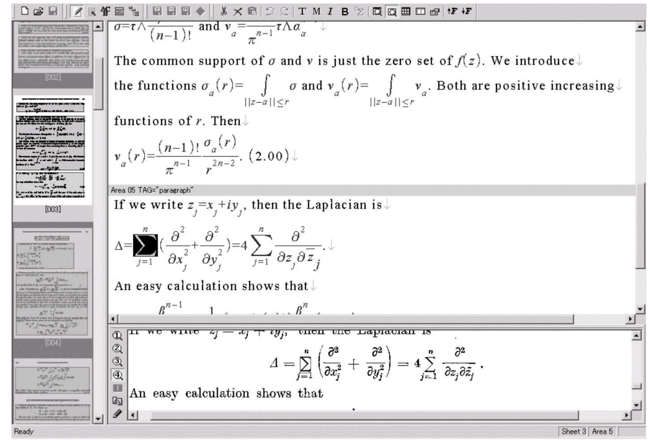
3. Retrieval for Math Notation
   - Mathematical Information Retrieval (MIR)
   - Query-by-Expression, combination with keywords



(a) Freehand Formula Entry System (FFES) [20, 136]

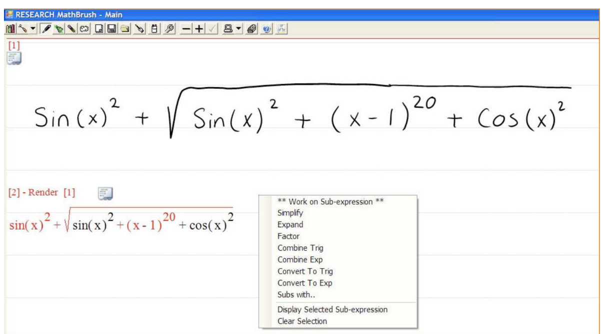


(b) XPRESS [116]

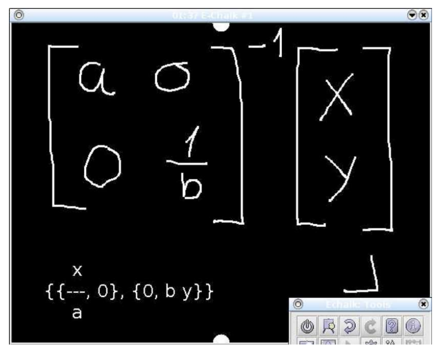


(c) InftyEditor/InftyReader [141]

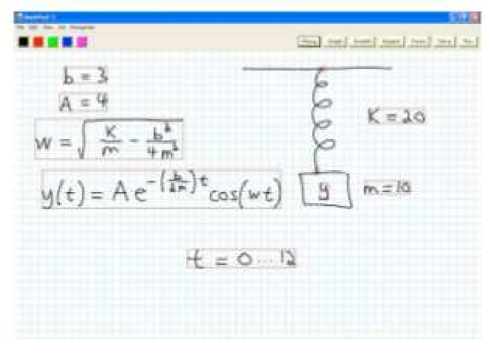
Fig. 1 *Math Entry Systems. FFES is pen-based, XPRESS supports mouse and keyboard entry, and InftyEditor/IntryReader supports OCR, pen, mouse and keyboard entry.*



(a) MathBrush [81]



(b) E-chalk [144]



(c) MathPad^2 [85]

(d) Li, Zeleznik et al. [89]

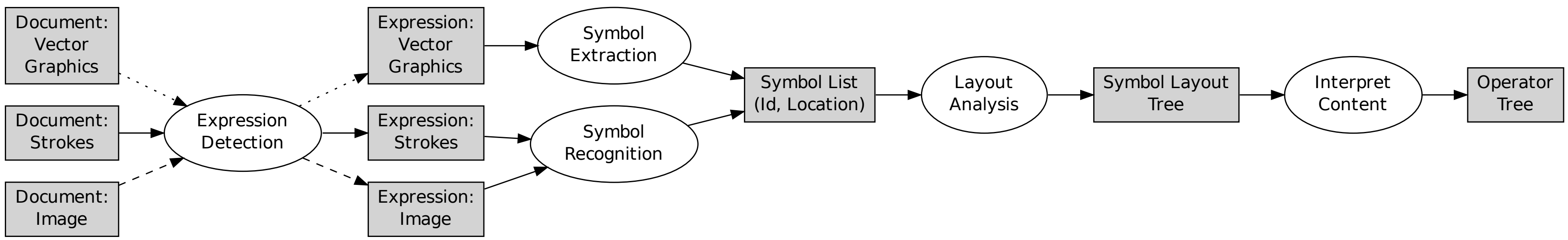
Fig. 2 *Systems for Pen-Based Computer Algebra and Sketching.*

Example: m_{in} system

Prototype Available Online:

http://saskatoon.cs.rit.edu/pen_entry/

Math Recognition



Key Steps

Expression Detection

Symbol Recognition/Extraction

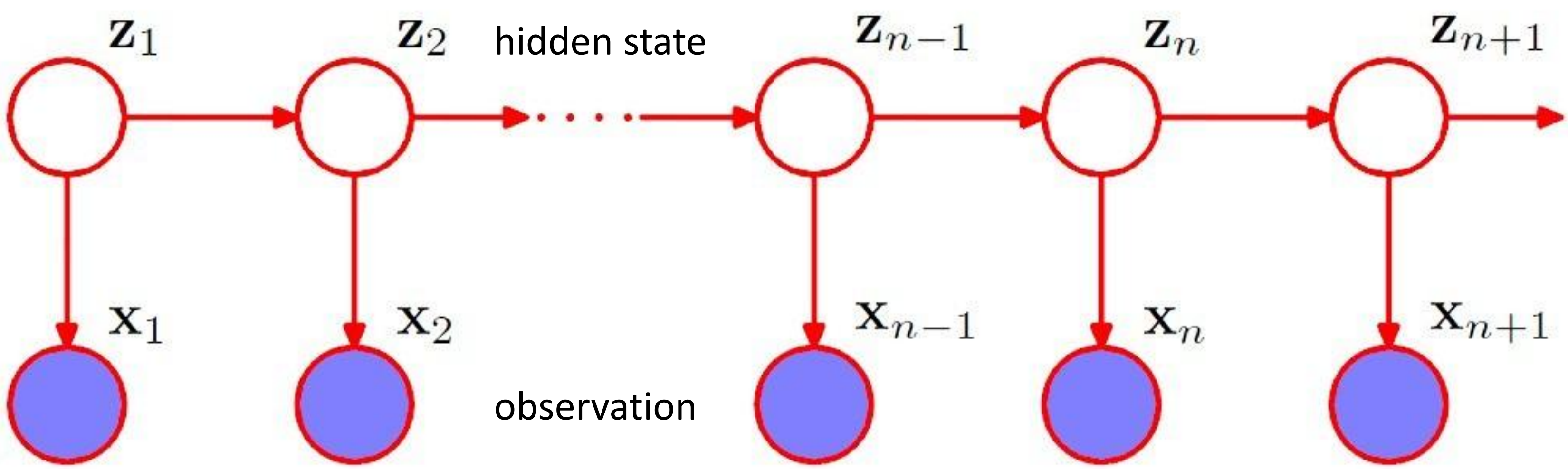
Layout Analysis (Parsing)

Interpretation of Content (Mathematical semantics)

Online Handwritten Math Symbol Recognition (L. Hu, R. Zanibbi)

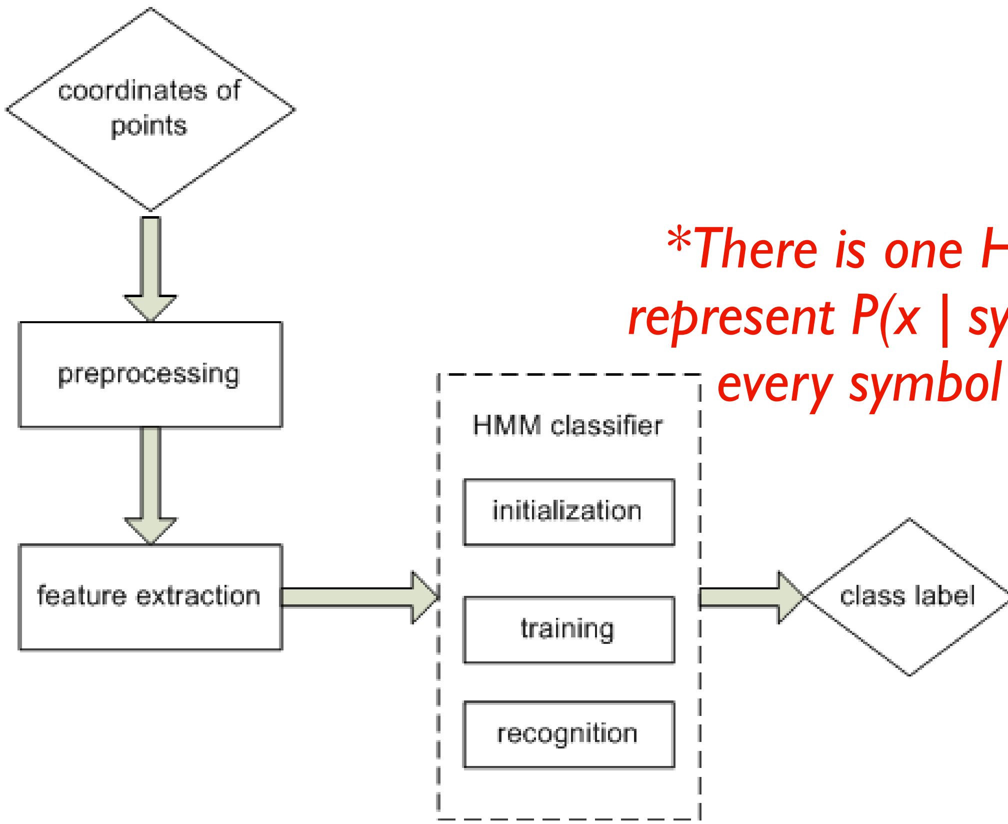
Hidden Markov Model

- An HMM process is a doubly stochastic process [2]
  - hidden state
  - observation



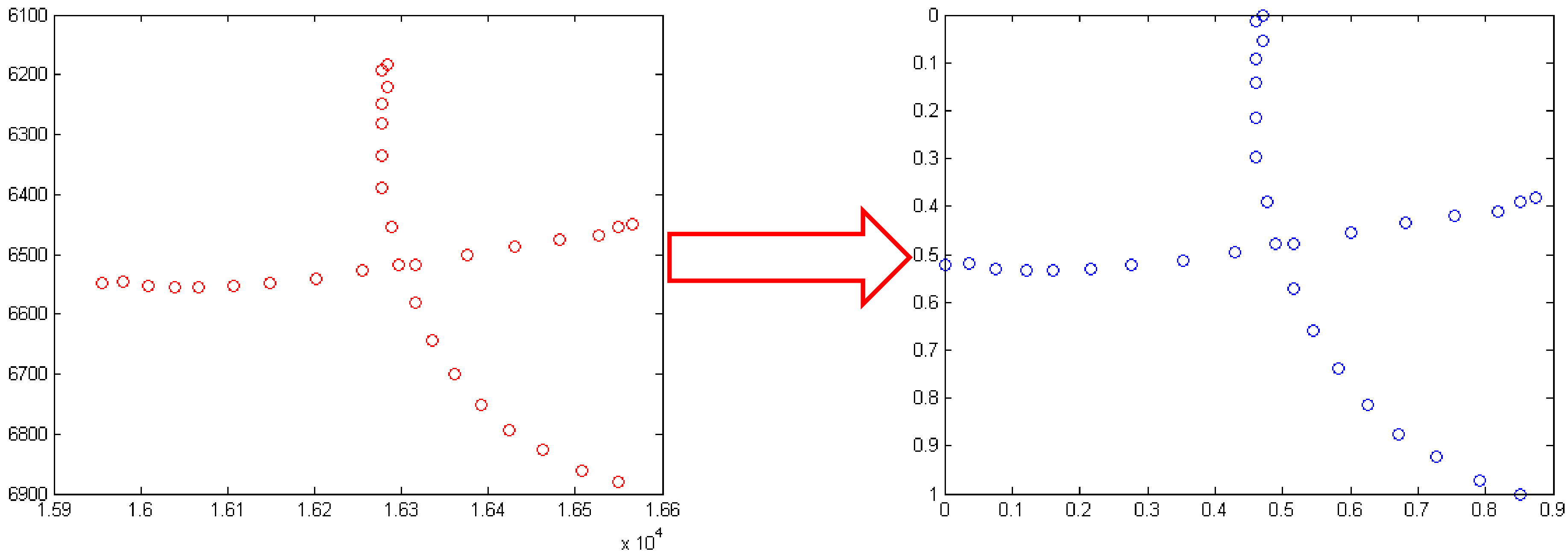
[2] L.Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," In Proc. IEEE, vol. 77, no. 2, pp. 257-286, 1989.

Flow Chart of System

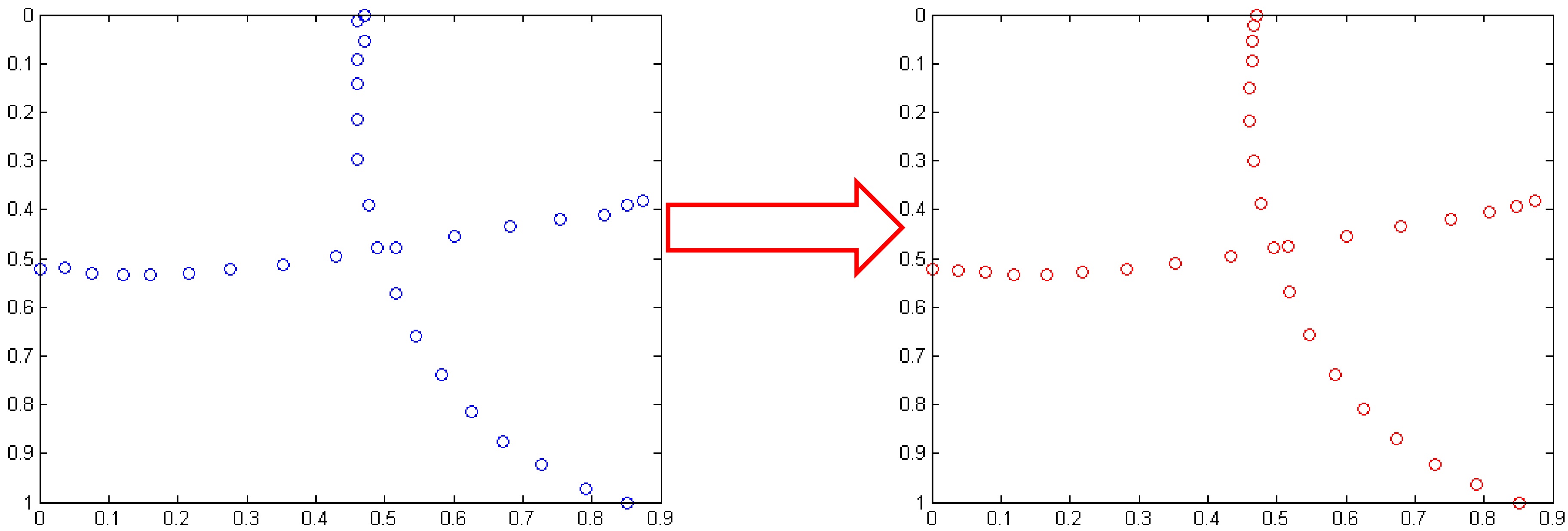


**There is one HMM to represent P(x | symbol) for every symbol type.*

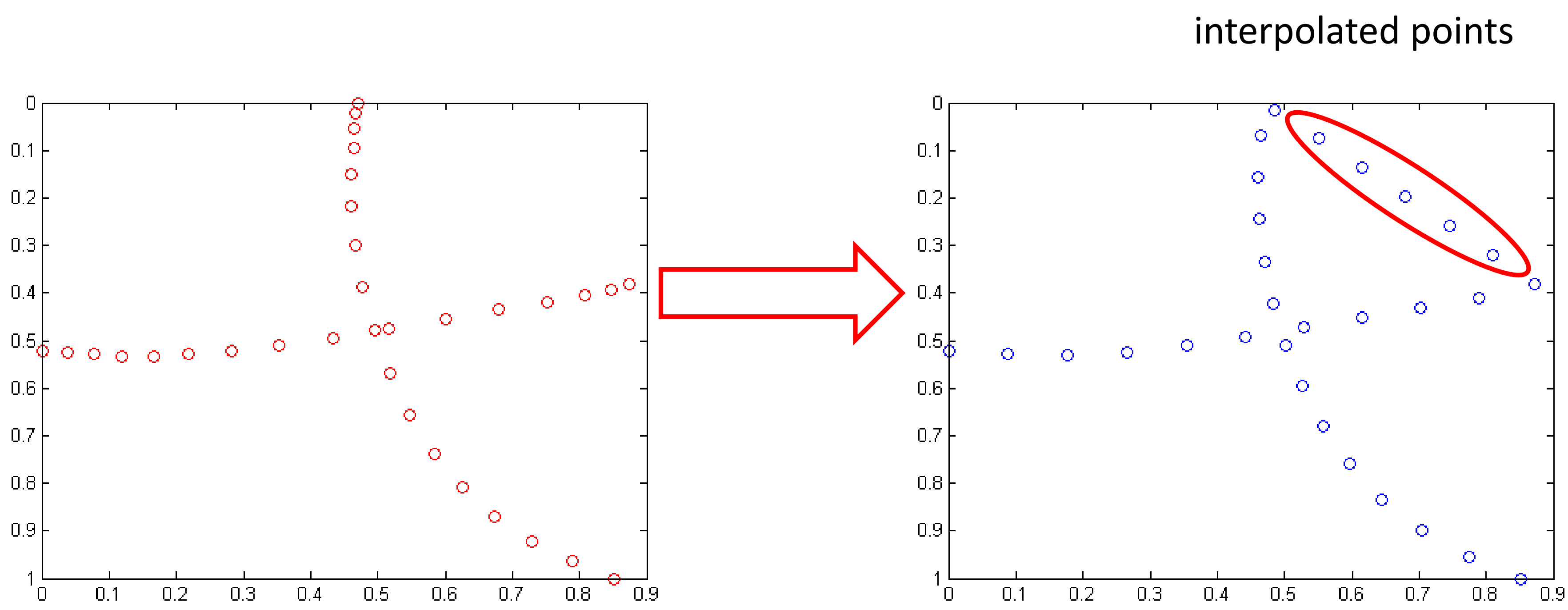
Size Normalization



Smoothing

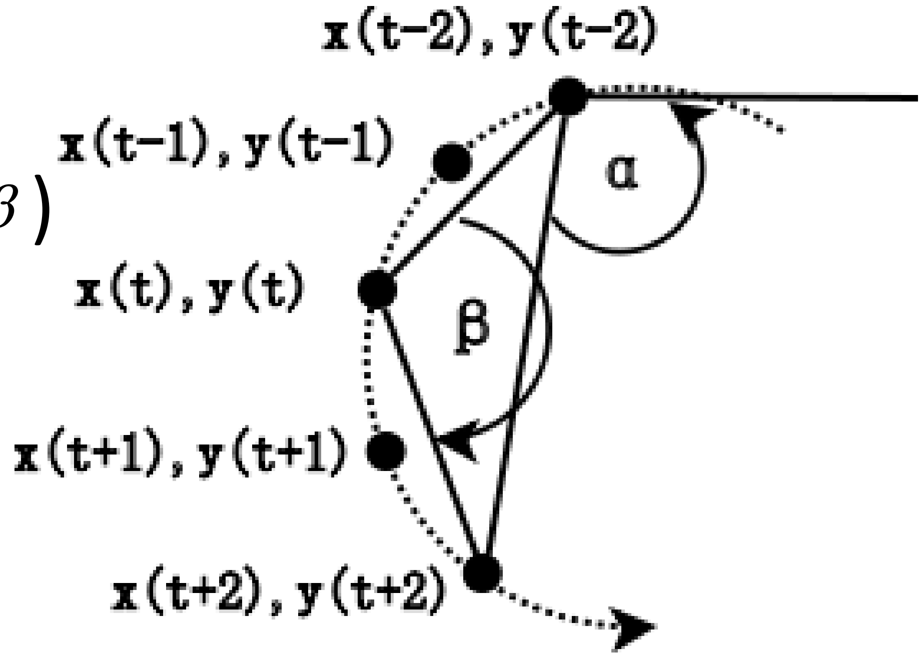


Resampling



Feature Extraction

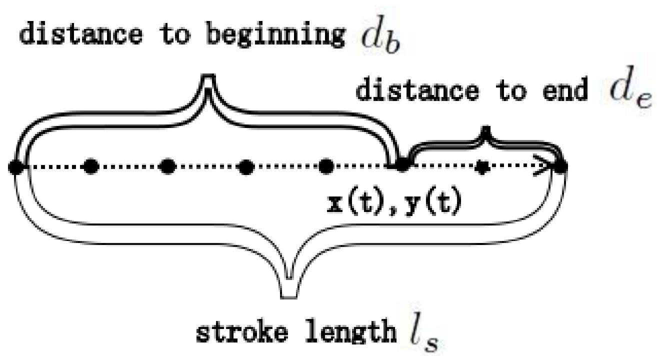
- Normalized y-coordinate
- Pen-up/down
- Cosine of slope (α)
- Sine of curvature (β)



Modified Pen-up/down Feature

Pen-up/down → Normalized distance to stroke edge (NDTSE)

$$NDTSE(s,t) = \begin{cases} 1 - \frac{|d_e - d_b|}{l_s}, & \text{for actual stroke} \\ \frac{|d_e - d_b|}{l_s} - 1, & \text{for interpolated stroke} \end{cases}$$



(Improves recognition by approx. 1%)

Dataset

- A new publicly available, ground-truthed corpus [3]
- 20281 samples for training
- 2202 samples for testing

93 TYPES OF SYMBOLS

0	1	2	3	4	5	6	7	8	9
a	b	c	d	e	f	g	h	i	j
k	l	m	n	o	p	q	r	s	t
u	v	w	x	y	z	A	B	C	D
E	F	G	H	I	J	K	L	M	N
O	P	Q	R	S	T	U	V	W	X
Y	Z	α	β	δ	Δ	ϵ	$=$	γ	Γ
$\geq$	$>$	-	∞	$\int$	[	(	μ	ϕ	π
Π	$+$	ψ	]	ρ	)	σ	Σ	$\surd$	τ
θ	ξ	ζ							

[3] S. MacLean, G. Labahn, E. Lank, M. Marzouk, and D. Tausky, "Grammar-based techniques for creating ground-truthed sketch corpora," International Journal on Document Analysis and Recognition, pp. 1–21, May 2010.

Results

Recognition Rate (over 93 classes): 82.9%

Top-5 Recognition Rate: 97.8%

Compares well with previous results (subset of symbols: single stroke only):

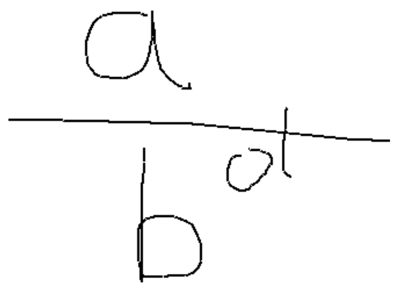
~85% top-1 (HMM-based less than 1% lower)

99% top 5 (2% higher than previous result)

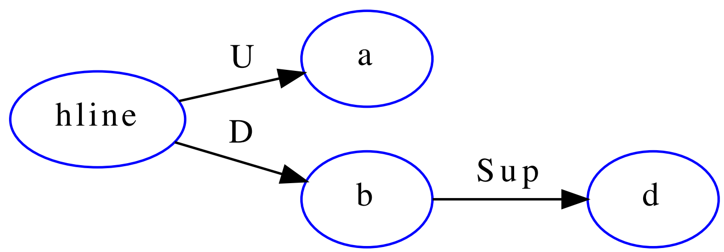
DOCUMENT AND PATTERN RECOGNITION LAB

Recognition Problem Representation and Evaluation

Handwritten Math Recognition Evaluation Using DAGs/Bipartite Graphs



`\frac{a}{b^d}`

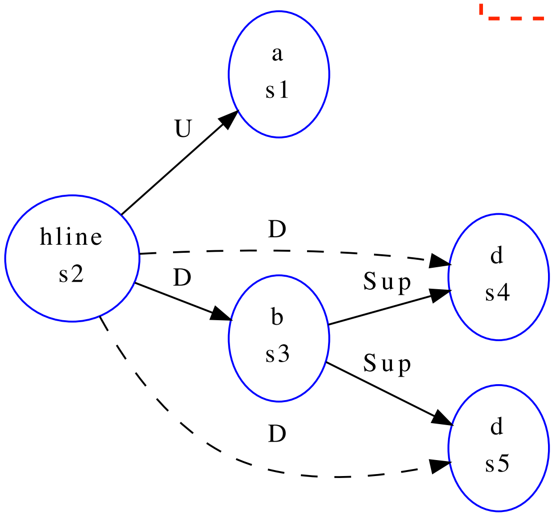


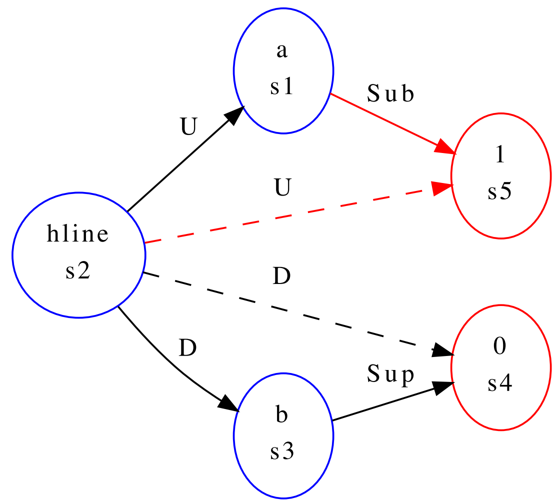
Goal Output ('Ground Truth')

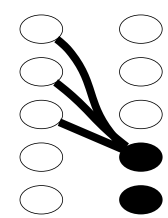
Does not capture:

- Input objects (strokes)
- Stroke segmentation
- Non-local spatial relationships

$$\frac{a_1}{b^0}$$







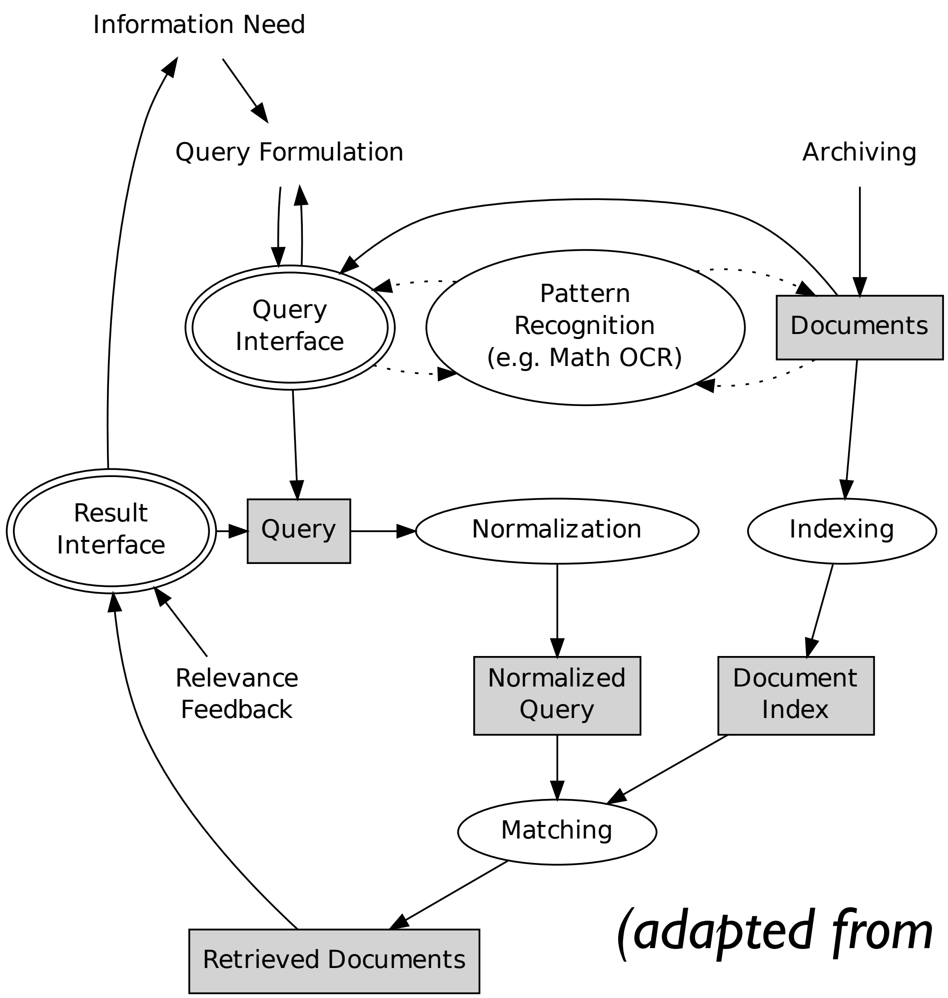
(missing: b -Sup-> d stem)

R. Zanibbi, A. Pillay, H. Mouchere, C. Viard-Gaudin, and D. Blostein. (2011) Stroke-Based Performance Metrics for Handwritten Mathematical Expressions. *Proc. Int'l Conf. Document Analysis and Recognition*, pp. 334-338, Beijing.

DOCUMENT AND PATTERN RECOGNITION LAB

Image-Based Mathematical Information Retrieval (MIR)

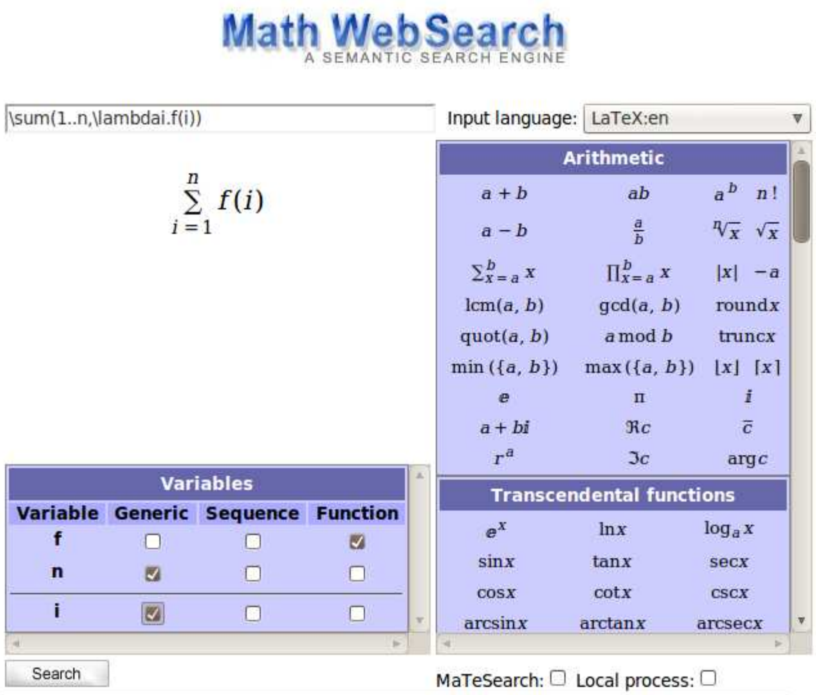
Information Retrieval



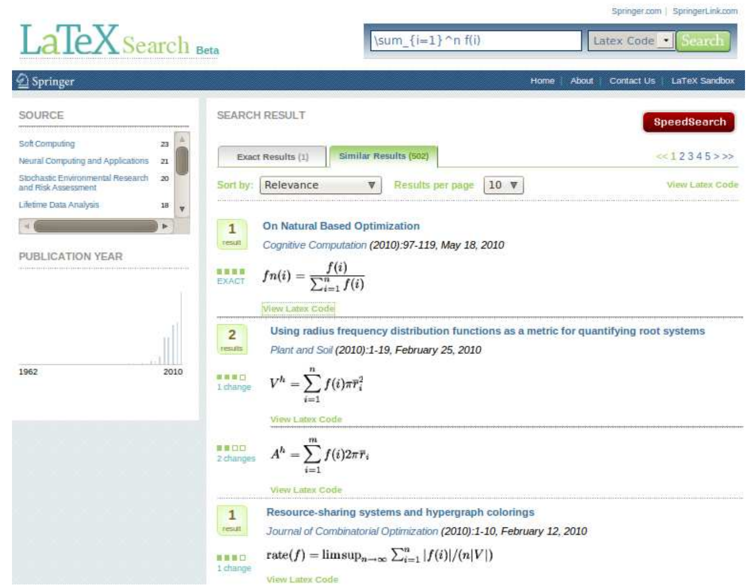
(adapted from Hiemstra, 2009)

Goal: Mathematical Information Retrieval (MIR)

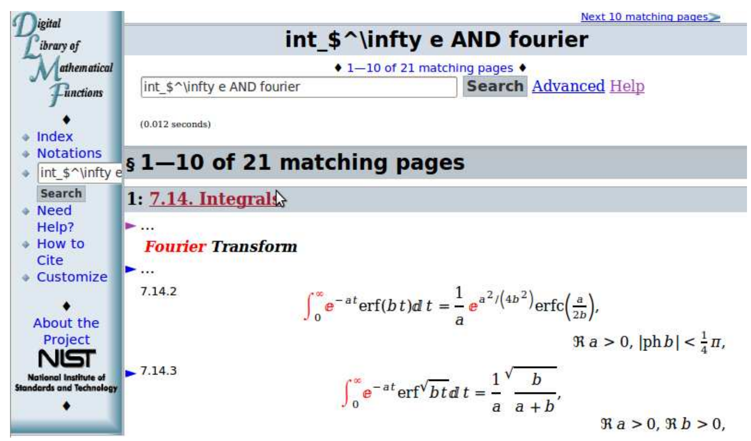
Query-by-Expression



a. Math WebSearch Interface [77, 78]. Queries are constructed via keyboard and templates on the right. Symbol types may also be constrained (bottom left)



b. Springer LaTeX Search. Results may be filtered by clicking on a publication year or source document type



c. NIST Digital Library of Mathematical Functions. Shown are results for a boolean query combining math and keywords [3,102]

Image-Based Querying

Q. Can we use an *image* of a handwritten expression for search?

Li Yu decided to look into this for his MSc (2009-2010), and the answer surprised us.

Math Spotting: Approach

Construct document index using recursive & depth-limited standard X-Y cutting of document image regions, along with smallest maximum upper/lower contour offset from top/bottom of image

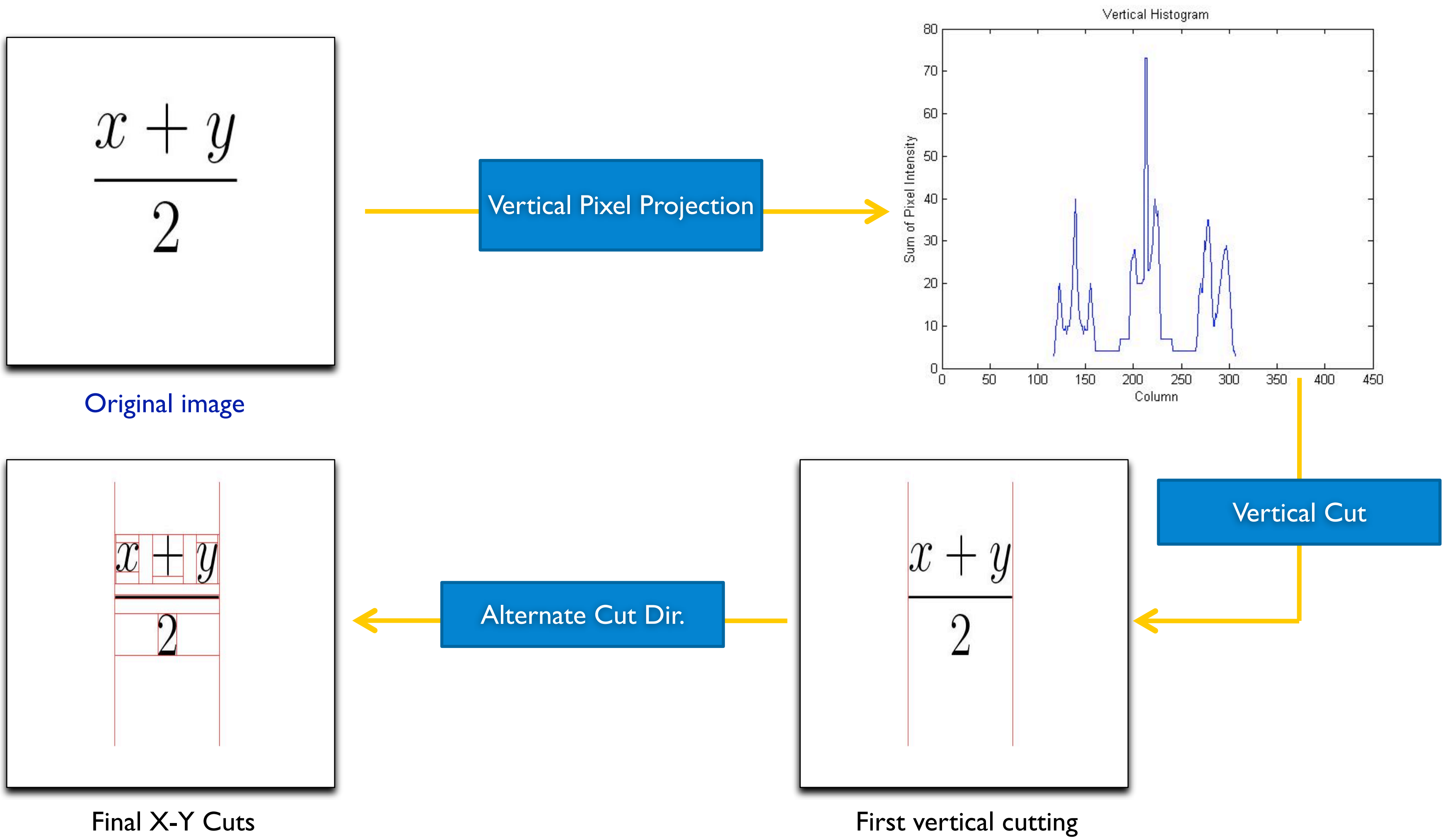
- Inspiration: Projection Profile Cutting of Okamoto et al.
- Regions stored in table contain X-Y cut attributes and image data

Retrieve matches (candidates) using similarity of X-Y structure and differences in largest upper/lower contour offsets

Rank matches by image similarity. Compute similarities via Dynamic Time Warping (DTW): compute cost of transforming upper and lower projection profiles from query to candidate regions

M. Okamoto and B. Miao. *Recognition of mathematical expressions by using the layout structures of symbols.* In Proc. Int'l Conf. Document Analysis and Recognition, volume 1, pages 242–250, Saint-Malo, France, 1991.

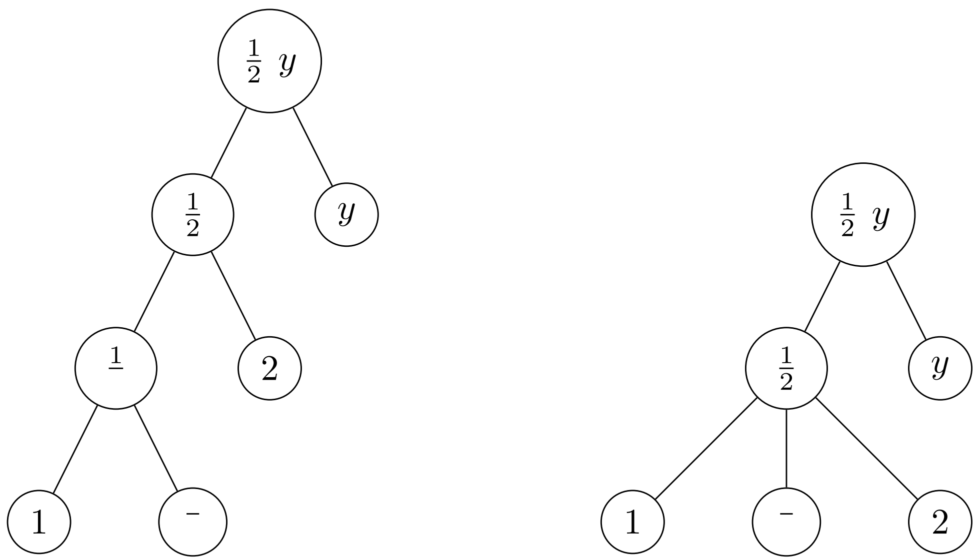
Structural Analysis: X-Y cutting



G. Nagy and S. Seth. *Hierarchical representation of optically scanned documents*. In Proc. Int'l Conf. Pattern Recognition, pages 347–349, Montreal, Canada, 1984.

Recursive and Standard X-Y Cutting

Recursive Cuts: (Ha, Haralick et al., ICDAR 1995) Cut at maximum projection gap each iteration; used for page segmentation

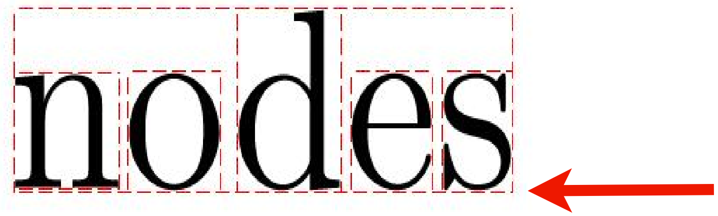


(b) Recursive X-Y Tree (c) Standard X-Y Tree

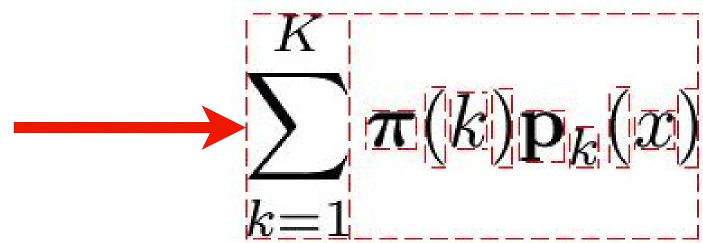
Standard cuts: (Nagy & Seth, ICPR 1984) alternate cutting all gaps in y and x directions; depth limited version used as features for regions stored in the index (1 vertical, 1 horizontal cut)

Edge Distance Feature

For connected components, obtain maximum vertical offset from top/bottom of image:

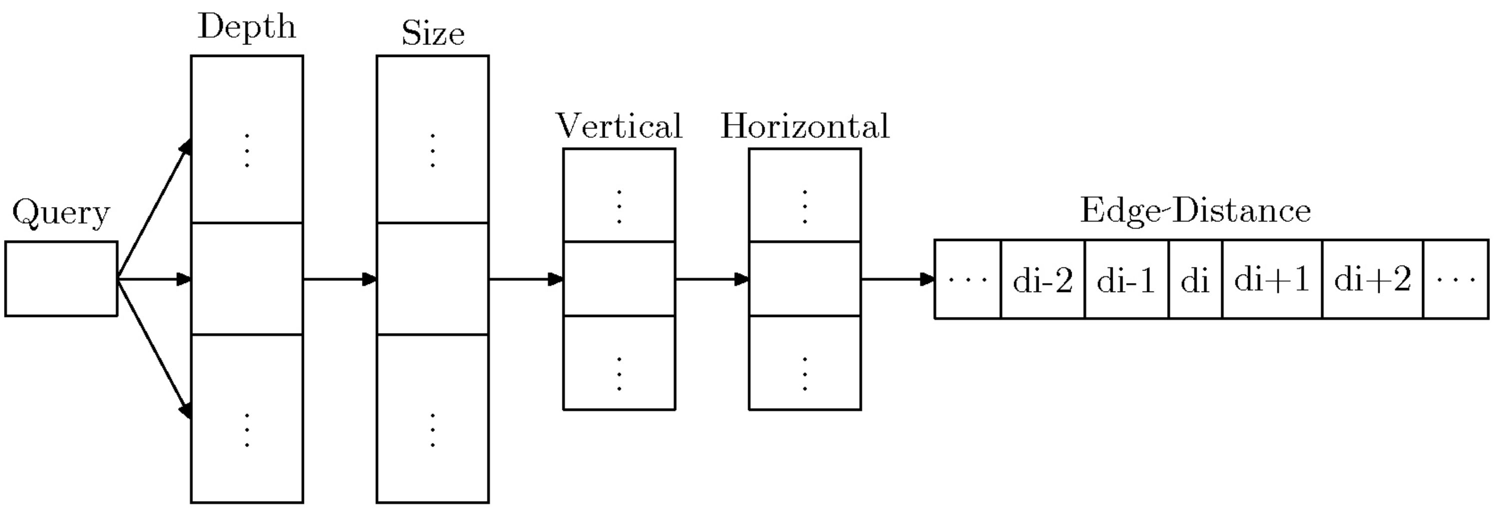


$$\frac{\min(\max(top), \max(bottom))}{Height}$$



Indexing and Retrieval

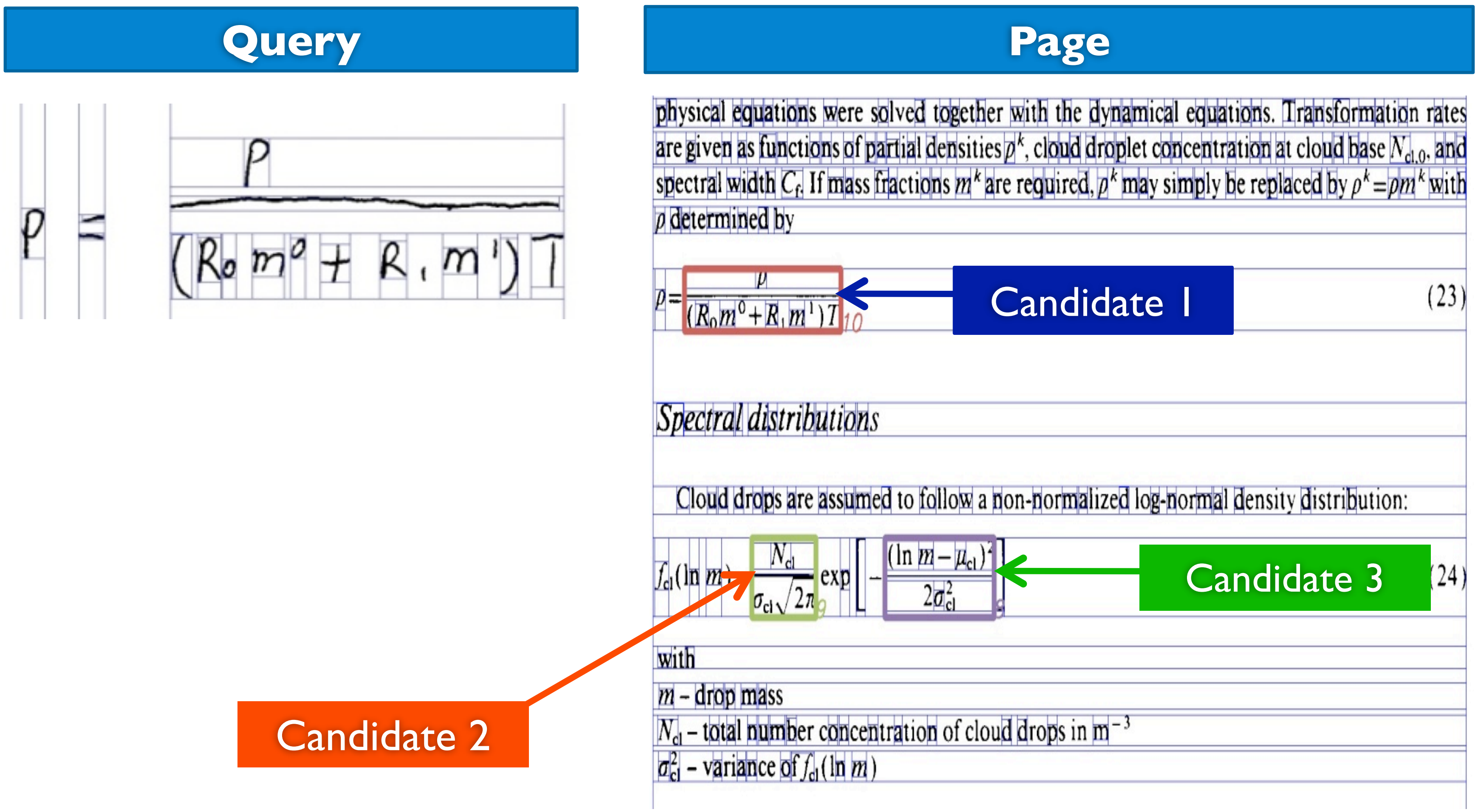
All recursive X-Y sub-trees with depth at least 2 and 90 or fewer nodes are stored in the index; cut directions ignored



Query Match Criteria (for selecting candidate matches):

1) Recursive X-Y tree: $R_d \in Q_d \pm \alpha \wedge R_s \in Q_s \pm \beta$
2) Standard X-Y cuts: $R_x \in Q_x \pm \gamma \wedge R_y \in Q_y \pm \gamma$
3) Contour offset: $R_o \in Q_o \pm \delta$
4) Aspect ratio: $Q_a/2 \leq R_a \leq 2Q_a$

Example: Retrieving Candidates



*Candidates ranked by image similarity with the query

Image Similarity: Dynamic Time Warping

Image (Dis)Similarity: $D(|F_Q|, |F_C|)$

$$D(i,j) = min \left\{ \begin{array}{r} D(i-1,j) \\ D(i,j-1) \\ D(i-1,j-1) \end{array} \right\} + d(i,j) \qquad (1)$$

$$d(i,j) = (u(F_Q[i]) - u(F_C[j]))^2 + (l(F_Q[i]) - l(F_C[j]))^2 \qquad (2)$$

where $D(0,0) = 0$, $D(x,0) = \infty$ for $1 \leq x \leq |F_Q|$, and $D(0,y) = \infty$ for $1 \leq y \leq |F_C|$.

Image similarity computed in terms of least cost squared sum for differences in projection profiles in upper and lower halves of each image (scaled to interval [0, 0.5])

T.M. Rath and R. Manmatha. *Word image matching using dynamic time warping*. In Proc. Computer Vision and Pattern Recognition, pages 521–527, Madison, WI, 2003.

Experimental Design

Task: retrieve a *specific* query expression at its original location in a document within the top-n (n = {1, 5, 10}) results

Corpus: CVPR 2008 collection (1688 pages)

- 400 pages selected randomly, 200 for training and 200 for testing
- Samples obtained through random sampling of expressions labeled according to structure (primarily horizontal, vertical, or roughly equal numbers of x-y cutting gaps; 1 per page containing math) - primarily offset expressions
- PDF files converted to JPG format at 300 DPI (noise-free)

Participants: 10 Graduate Students from Computer Science Department, RIT

- 5 men and 5 women
- Each asked to write 40 queries
    - 20 from training set, 20 from testing set

$$\frac{\kappa}{4\pi \sinh \kappa} \cosh(\kappa \boldsymbol{\mu}^T \mathbf{x})$$

$$f(x,y) \approx \sum_{j=0}^{n} a_j \phi_j(x,y)$$

$$\frac{35x^4 - 30x^2 + 3}{8}$$

$$\sum_{k=1}^{K} \boldsymbol{\pi}(k) \mathbf{p}_k(x)$$

$$\nu = \frac{2\nu_+ \nu_-}{\nu_+ + \nu_-}$$

$$u_i = \sum_{x \in S_i} p(x) \in [0,1]$$

$$\frac{dc_i}{dt} = -\lambda_i^\alpha c_i$$

$$\mathbf{f}_\mathrm{s} = \frac{1}{2\pi} \int_0^{2\pi} g_\phi \mathbf{f} \, d\phi.$$

$$\int_{X \times X} e_\varphi(x, x') dx dx'$$

$$A(\widehat{W}_A - V_A) + N(W_N - V_N)$$

$$-\log r(g|\partial g)$$

$$\frac{1}{2\mu_0} \exp(-\frac{|\delta(u)|}{\mu_0})$$

$$\left(\frac{1}{a^2} + \frac{1}{b^2} \right)$$

$$\Delta[n] = \hat{r}_{1,K}^{(1)}[n] - \hat{r}_{1,K}^{(2)}[n],$$

$$C^{(+)} = \begin{pmatrix} P & Q \\ Q^T & R \end{pmatrix}^{-1}$$

$$\frac{1}{1 + \exp\{Af(x) + B\}}$$

$$p(k) = e^{-E_t} \cdot \frac{E_t^k}{k!}$$

$$\frac{1}{2}(f_i - f_{\hat{j}})$$

$$\frac{1}{M} \cdot \mathbf{B} \cdot \underline{\mathbf{P}}(Y_m = n)$$

$$\boldsymbol{g} = \begin{pmatrix} 1 & 0 \\ 0 & \sin^2 \theta \end{pmatrix}$$

Test Set (20 Expressions)

Sample Queries

$$\frac{k}{4\pi \sinh x} \cosh(k u^T x)$$

$$f(x, y) \approx \sum_{j=0}^{n} a_j \phi_j(x, y)$$

$$\frac{35x^4 - 30x^2 + 3}{8}$$

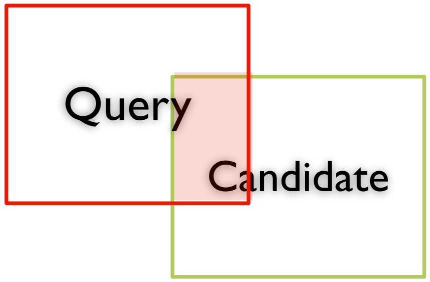
$$\sum_{K=1}^{K} \pi(K) p_K(x)$$

$$A(\widehat{W_A} - V_A) + N(W_N - V_N)$$

Evaluation Metrics

Area Recall

$$A_{recall} = \frac{A_{query} \cap A_{return}}{A_{query}}$$



consider match with highest area recall in top-n results

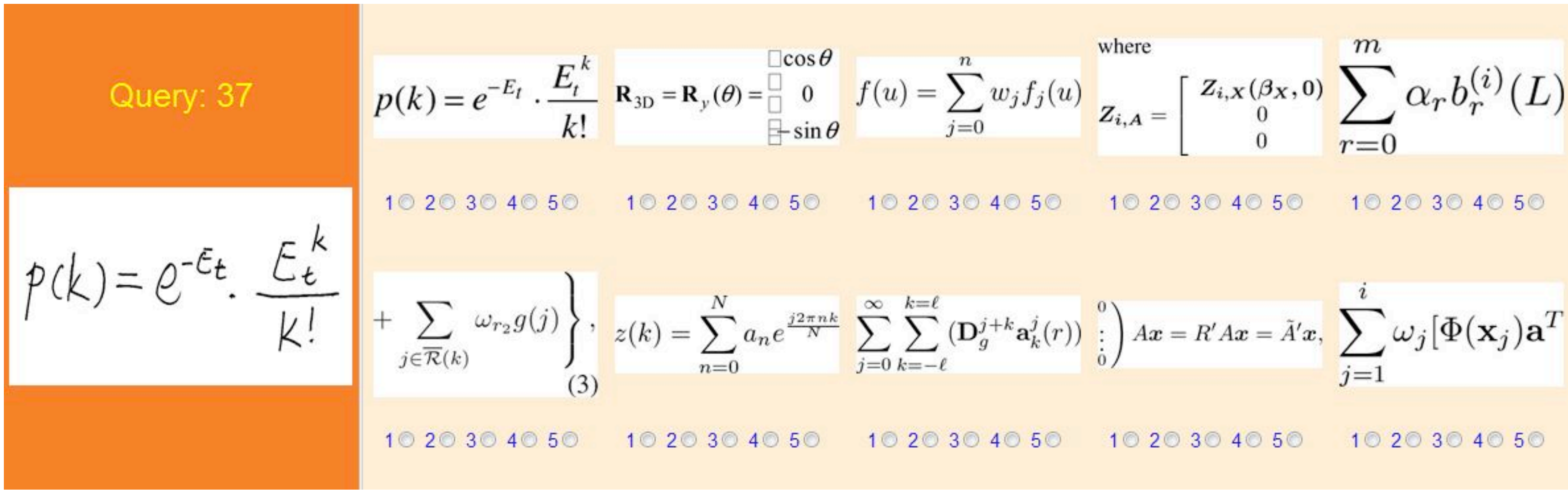
Page Recall

$$P_{recall} = \frac{P_{match}}{Queries}$$

number of queries returned with page for query in the top-n results

Note: metrics are conservative; we consider only matches for the expression at the original location of the selected query expression

Relevance Evaluation



Evaluation Using a 5-point Likert Scale:

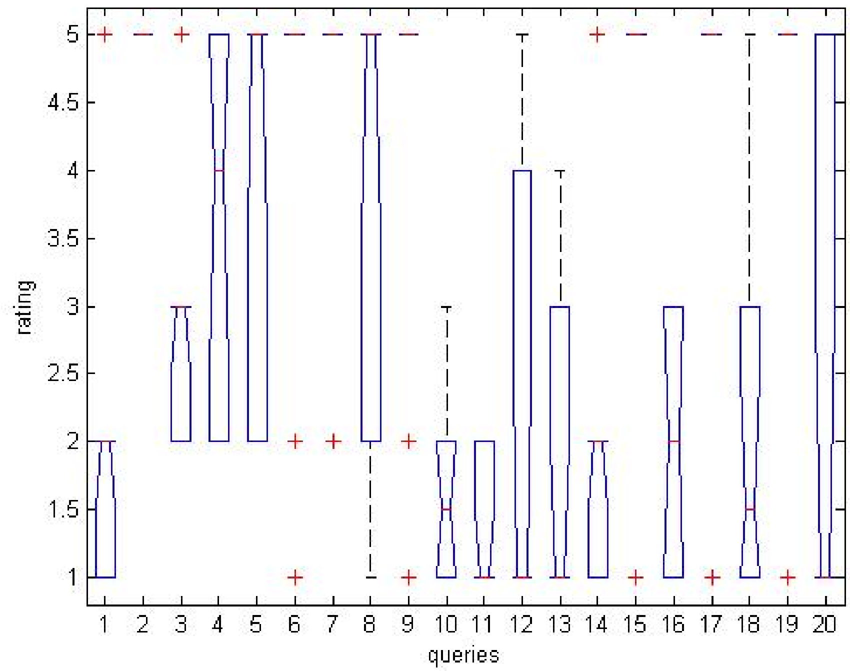
1. No match
2. Less than 1/2 the query is matched
3. Roughly half the query is matched
4. More than half the query is matched
5. The query is completely matched

Results

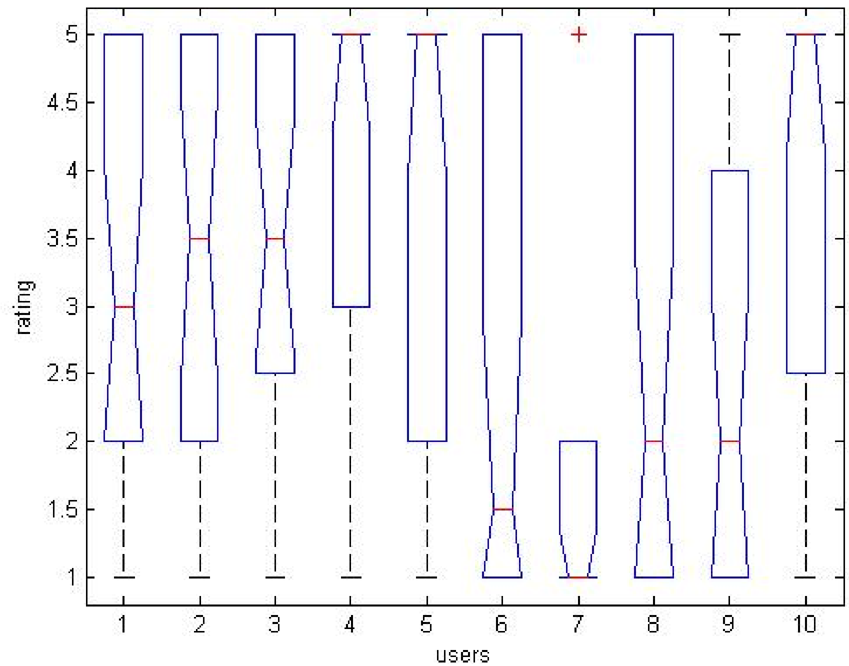
RETRIEVAL ACCURACY FOR TEST SET ($\alpha = 4,\ \beta = 16,\ \gamma = 2,\ \delta = 0.2$). RESULTS ARE SHOWN FOR 20 QUERIES WRITTEN BY TEN PARTICIPANTS, ALONG WITH THE ORIGINAL QUERY IMAGES

Top	P_{recall} (%) μ	P_{recall} (%) σ	A_{recall} (%) μ	A_{recall} (%) σ	$Part.$ (1-5) μ	$Part.$ (1-5) σ
	HANDWRITTEN QUERY IMAGES					
1	38.6	11.7	26.7	13.3	2.06	0.63
5	54.9	14.2	39.8	13.8	2.97	0.77
10	63.2	14.9	43.3	14.0	3.15	0.71
	ORIGINAL QUERY IMAGES					
1			90.0	30.0	4.65	0.08
5			90.0	30.0	4.83	0.05
10			90.0	30.0	4.83	0.05

Participant Evaluations



Ratings by Query



Ratings by Participant

Participant 7: $A(\widehat{W_A} - V_A) + N(W_N - V_N)$

$$\frac{\kappa}{4\pi \sinh \kappa} \cosh(\kappa \boldsymbol{\mu}^T \mathbf{x})$$

$$f(x,y) \approx \sum_{j=0}^{n} a_j \phi_j(x,y)$$

$$\frac{35x^4 - 30x^2 + 3}{8}$$

$$\sum_{k=1}^{K} \boldsymbol{\pi}(k) \mathbf{p}_k(x)$$

$$\nu = \frac{2\nu_+ \nu_-}{\nu_+ + \nu_-}$$

$$u_i = \sum_{x \in S_i} p(x) \in [0,1]$$

$$\frac{dc_i}{dt} = -\lambda_i^\alpha c_i$$

$$\mathbf{f}_s = \frac{1}{2\pi} \int_0^{2\pi} g_\phi \mathbf{f} \, d\phi.$$

$$\int_{X \times X} e_\varphi(x, x') dx dx'$$

$$A(\widehat{W}_A - V_A) + N(W_N - V_N)$$

$$-\log r(g|\partial g)$$

$$\frac{1}{2\mu_0} \exp\left(-\frac{|\delta(u)|}{\mu_0}\right)$$

$$\left(\frac{1}{a^2} + \frac{1}{b^2}\right)$$

$$\Delta[n] = \hat{r}^{(1)}_{1,K}[n] - \hat{r}^{(2)}_{1,K}[n],$$

$$C^{(+)} = \begin{pmatrix} P & Q \\ Q^T & R \end{pmatrix}^{-1}$$

$$\frac{1}{1 + \exp\{Af(x) + B\}}$$

$$p(k) = e^{-E_t} \cdot \frac{E_t^{\,k}}{k!}$$

$$\frac{1}{2}(f_i - f_{\hat{j}})$$

$$\frac{1}{M} \cdot \mathbf{B} \cdot \underline{\mathbf{P}}(Y_m = n)$$

$$\boldsymbol{g} = \begin{pmatrix} 1 & 0 \\ 0 & \sin^2\theta \end{pmatrix}$$

Blue: median raking of 5 (handwritten)

Red: median rank of 1 (no match, handwritten)

Dashes: two original query images not located

Test Set (20 Expressions)

Summary

X-Y cutting and word spotting techniques applied to query-by-expression using handwritten queries; in experimental results, on average one candidate region in the top-10 matches overlaps 43% of target expression (std. deviation = 14%)

Original query images were retrieved more reliably than handwritten queries (90% average area match, 30% std. deviation (top 1!))

Future Directions

Improving indexing, retrieval and feature representations for handwritten and typeset expression image queries

Extension for other languages and graphical objects (tables, charts, etc.)

Modifications to X-Y cutting (e.g. removing largest CC), alt. region segmentation techniques

A. Raja, M. Rayner, A.P. Sexton, and V. Sorge. **Towards a parser for mathematical formula recognition**. In Mathematical Knowledge Management, volume 4108 of LNAI, pages 139–151, 2006.

Other Retrieval Work

LaTeX-based Search (e.g. for the arXiv)

- R. Zanibbi and B. Yuan (DRR 2011): using tf-idf keyword-based search, in isolation and paired with simple image-based matching
- T. Schellenberg, B. Yuan and R. Zanibbi (DRR 2012): using substitution index trees with graph based penalty metric for retrieval

R. Zanibbi and B. Yuan. (2011) **Keyword and image-based retrieval for mathematical expressions.** *Proc. Document Recognition and Retrieval XVIII*, vol. 7874 of Proc. SPIE, pp. OI1-OI9, San Francisco, CA.

T. Schellenberg, B. Yuan, and R. Zanibbi. (2012). **Layout-based substitution tree indexing and retrieval for mathematical expressions**, *Proc. Document Recognition and Retrieval XIX*, San Francisco (to appear).

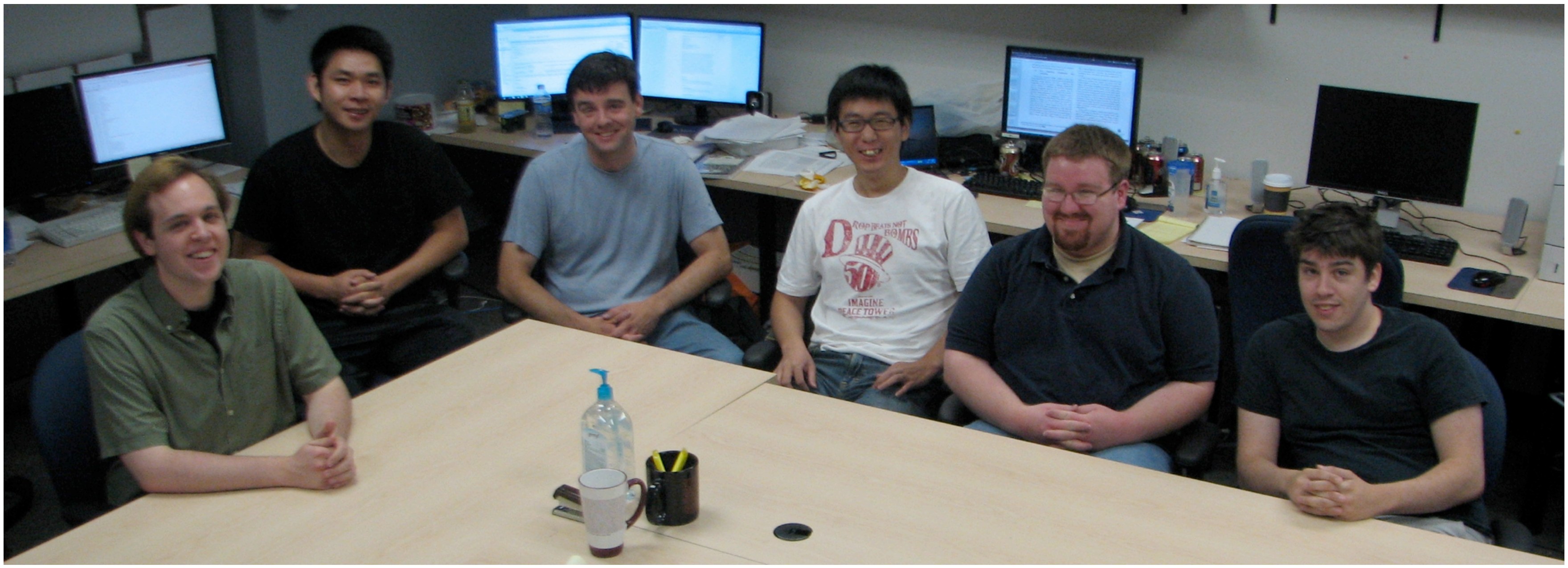
Table 6.5: Top 20 Results for Query 2: $L_1 \times L_2 \times L_3$

Result	Lucene	Sub. Tree	Sub. Tree Rank
1	$L_1 \times L_2 \times L_3$	$L_1 \times L_2 \times L_3$	100.0
2	$E_1 \times E_2 \times E_3$	$E_1 \times E_2 \times E_3$	77.3
3	GL_3	$G = G_1 \times G_2$	60.2
4	$L_1 \gg L_2$	q_1, q_2, q_3	58.6
5	$\boldsymbol{L} = (L_1, L_2, L_3)$	$\sigma_1, \sigma_2, \sigma_3$	58.6
6	L_2	f_1, f_2, f_3	58.6
7	L_2	e_1, e_2, e_3	58.6
8	L_2	t_1, t_2, t_3	58.6
9	L_1	E_1, E_2, E_3	58.6
10	(L_1, L_2, L_2)	j_1, j_2, j_3	58.6
11	$k_2 = (0, n_2/L_2, n_3/L_2)$	i_1, i_2, i_3	58.6
12	$L_1 = \pi^{-1}(U_1) \cong U_1 \times \mathbb{C}_{t_1}$	e_1, e_2, e_3	58.6
13	$L_2 = \pi^{-1}(U_2) \cong U_2 \times \mathbb{C}_{t_2}$	a_1, a_2, a_3	58.6
14	$E_0 \otimes_K L = E$	$a_1 < a_2 = a_3$	58.6
15	Λ_1	e_1, e_2, f_3	58.6
16	$U^{perp} subset U.$	e_2, e_3, f_1	56.9
17	$\Phi = \{\alpha_1, \alpha_2\} \subset \Lambda = \{\alpha_1, \alpha_2, \alpha_3\}$	$L^1(I \times \Omega)$	56.8
18	L_p	$t_1 t_2 t_3 = c$	56.7
19	y_L	$F_2 = L_2 \cdot o$	56.4
20	L_0	$D_1 \supset D_2 \supset D_3 \cdots$	55.8

Table 6.4: Top 20 Results for Query 1: $d,$

Result	Lucene	Sub. Tree	Sub. Tree Rank
1	d	$d,$	100.0
2	d	$X,$	73.8
3	d	$S,$	73.8
4	d	$I,$	73.8
5	d	$\rho,$	73.8
6	d	$s,$	73.8
7	d	$A,$	73.8
8	d	$E,$	73.8
9	d	$V,$	73.8
10	d	$G,$	73.8
11	d	$x,$	73.8
12	d	$V,$	73.8
13	dn	$n,$	73.8
14	d	$K,$	73.8
15	d	$\Omega,$	73.8
16	$d,$	$g,$	73.8
17	$d.$	$W,$	73.8
18	d	$T,$	73.8
19	d	$A,$	73.8
20	d	$s,$	73.8

Document and Pattern Recognition Lab



DPRL Lab Members, July 2011. From left to right: Thomas Schellenberg, Lei Hu, Richard Zanibbi, Bo Ding, Kevin Hart, and Richard Pospesel (not shown: Benjamin Holm and Lane Lawley)

Siyu Zhu (Phd, Imaging Science) joined in September 2011.

Thank you.

Acknowledgements:

National Science Foundation (USA), NSERC Canada

Xerox Foundation (University Affairs Committee Grant)

CEIS/NYSTAR

This material is based upon work supported by the National Science Foundation under Grant No. IIS-1016815. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

UNIVERSITÉ DE NANTES

NSERC CRSNG

xerox

NYSTAR

Queen's UNIVERSITY

dprl [DOCUMENT AND PATTERN RECOGNITION LAB]

R·I·T