Part-Structured Inkball Models for One-Shot Handwritten Word Spotting

Nicholas R. Howe
Word Spotting (by Example)

To the Honourable Robert Dinwiddie, Esquire, Governor.
Word Spotting (by Example)
Word Spotting (by Example)
Word Spotting (by Example)

5. To the Honourable Robert Dinwiddie, Esquire, Governor.

28. To ensign Fleming, of the Virginia Regiment.
You are hereby ordered to repair to Captain Hogg’s company at Fort Dinwiddie.

... etc.
One-Shot Learning

Single example is all you get (usually)
One-Shot Learning

Single example is all you get (usually)

Handwriting varies – must generalize to match
One-Shot Learning

Single example is all you get (usually)

Handwriting varies – must generalize to match

Flexibility is essential – no planar transformations
Part-Structured Models

- Used for photographic object recognition
- Detected parts arranged in approximate spatial configuration
Part-Structured Models

• Used for photographic object recognition
• Detected parts arranged in approximate spatial configuration
• Successful fit identifies required parts near expected position
Inkball Models

- Model = Closely spaced inkballs forming curve
- Part = Ball of ink
- Tree structure
Inkball Models

• Model = Closely spaced inkballs forming curve
• Part = Ball of ink
• Tree structure
• Connections are flexible links
Part-Structured Inkball Models for One-Shot Handwritten Word Spotting

So, now you know.

...but how do we use these models for word spotting?
Configurations

• Configuration = 2D position for each ball
• Rest/default configuration derived from example
• Altering configuration modifies shape

Rest Configuration  Alternate Configurations
Configuration Energy

• Match of model to image has two terms:

  Internal deformation: how far from default?
  \[ E_\xi (Q, C) \]

  Observational deformation: how far from ink skeleton?
  \[ E_\omega (C, \Omega) \]
Configuration Energy

• Match of model to image has two terms:

  Internal deformation: how far from default?
  \[ E_\xi(Q, C) \]

  Observational deformation: how far from ink skeleton?
  \[ E_\omega(C, \Omega) \]

\[ E(Q, C, \Omega) = E_\xi(Q, C) + E_\omega(C, \Omega) \]

\[ E_\xi(Q, C) = \sum_{j=2}^{m} \frac{\| (\hat{v}_j - \hat{v}_j^\uparrow) - \hat{t}_j \|^2}{2\sigma_j^2} \]

\[ E_\omega(C, \Omega) = \sum_{i=1}^{m} \min_{\tilde{s} \in S} \frac{\| s - \tilde{v}_i \|^2}{2\sigma_i^2} \]
One-Shot Word Spotting

1. Infer inkball model from word sample

2. Efficiently identify model configurations with low energy in target document

3. Confirm candidates via reverse match
Efficient Energy Minimization

• Consider simplest case: single-node model
  – Observation deformation is only term in play
  – Compute the energy for all possible configurations
    Distance to closest ink is just a distance transform
Efficient Energy Minimization

• Consider simplest case: single-node model
  – Observation deformation is only term in play
  – Compute the energy for all possible configurations

  *Distance to closest ink is just a distance transform*
Efficient Energy Minimization

• Consider simplest case: single-node model
  – Observation deformation is only term in play
  – Compute the energy for all possible configurations

*Distance to closest ink is just a distance transform*

Target image
Efficient Energy Minimization

• Consider simplest case: single-node model
  – Observation deformation is only term in play
  – Compute the energy for all possible configurations

*Distance to closest ink is just a distance transform*
Efficient Energy Minimization

• Consider simplest case: single-node model
  – Observation deformation is only term in play
  – Compute the energy for all possible configurations

*Distance to closest ink is just a distance transform*
Efficient Energy Minimization

• Consider simplest case: single-node model
  – Observation deformation is only term in play
  – Compute the energy for all possible configurations

*Distance to closest ink is just a distance transform*
Efficient Energy Minimization

- Consider simplest case: single-node model
  - Observation deformation is only term in play
  - Compute the energy for all possible configurations

*Distance to closest ink is just a distance transform*
Efficient Energy Minimization

• Slightly harder case: **barbell model**
  – Still observation terms only (fixed separation)
  – Energy is sum of offset distance transforms:
Efficient Energy Minimization

• Slightly harder case: **barbell model**
  – Still observation terms only (fixed separation)
  – Energy is sum of offset distance transforms:
Efficient Energy Minimization

• Slightly harder case: barbell model
  – Still observation terms only (fixed separation)
  – Energy is sum of offset distance transforms:

\[
\text{Sum}
\]
Efficient Energy Minimization

• Slightly harder case: **barbell model**
  – Still observation terms only (fixed separation)
  – Energy is sum of offset distance transforms:

  Offset equals default separation of nodes in model

  Sum
Efficient Energy Minimization

• Slightly harder case: **barbell model**
  – Still observation terms only (fixed separation)
  – Energy is sum of offset distance transforms:

Offset equals default separation of nodes in model

Energy functional:
Shows energy of model w.r.t. possible root node placements
Efficient Energy Minimization

• More complication: springy barbell
  – Internal deformation term enters picture
  – Use *generalized distance transform* on offset energy
(Squared) Distance Transform

• Minimum of upward paraboloids extending from ink pixels only, rooted at zero

1D Example:

```
  4 1 0 1 4 4 1 0 0 1 1
```

Note: Computational complexity grows linearly with number of pixels
Generalized Distance Transform

• Minimum of upward paraboloids \textit{at every pixel} but rooted at pixel value
  – \textit{Still linear complexity in number of pixels}

- Intuition:
  \textit{The local value can be beaten by a better one nearby}
Efficient Energy Minimization

• General case: node + arbitrary structure
  – Translate energy of child structure(s) by offset
  – Apply generalized distance transform
  – Add to single-node energy
Model Matching Visualization

• Demonstration with simple example:
  Match model a to image
Model Matching Visualization

Single Node
Model Matching Visualization
Model Matching Visualization

GDT
Model Matching Visualization
Model Matching Visualization

Single Node  Sum
Model Matching Visualization
Model Matching Visualization
Model Matching Visualization

Single Node

GDT

© Nicholas R. Howe, Smith College
Model Matching Visualization

Single Node

Sum
Model Matching Visualization

Single Node
Model Matching Visualization
Model Matching Visualization

GDT

Single Node
Model Matching Visualization
Model Matching Visualization

Subtree

Sum
Model Matching Visualization

Translate

Translate

© Nicholas R. Howe, Smith College
Model Matching Visualization
Model Matching Visualization
Model Matching Visualization
Model Matching Visualization
Model Matching Visualization
Model Matching Visualization
Model Matching Visualization
Model Matching Visualization
Parallel GDT

• Optimum model fit requires:
  – One translation per node
  – One GDT per node
• Work scales with number of image pixels
• Fast parallel computation on graphics processing unit (GPU)
Configuration Recovery

• Energy optimization/model matching is just big dynamic programming problem
• Trace back DP winner to recover configuration
• Useful for display/localization

A quick brown fox jumps over the lazy dog.
Jackdaws love my big sphinx of quartz.
Pack my box with five dozen liquor jugs.
Configuration Recovery

• Energy optimization/model matching is just big dynamic programming problem
• Trace back DP winner to recover configuration
• Useful for display/localization

A quick brown fox jumps over the lazy dog. Jackdaws love my big sphinx of quartz. Pack my box with five dozen liquor jugs.
India, officially the Republic of India, is a country in South Asia. It is the seventh-largest country by geographical area, the second-most populous country, and the most populous democracy in the world. Bounded by the Indian Ocean on the south, the Arabian Sea on the west, and the Bay of Bengal on the east, India has a coastline of 7,517 kilometers. It is bordered by Pakistan to the west, China, Nepal, and Bhutan to the north, and Bangladesh and Burma to the east. India is in the vicinity of Sri Lanka and the Maldives in the Indian Ocean, home to the Indus Valley Civilization and a region of historic trade routes and vast empires. The Indian subcontinent was identified with its commercial and cultural wealth for much of its long history. Four major religions, Hinduism, Buddhism, Jainism, and Sikhism originated here, while Zoroastrianism, Judaism, Christianity, and Islam arrived in the first millennium CE and shaped the region's diverse culture.

Note: left/right color scales do not match.
Sample Result: Query = democracy

India, officially the Republic of India, is a country in South Asia. It is the seventh-largest country by geographical area, the second-most populous country, and the most populous democracy in the world. Bounded by the Indian Ocean on the south, the Arabian Sea on the west, and the Bay of Bengal on the east, India has a coastline of 7,516 kilometers. It is bordered by Pakistan to the west; China, Nepal, and Bhutan to the north; and Bangladesh and Burma to the east. India is in the vicinity of Sri Lanka, and the Nicobar Islands in the Indian Ocean, home to the Indus Valley Civilization and a region of historic trade routes and vast empires, the Indian subcontinent was identified with its commercial and cultural wealth for much of its long history. Four major religions, Hinduism, Buddhism, Jainism, and Sikhism originated here, while Zoroastrianism, Judaism, Christianity, and Islam arrived in the first millennium CE and shaped the region's diverse culture.

Note: left/right color scales do not match.
Sample Result: Query = democracy

India, officially the Republic of India, is a country in South Asia. It is the seventh-largest country by geographical area, the second-most populous country, and the most populous democracy in the world. Bounded by the Indian Ocean on the south, the Arabian Sea on the west, and the Bay of Bengal on the east, India has a coastline of 7,516 kilometers. It is bordered by Pakistan to the west; China, Nepal, and Bhutan to the north, and Bangladesh and Burma to the east. India is in the vicinity of Sri Lanka, and the Maldives in the Indian Ocean, home to the Indian Ocean

Civilisation and a region of historic trade routes and vast empires, the Indian subcontinent was identified with its commercial and cultural wealth for much of its long history. Four major religions, Hinduism, Buddhism, Jainism and Sikhism originated here, while Zoroastrianism, Judaism, Christianity and Islam arrived in the first millennium CE and shaped India, officially the Republic of India, is a country in South Asia. It is the seventh-largest country by geographical area, the second-most populous country, and the most populous democracy in the world. Bounded by the Indian Ocean on the south, the Arabian Sea on the west, and the Bay of Bengal on the east, India has a coastline of 7,516 kilometers. It is bordered by Pakistan to the west; China, Nepal, and Bhutan to the north, and Bangladesh and Burma to the east. India is in the vicinity of Sri Lanka, and the Maldives in the Indian Ocean, home to the Indian Ocean

Note: left/right color scales do not match.
Match Confirmation

• Model matches ink, ignores noise/context
  – Will match **and** to **Alexandria**:
  – Will match **bird** to **bind**:
• Whitespace not considered in model
• Expedient heuristic: Confirm top hits by reverse match
  – Build model of target area & match to query
  – Match energy is greater of the two directions (scaled by number of nodes)
Experimental Data Sets

George Washington (GW20)
- 20 pages; 4685 words
- English cursive script

Parzival
- 47 pages; 18,918 words
- German medieval lettering
Methodology

• Used train/test split from Frinken et al. [PAMI'12]
• Each non-stopword in training set is a query
  – Some appear multiple times in training set
  – Run retrieval on all instances & take high scores
• Reverse match uses segmented words
• Recall-Precision curves averaged for all queries
  – Threshold may vary from query to query
  – Cross-query calibration still requires research
George Washington

![Graph showing precision and recall](image-url)
George Washington

Notes:
• 93.4% Average Precision [84% is prior best]
• 78.9% Precision at 100% Recall
Parzival

![Precision vs Recall Graph](image-url)
Parzival

Notes:
- 88.2% Average Precision [94% is prior best]
- 68.4% Precision at 100% Recall
Parzival

Notes:
• 88.2% Average Precision [94% is prior best]
• 68.4% Precision at 100% Recall

Not bad for such a simple model!
• No learning...
• No language model...
...different yet still good.
Caveat Lector

• Some dependence on handwriting style
  – Intrinsic letter forms can vary
  – Cross-style spotting requires more research

• Limited invariance to scale & rotation
  – Match model scale to text in document
  – Correct skew/rotation prior to spotting

• Speed not yet real-time for large collections
  – Roughly 2 Mpixel/second for most words
Part-Structured Promise

• Powerful matching/retrieval tool
  – Part models could be more complex

• Requires no training, language modeling, etc.
  – Easily applied to new languages, figures, etc.

• Intuitive pixel-level correspondences
  – Starting point for further processing?

• Reference code on my web page
  – I welcome opportunities to collaborate!

http://cs.smith.edu/~nhowe/research/code/
Thank You
Rare Words

• Performs well with single training examples

GW20: 25.4% of queries are singletons ➞ 60.2% precision at full recall
Parzival: 31.8% of queries are singleton ➞ 69.5% precision at full recall
Building PSM from Image

1. Find skeleton
Building PSM from Image

1. Find skeleton
2. Select endpoints & junctions
Building PSM from Image

1. Find skeleton
2. Select endpoints & junctions
3. Add points chosen 2r from included points
Building PSM from Image

1. Find skeleton
2. Select endpoints & junctions
3. Add points chosen $2r$ from included points
4. Additional points to fill remaining gaps
Building PSM from Image

1. Find skeleton
2. Select endpoints & junctions
3. Add points chosen $2r$ from included points
4. Additional points to fill remaining gaps
5. Form tree by greedily connecting closest pairs
Online vs. Offline Models

- Online query allows model structure to follow actual stroke
- Offline query must use *ad hoc* model structure
Some Matches

Fredericksburgh

Fredericksburgh

Fredericksburgh

Fredericksburgh
Caveat Lector

- Some dependence on handwriting style
  - Intrinsic letter forms can vary
  - Cross-style spotting requires more research
- Limited invariance to scale & rotation
  - Match model scale to text in document
  - Correct skew/rotation prior to spotting
- Speed not yet real-time for large collections
  - Roughly 2-3 Mpixel/second for most words