technische universität
dortmund

Bag of Features for Text Detection in the Natural Scene Images

Rashik Thalappully

Master Student,
Informatik XII, Technische Universität Dortmund
17. Mai 2015

Overview

- Text Detection
- Bag Of Features
- Text Detection Using Descriptor formulated by Coates et. al.
- Text Detection Using SIFT Descriptor
- Evaluation
- Results
- Scale Detection Using Maximally Stable Extremal Regions

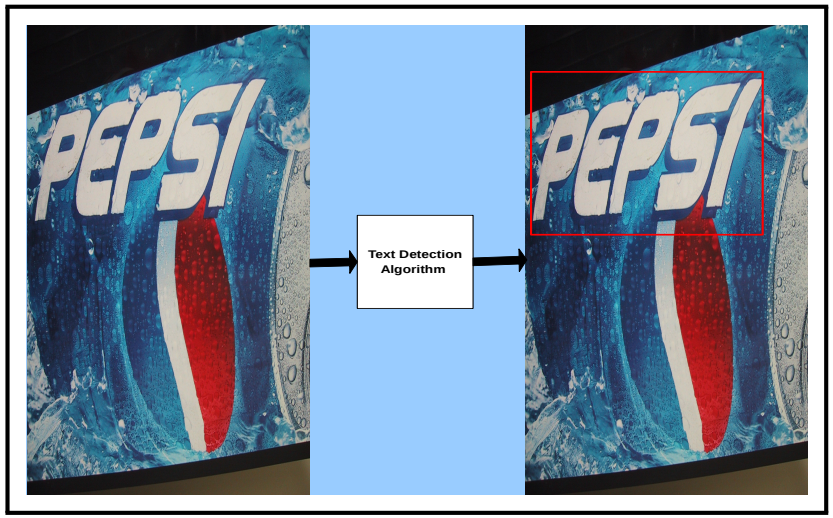


Image Credits: *ICDAR 2003*

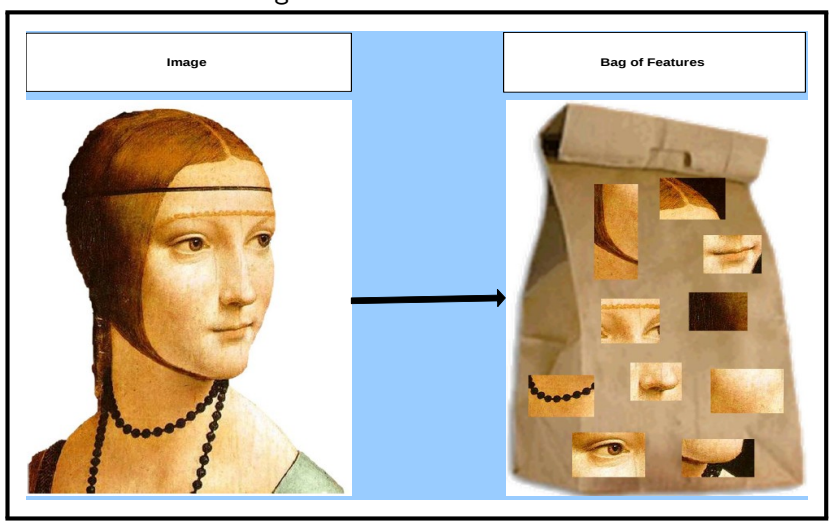


Image Credits: *sensblogs.wordpress.com*

Text Detection in Images

- Recognizing the text regions in an arbitrary image.

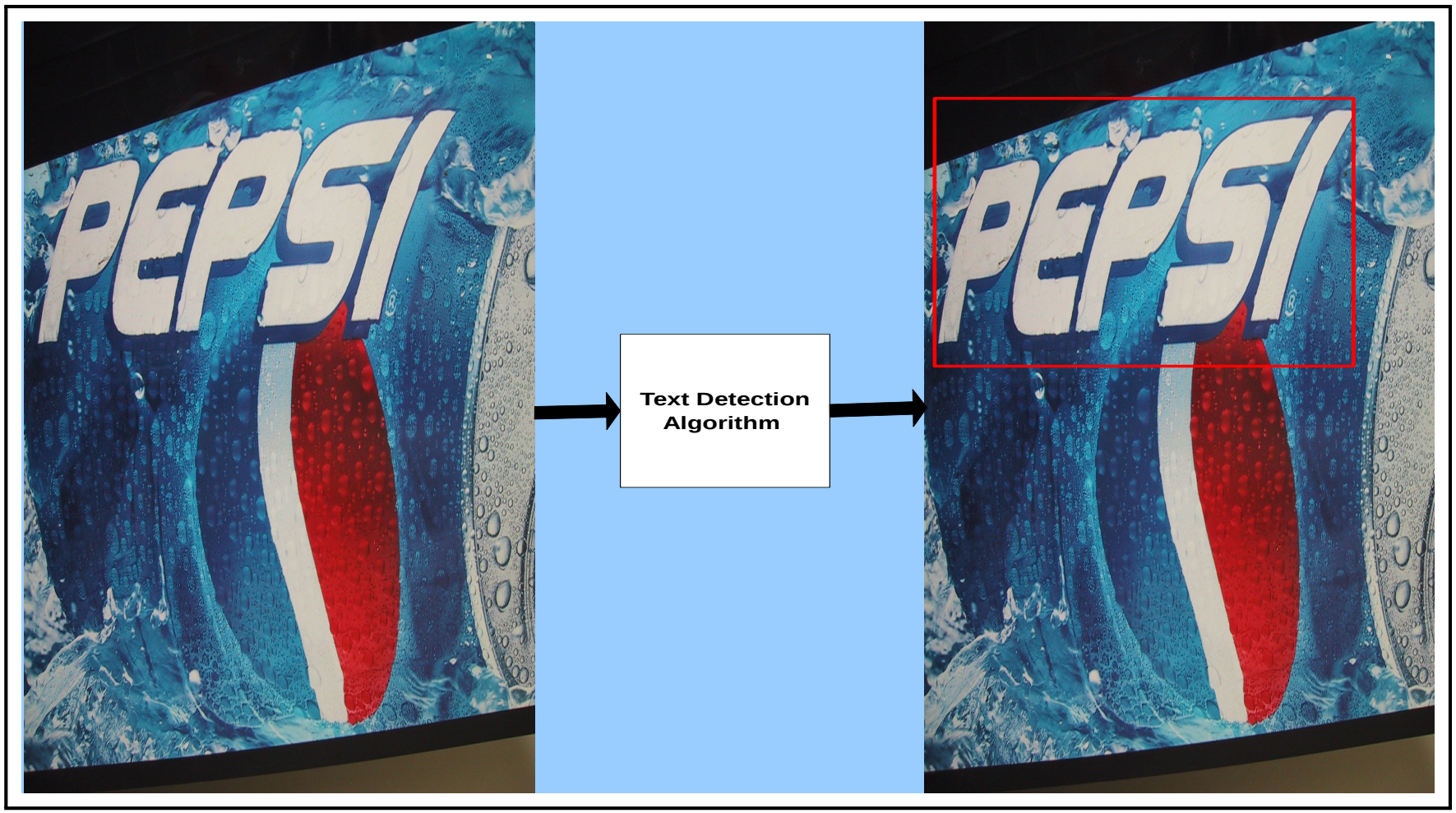


Image Credits: *ICDAR 2003*

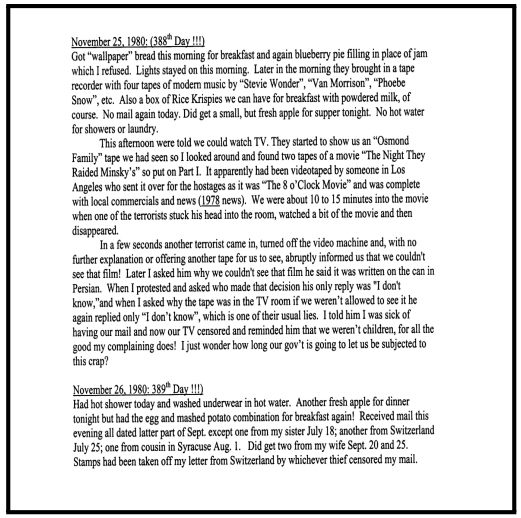
Text Detection in Images

- Problem is quite similar to Optical Character Recognition (OCR)
- Numerous solutions available for OCR for images of digital documents
- Use same solutions for natural scene images?

Image Credits: *mybigguide.com*

Natural Scene Image OCR != Traditional OCR

Image Credits: *http://www.northernsound.ie/*



November 25, 1980: (388th Day !!!)
Got "wallpaper" bread this morning for breakfast and again blueberry pie filling in place of jam which I refused. Lights stayed on this morning. Later in the morning they brought in a tape recorder with four tapes of modern music by "Stevie Wonder", "Van Morrison", "Phoebe Snow", etc. Also a box of Rice Krispies we can have for breakfast with powdered milk, of course. No mail again today. Did get a small, but fresh apple for supper tonight. No hot water for showers or laundry.

This afternoon were told we could watch TV. They started to show us an "Osmond Family" tape we had seen so I looked around and found two tapes of a movie "The Night They Raided Minsky's" so put on Part I. It apparently had been videotaped by someone in Los Angeles who sent it over for the hostages as it was "The 8 o'Clock Movie" and was complete with local commercials and news (1978 news). We were about 10 to 15 minutes into the movie when one of the terrorists stuck his head into the room, watched a bit of the movie and then disappeared.

In a few seconds another terrorist came in, turned off the video machine and, with no further explanation or offering another tape for us to see, abruptly informed us that we couldn't see that film! Later I asked him why we couldn't see that film he said it was written on the can in Persian. When I protested and asked who made that decision his only reply was "I don't know,"and when I asked why the tape was in the TV room if we weren't allowed to see it he again replied only "I don't know", which is one of their usual lies. I told him I was sick of having our mail and now our TV censored and reminded him that we weren't children, for all the good my complaining does! I just wonder how long our gov't is going to let us be subjected to this crap?

November 26, 1980: 389th Day !!!)
Had hot shower today and washed underwear in hot water. Another fresh apple for dinner tonight but had the egg and mashed potato combination for breakfast again! Received mail this evening all dated latter part of Sept. except one from my sister July 18; another from Switzerland July 25; one from cousin in Syracuse Aug. 1. Did get two from my wife Sept. 20 and 25. Stamps had been taken off my letter from Switzerland by whichever thief censored my mail.

Image Credits: *http://www.docstoc.com/*

Natural Scene images

- Artistic fonts
- Extreme lighting variation
- Large variation in color and texture
- Wide range of viewing angles

Images of digital documents

- Typical fonts
- Structured text
- Captured under controlled settings

Natural Scene Image OCR != Traditional OCR

- Additional challenges
- Solutions of traditional OCR can't be applied to natural scene image OCR

Input Image	Traditional OCR	Input Image	Natural Image OCR
	0 i-n C D 0 iz z 0 0 *CD* H) Oc 0 m *(I)* *CD* U 0 *(1')* CDCD >< ITW I • I		University of Essex Day Nursery The Houses Keynes R gh Tawney and William Morris Towers Wolfson Court Buses and Cycles Only CONFERENCE CAR PARK EntTance 4

Image Credits: *web.iit.edu*

Applications - Text Detection in Natural Scene Images

Self Driving Automobiles

Content based Image retrieval

Robotics

Image Credits: *Google Images*

Bag of Features [1]

- Represent images as an order less collection of local features
- Inspired from the Bag of Words representation
- Lacks any spatial information

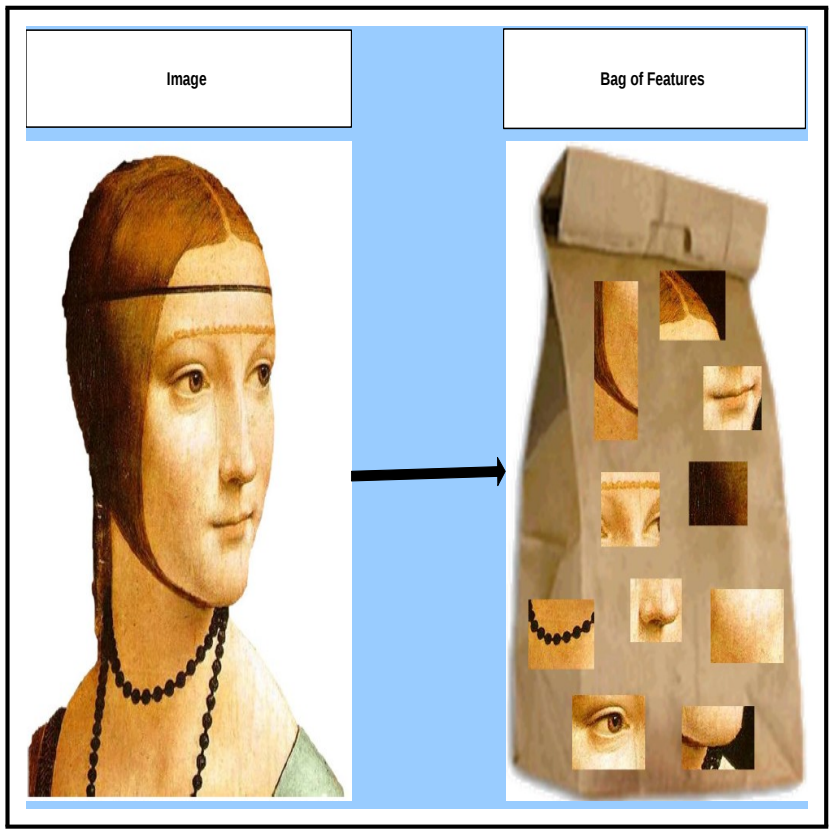


Image Credits: *sensblogs.wordpress.com*

[1] S. O'Hara und B. A. Draper: *Introduction to the bag of features paradigm for image classification and retrieval*. arXiv preprint arXiv:1101.3354, 2011

Process for Bag of Features [2]

- Building Visual Vocabulary

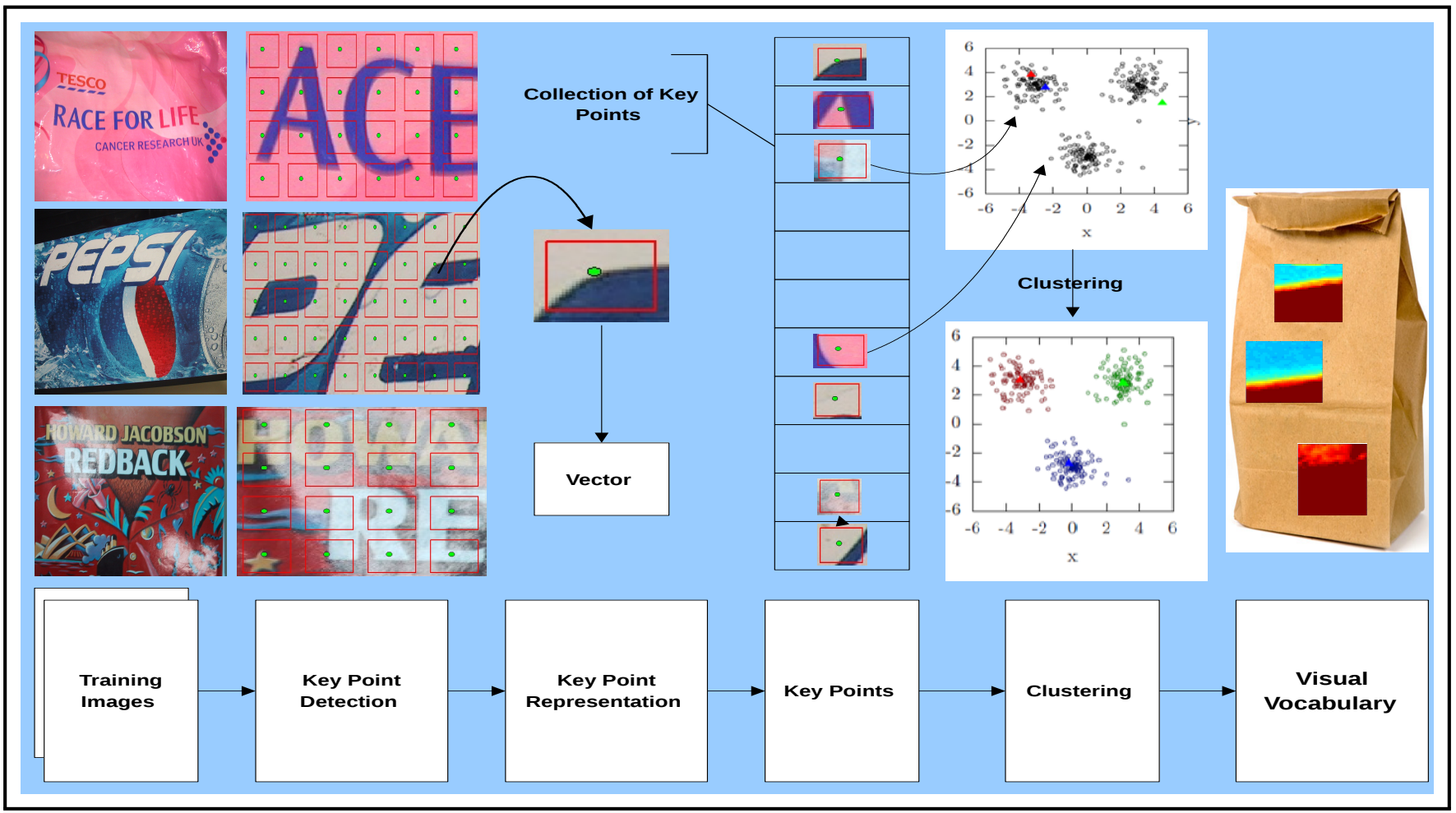


Image Credits: *ICDAR 2003*

[2] S. O'Hara und B. A. Draper: *Introduction to the bag of features paradigm for image classification and retrieval*. arXiv preprint arXiv:1101.3354, 2011

Process for Bag of Features [3]

- Generating Term Vector or Feature Vector

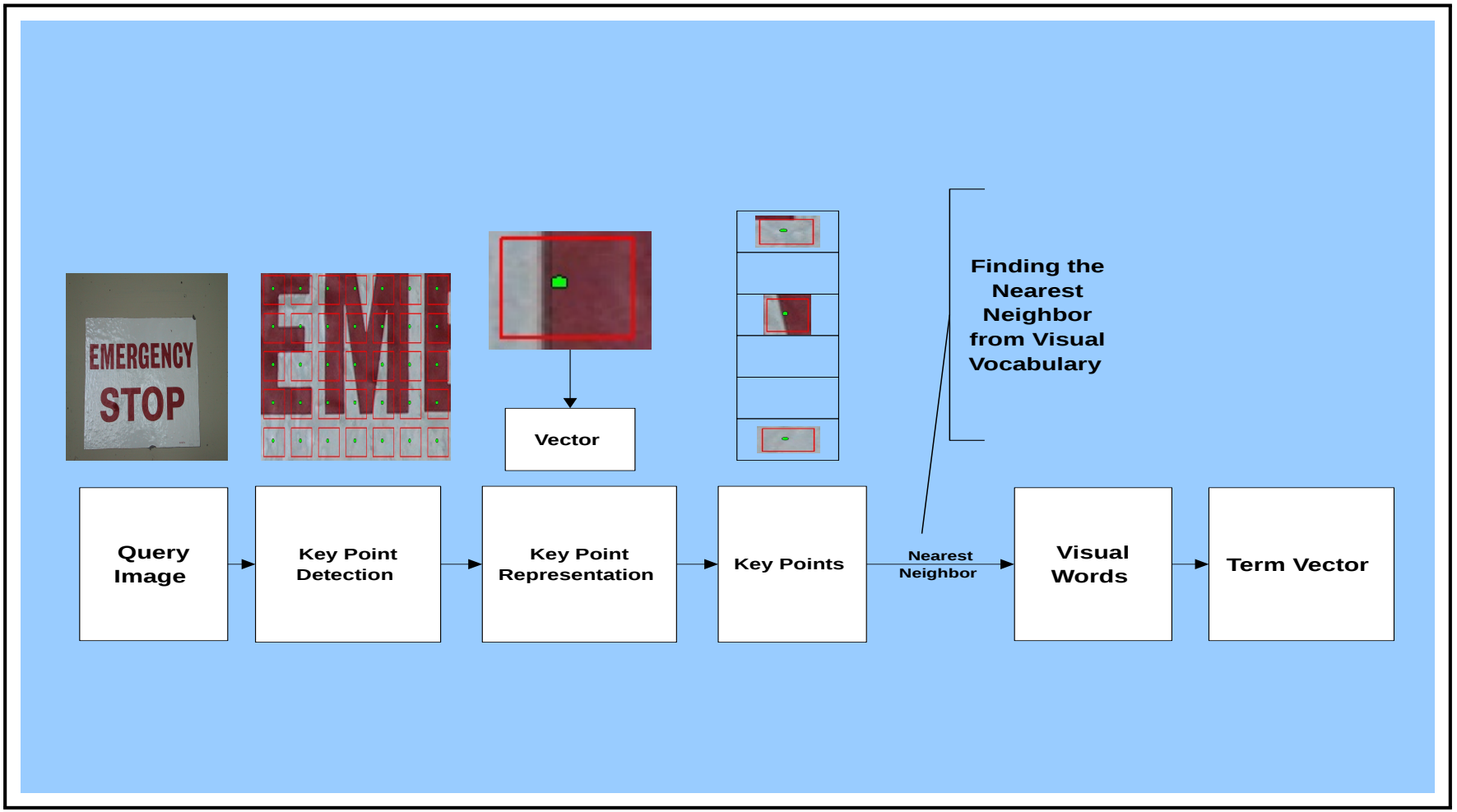


Image Credits: *ICDAR 2003*

[3] S. O'Hara und B. A. Draper: *Introduction to the bag of features paradigm for image classification and retrieval*. arXiv preprint arXiv:1101.3354, 2011

BoF approach proposed by Coates et. al. [4]

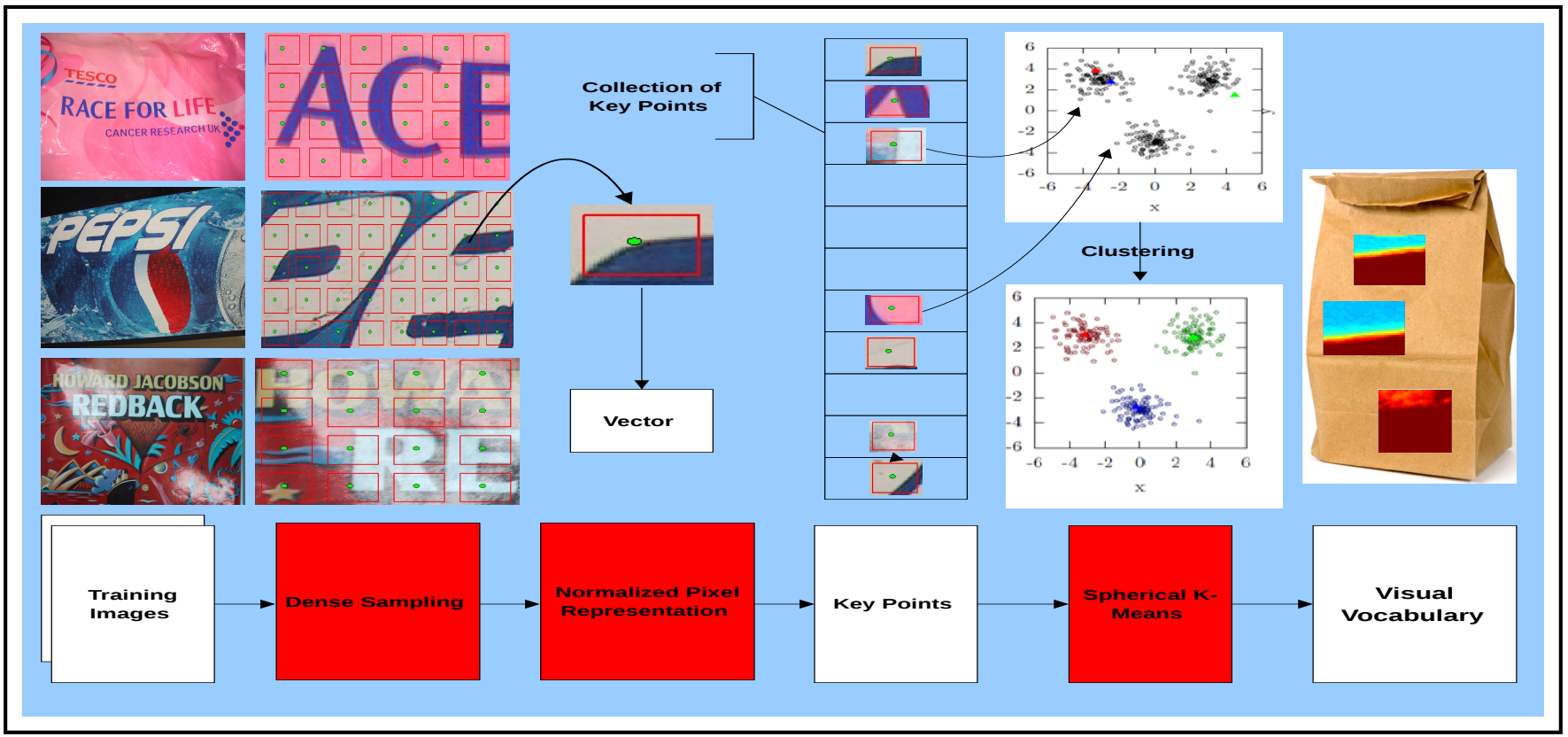


Image Credits: *ICDAR 2003*

- Key point detection - Dense Sampling
- Key point representation - Normalized Pixel Representation
- Clustering algorithm - Spherical K-means

[4] A. Coates, B. Carpenter, C. Case, S. Satheesh, B. Suresh, T. Wang, D. J. Wu und A. Y. Ng: *Text detection and character recognition in scene images with unsupervised feature learning*. In: *Document Analysis and Recognition (ICDAR), 2011 International Conference on*, S. 440–445. IEEE, 2011

BoF approach proposed by Coates et. al. [5]

- Generation of Feature vector

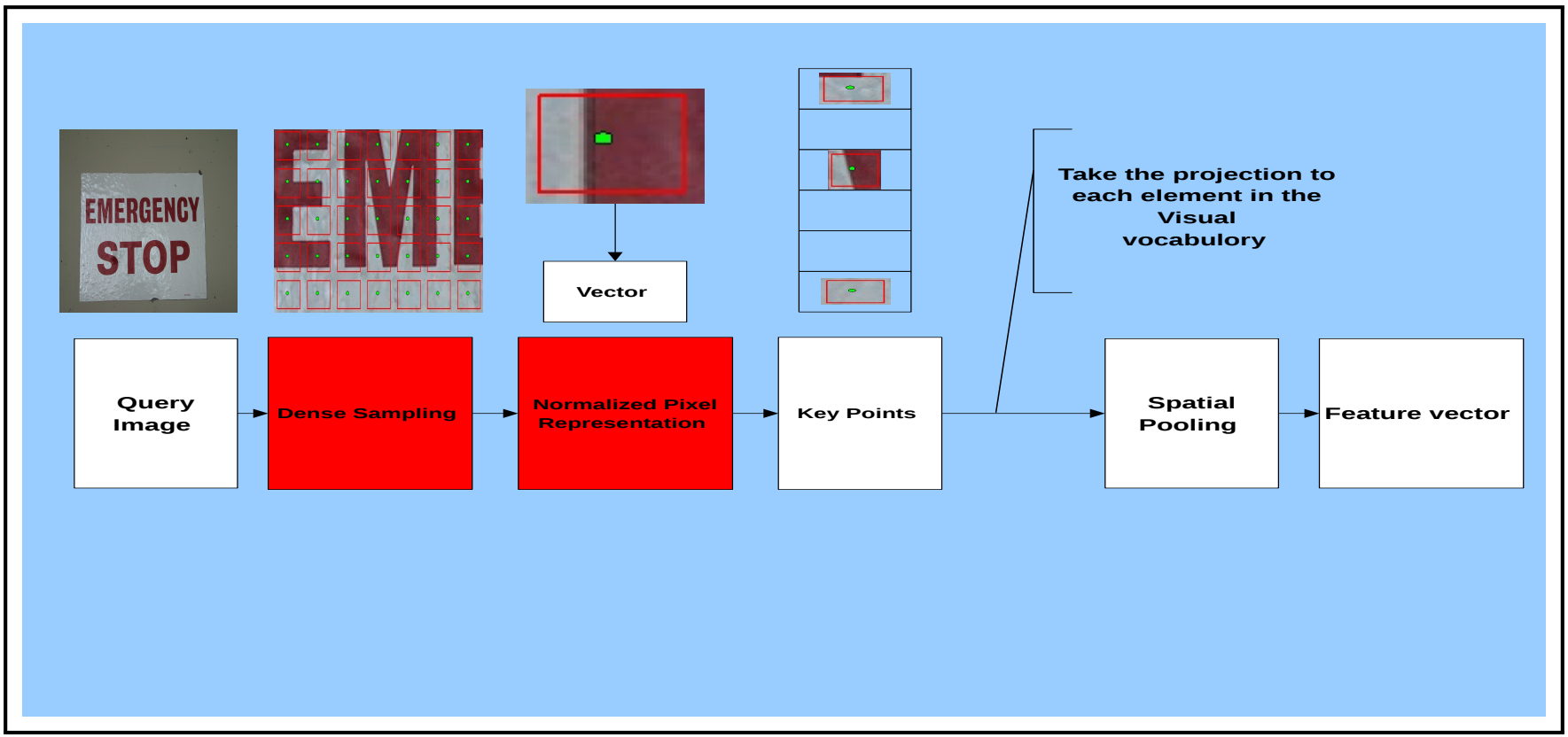


Image Credits: *ICDAR 2003*

[5] A. Coates, B. Carpenter, C. Case, S. Satheesh, B. Suresh, T. Wang, D. J. Wu und A. Y. Ng: *Text detection and character recognition in scene images with unsupervised feature learning*. In: *Document Analysis and Recognition (ICDAR), 2011 International Conference on*, S. 440–445. IEEE, 2011

Normalization And ZCA Whitening[6]

- Comparison of raw and preprocessed patches

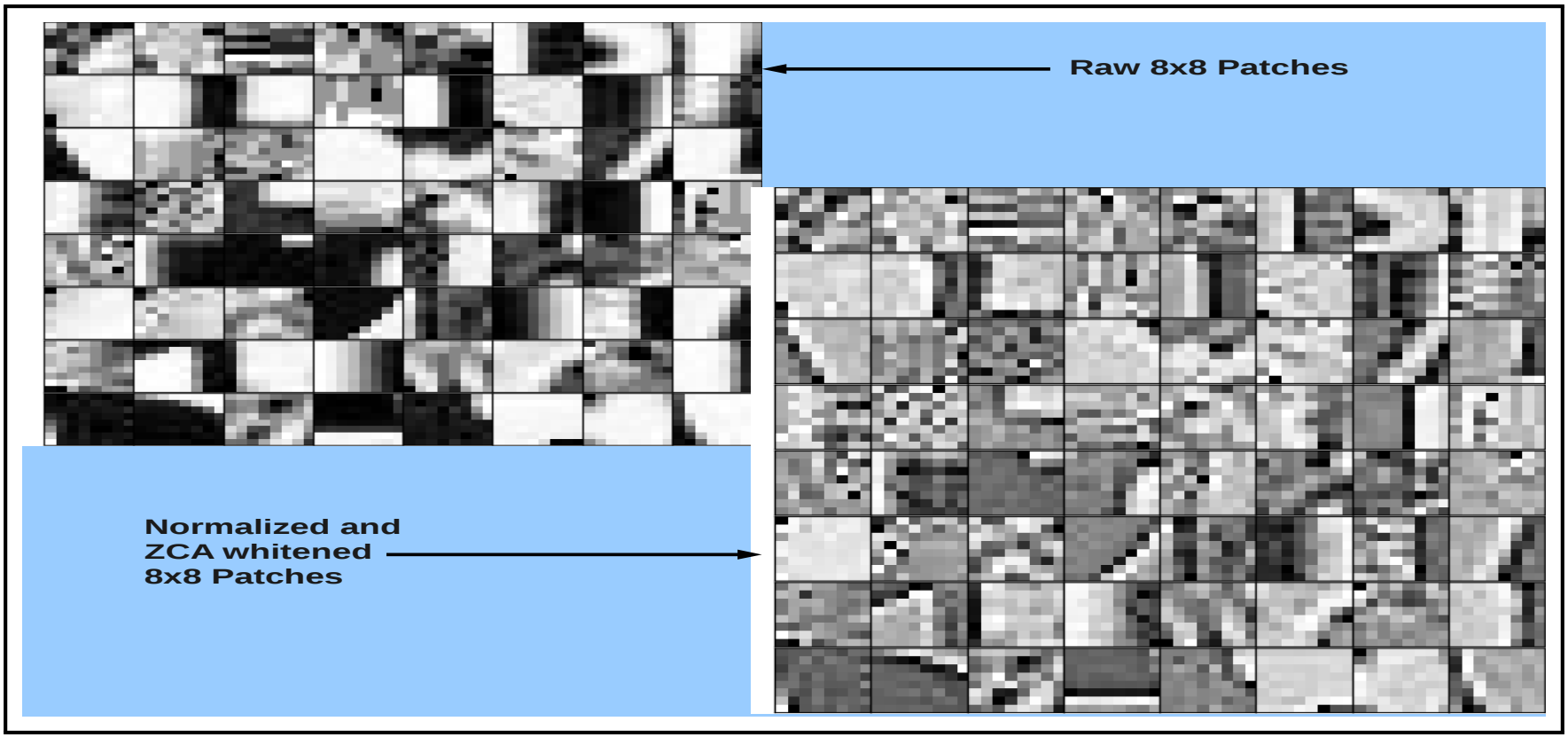
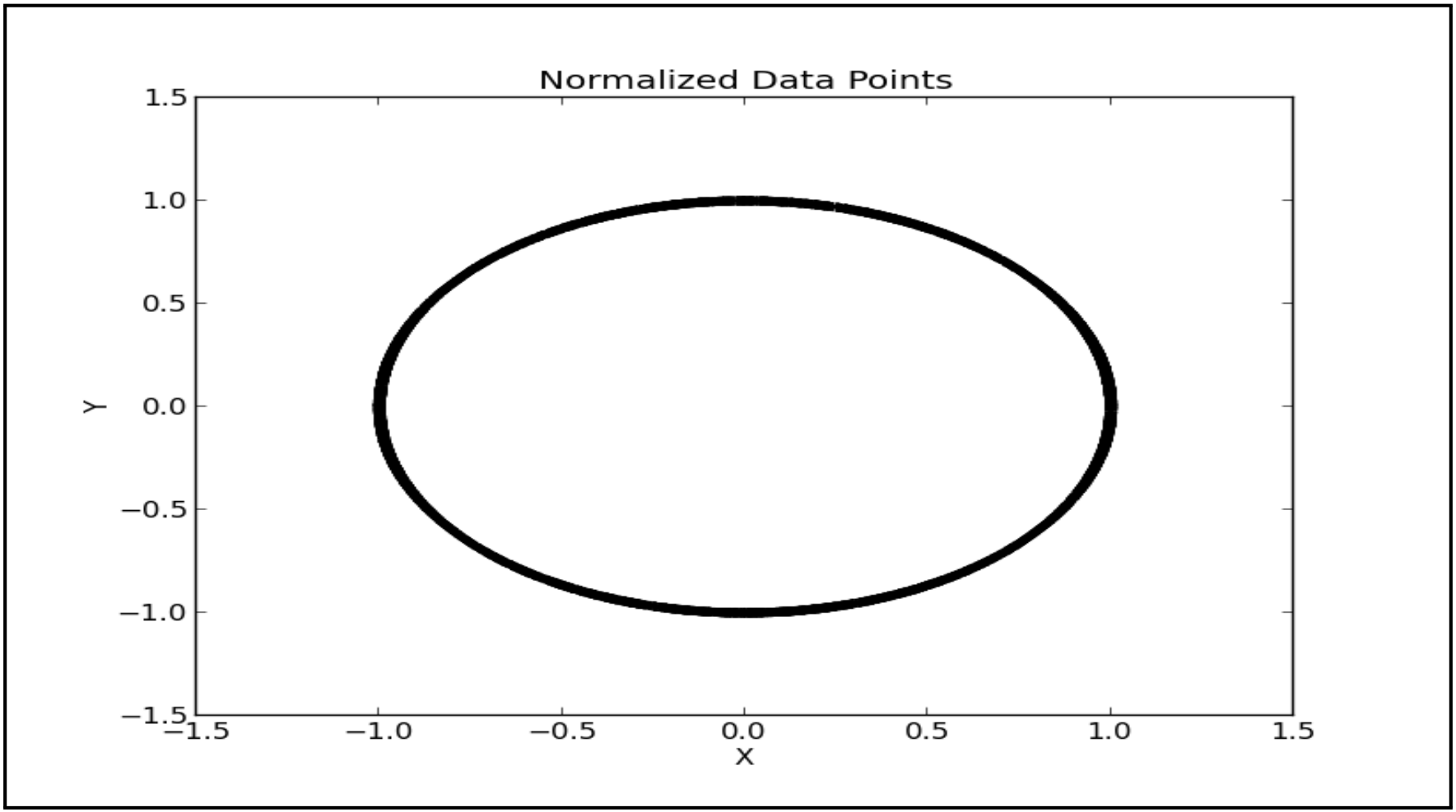


Image Credits: *ICDAR 2003*

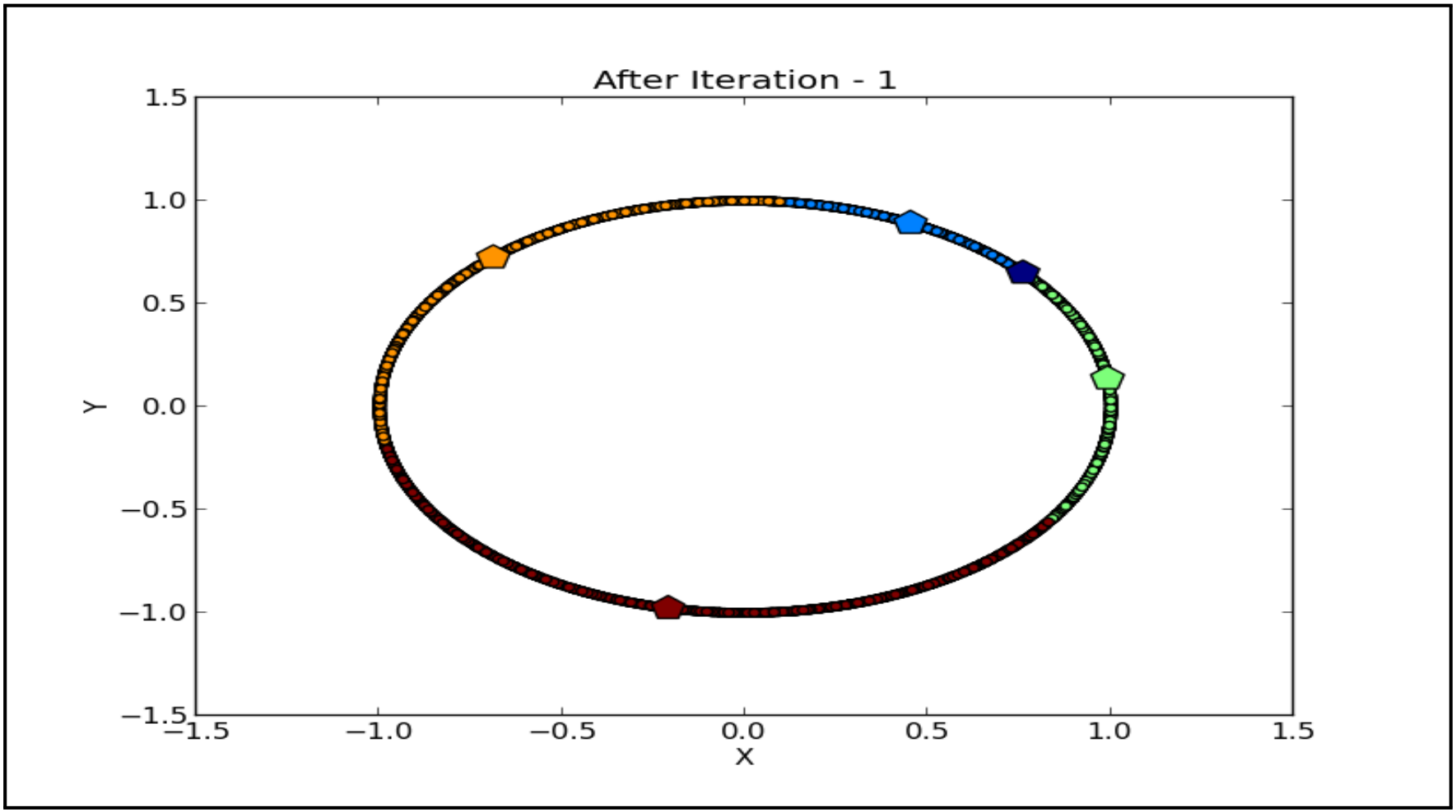
[6] A. Coates und A. Y. Ng: *Learning feature representations with k-means*. In: *Neural Networks: Tricks of the Trade*, S. 561–580. Springer, 2012

Clustering Algorithms - Spherical K Means Algorithm [7]



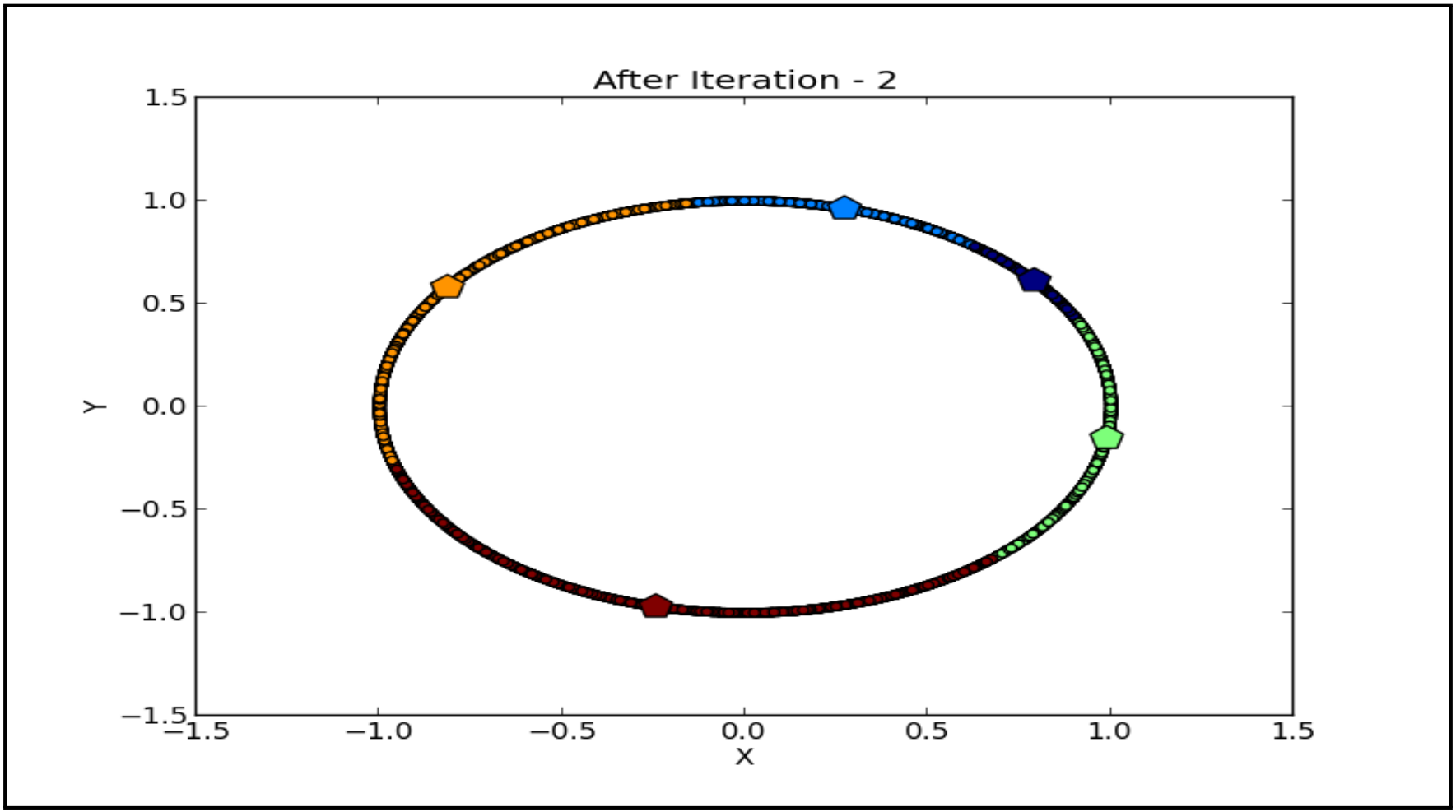
[7] A. Coates und A. Y. Ng: *Learning feature representations with k-means*. In: *Neural Networks: Tricks of the Trade*, S. 561–580. Springer, 2012

Clustering Algorithms - Spherical K Means Algorithm [8]



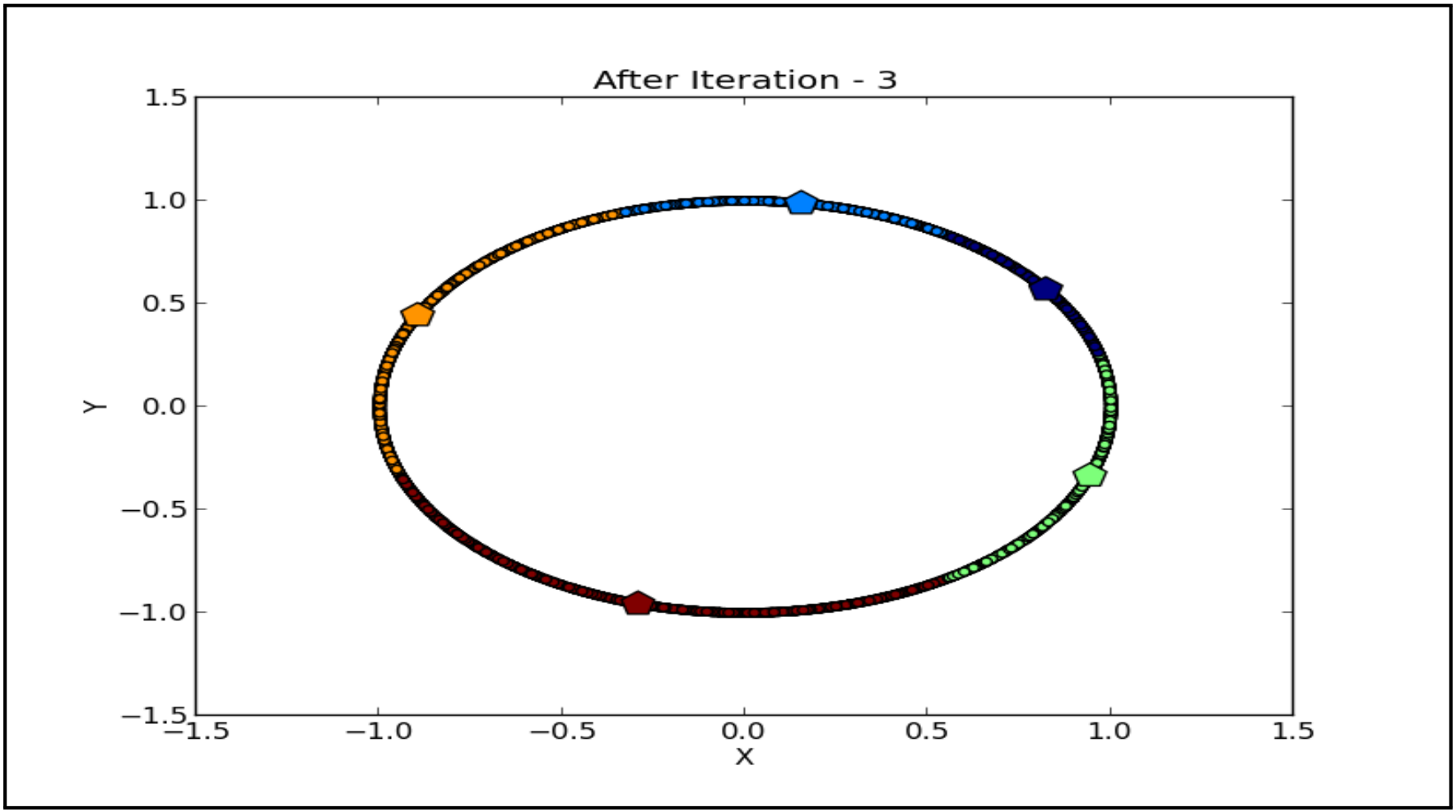
[8] A. Coates und A. Y. Ng: *Learning feature representations with k-means*. In: *Neural Networks: Tricks of the Trade*, S. 561–580. Springer, 2012

Clustering Algorithms - Spherical K Means Algorithm [9]



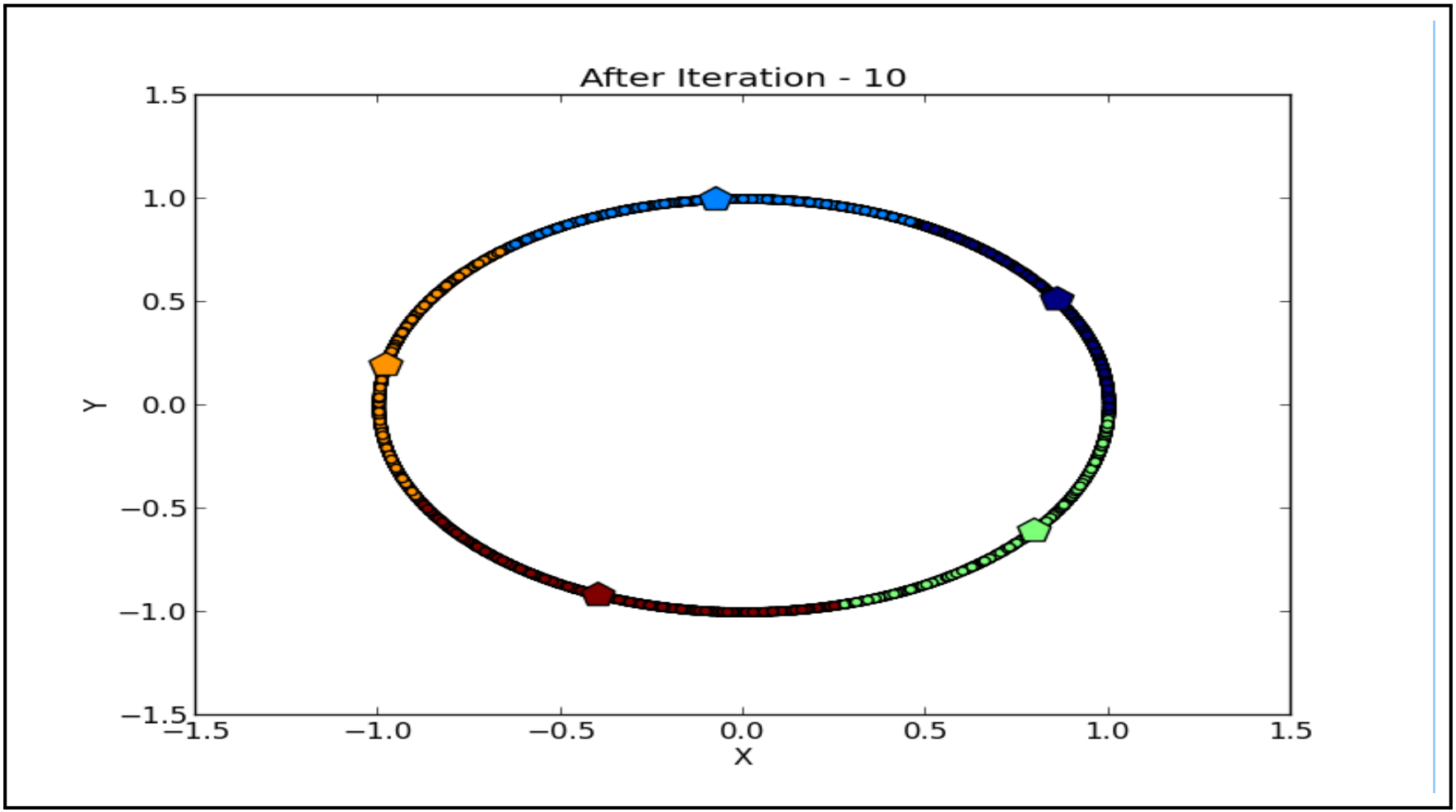
[9] A. Coates und A. Y. Ng: *Learning feature representations with k-means*. In: *Neural Networks: Tricks of the Trade*, S. 561–580. Springer, 2012

Clustering Algorithms - Spherical K Means Algorithm [10]



[10] A. Coates und A. Y. Ng: *Learning feature representations with k-means*. In: *Neural Networks: Tricks of the Trade*, S. 561–580. Springer, 2012

Clustering Algorithms - Spherical K Means Algorithm [11]



[11] A. Coates und A. Y. Ng: *Learning feature representations with k-means*. In: *Neural Networks: Tricks of the Trade*, S. 561–580. Springer, 2012

Creating Visual Vocabulary [12]

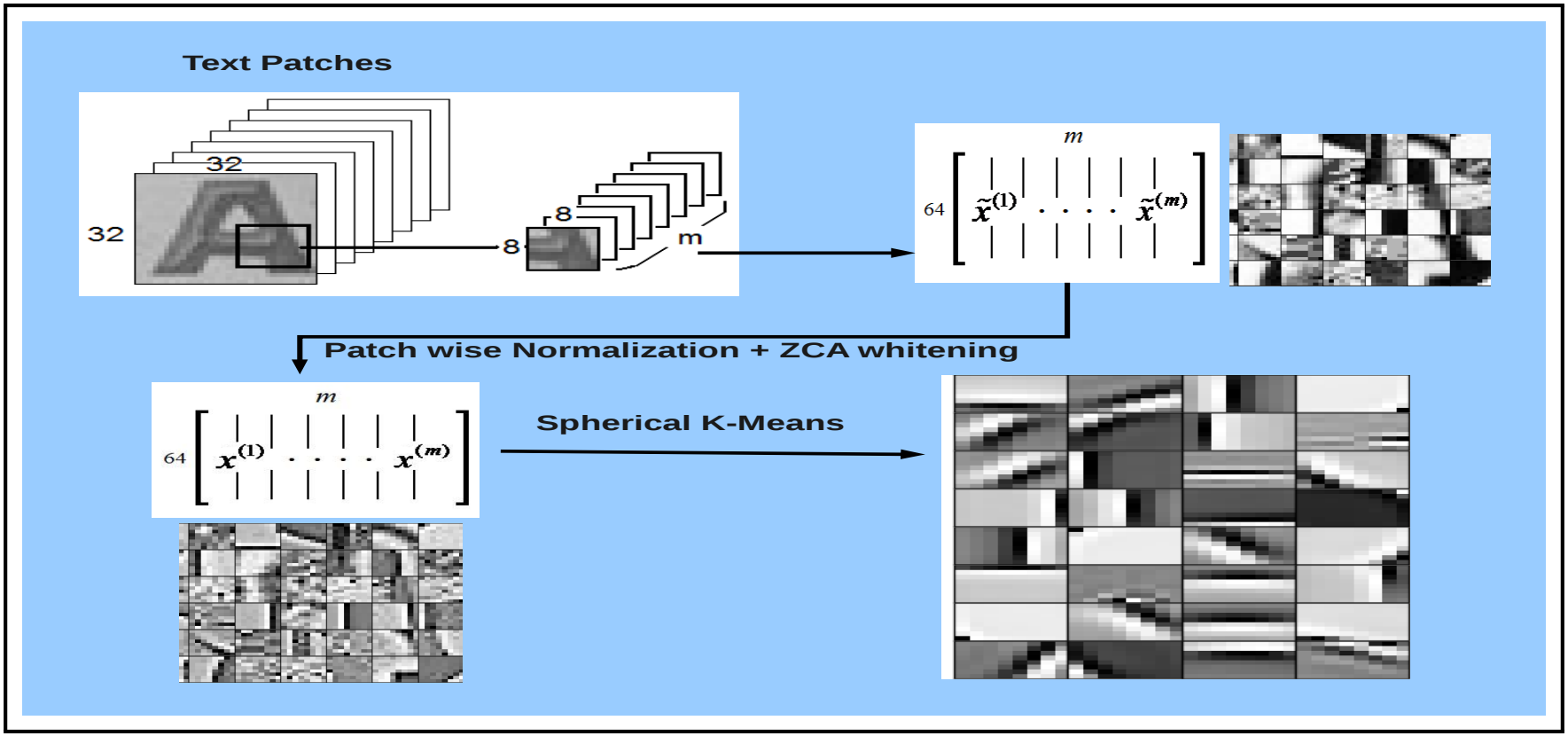


Image Credits: *ICDAR 2003*

[12] A. Coates, B. Carpenter, C. Case, S. Satheesh, B. Suresh, T. Wang, D. J. Wu und A. Y. Ng: *Text detection and character recognition in scene images with unsupervised feature learning*. In: *Document Analysis and Recognition (ICDAR), 2011 International Conference on*, S. 440–445. IEEE, 2011

Feature Extraction [13]

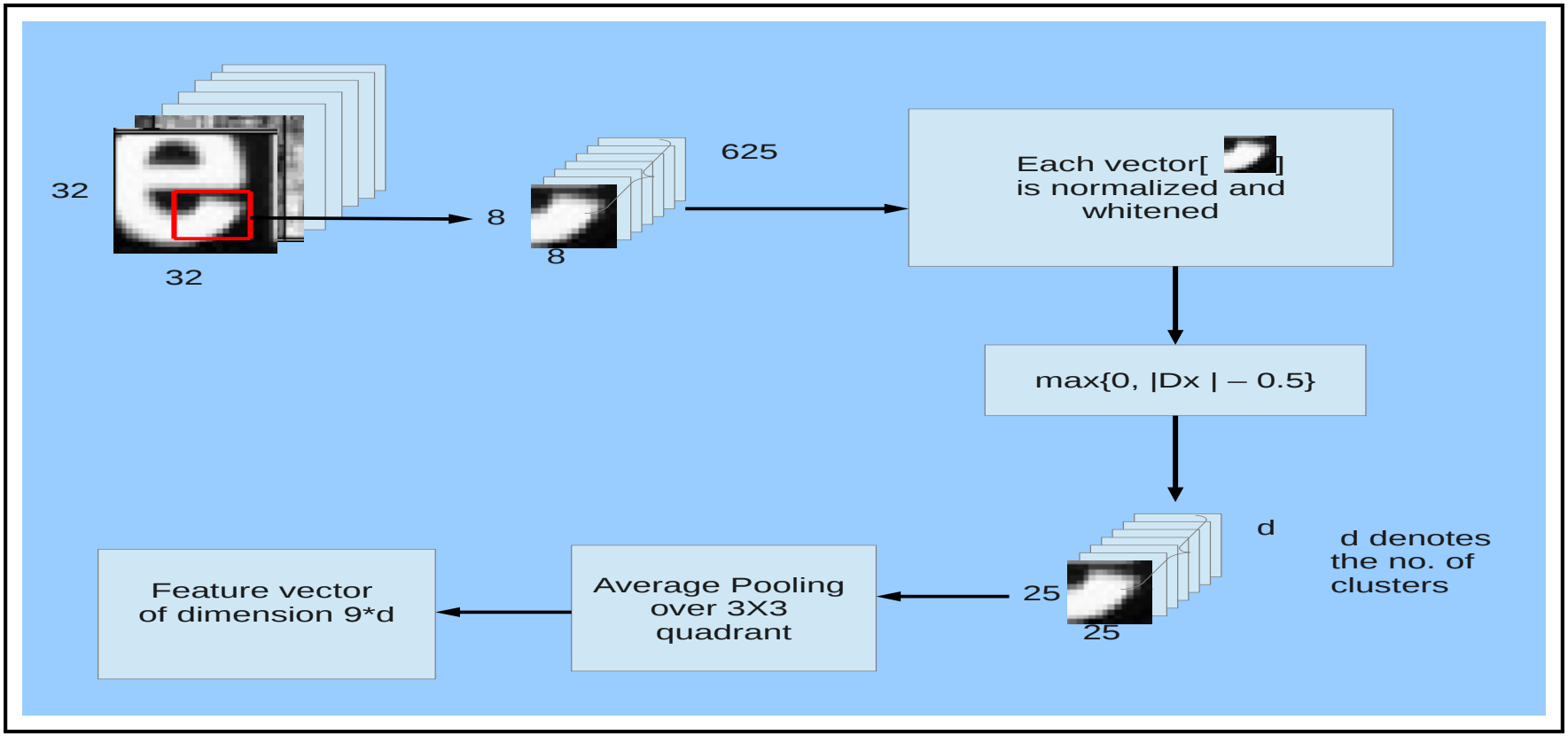


Image Credits: *ICDAR 2003*

[13] A. Coates, B. Carpenter, C. Case, S. Satheesh, B. Suresh, T. Wang, D. J. Wu und A. Y. Ng: *Text detection and character recognition in scene images with unsupervised feature learning*. In: *Document Analysis and Recognition (ICDAR), 2011 International Conference on*, S. 440–445. IEEE, 2011

Training Data Generation [14]

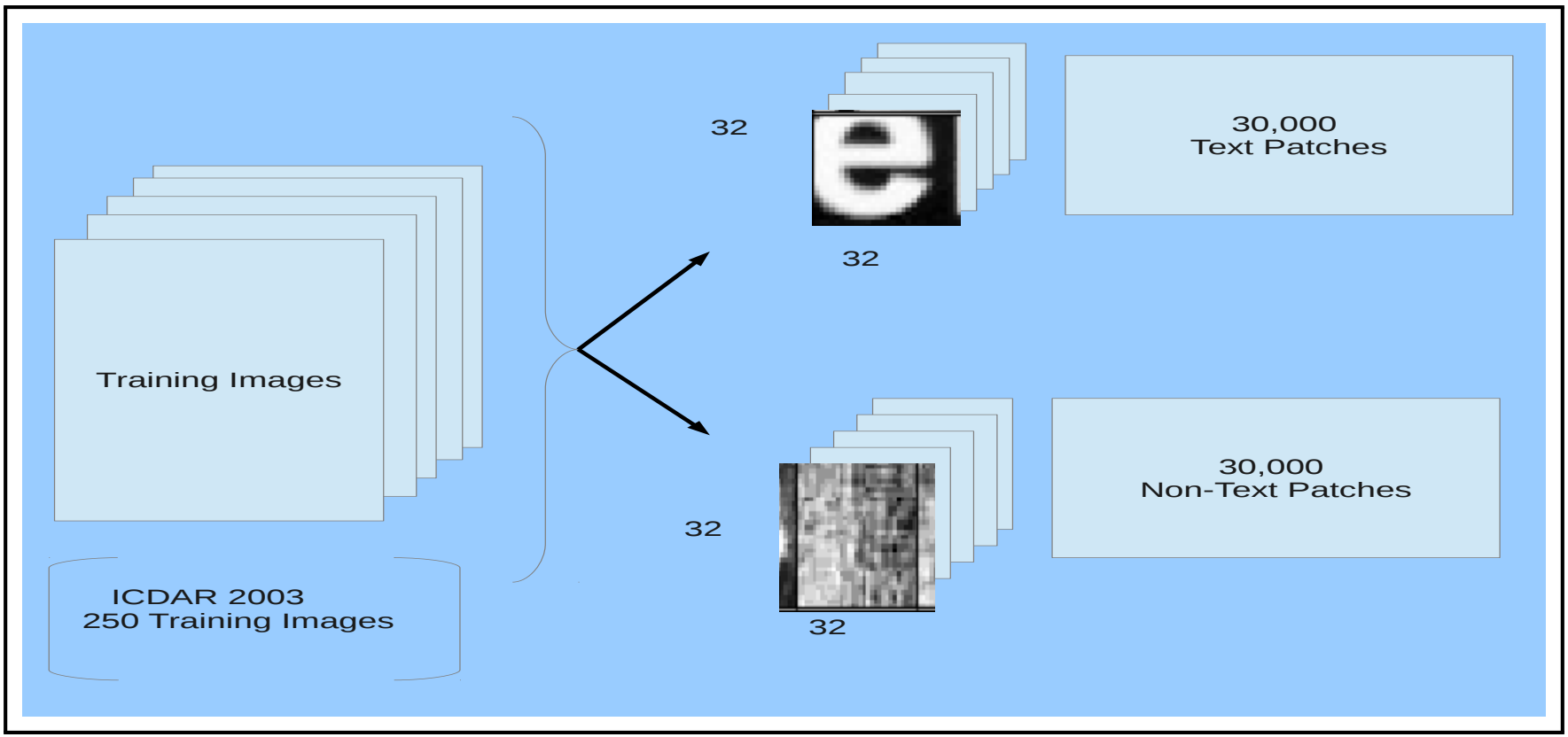


Image Credits: *ICDAR 2003*

[14] A. Coates, B. Carpenter, C. Case, S. Satheesh, B. Suresh, T. Wang, D. J. Wu und A. Y. Ng: *Text detection and character recognition in scene images with unsupervised feature learning*. In: *Document Analysis and Recognition (ICDAR), 2011 International Conference on*, S. 440–445. IEEE, 2011

Text Patches

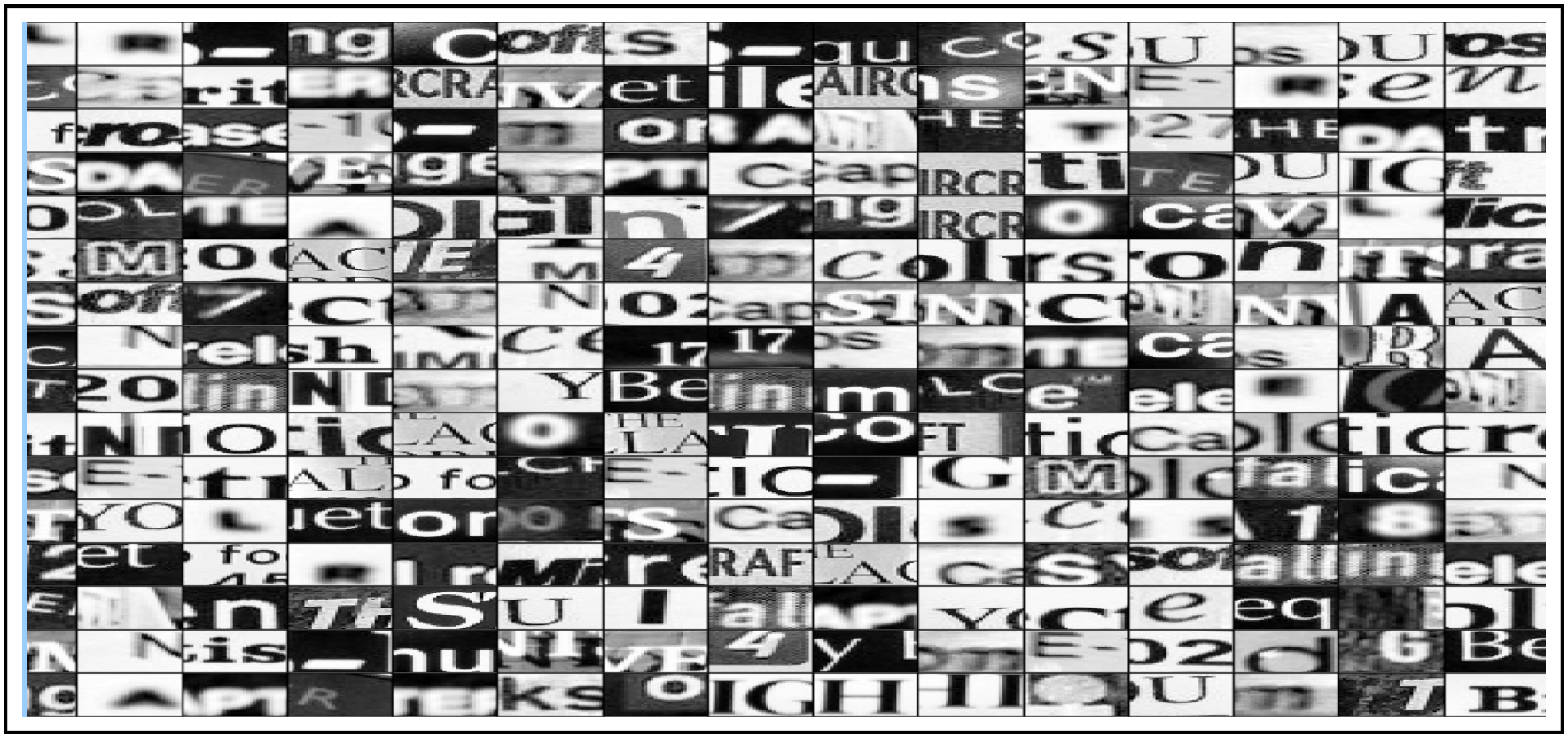


Image Credits: *ICDAR 2003*

Non Text Patches

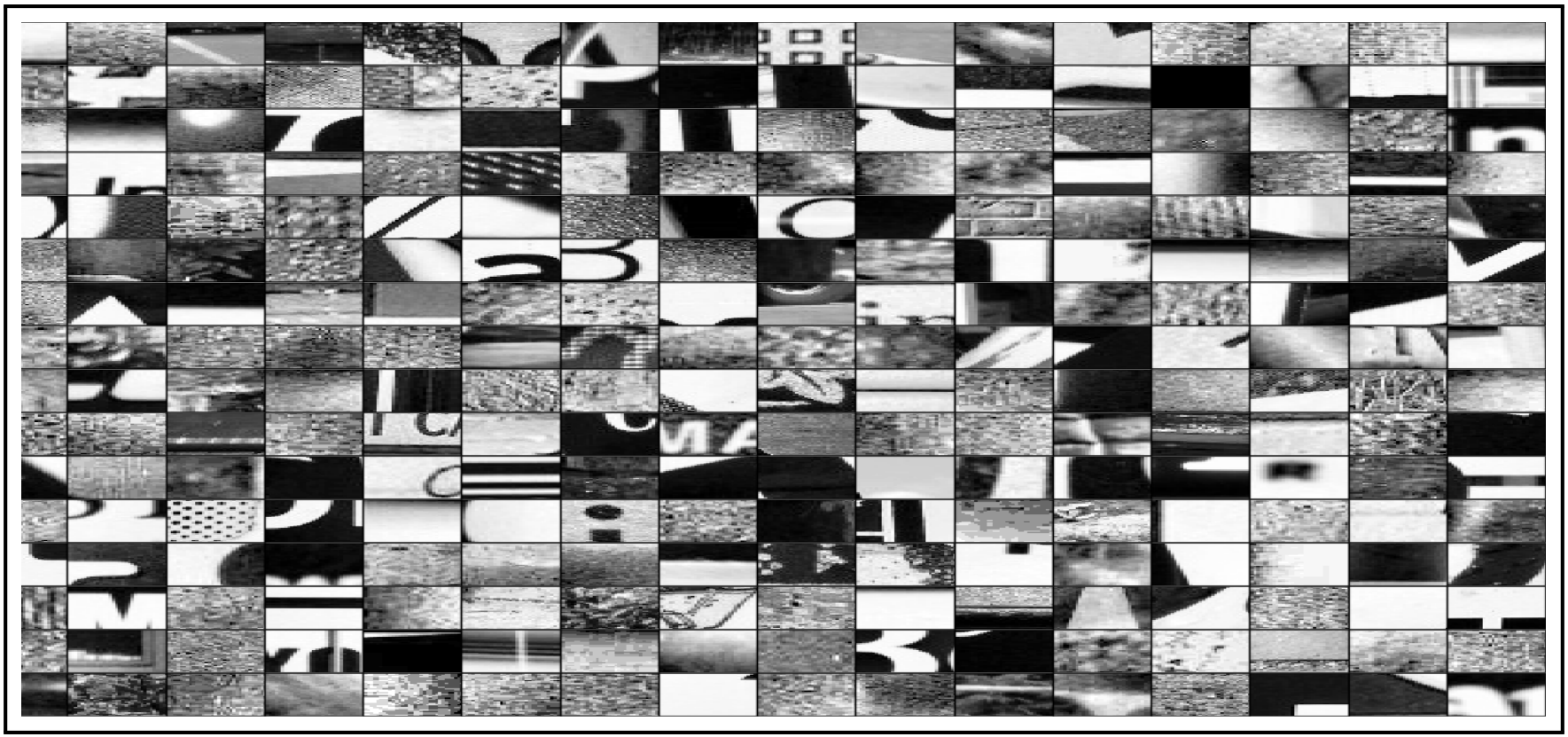


Image Credits: *ICDAR 2003*

SVM Classifier Training [15]

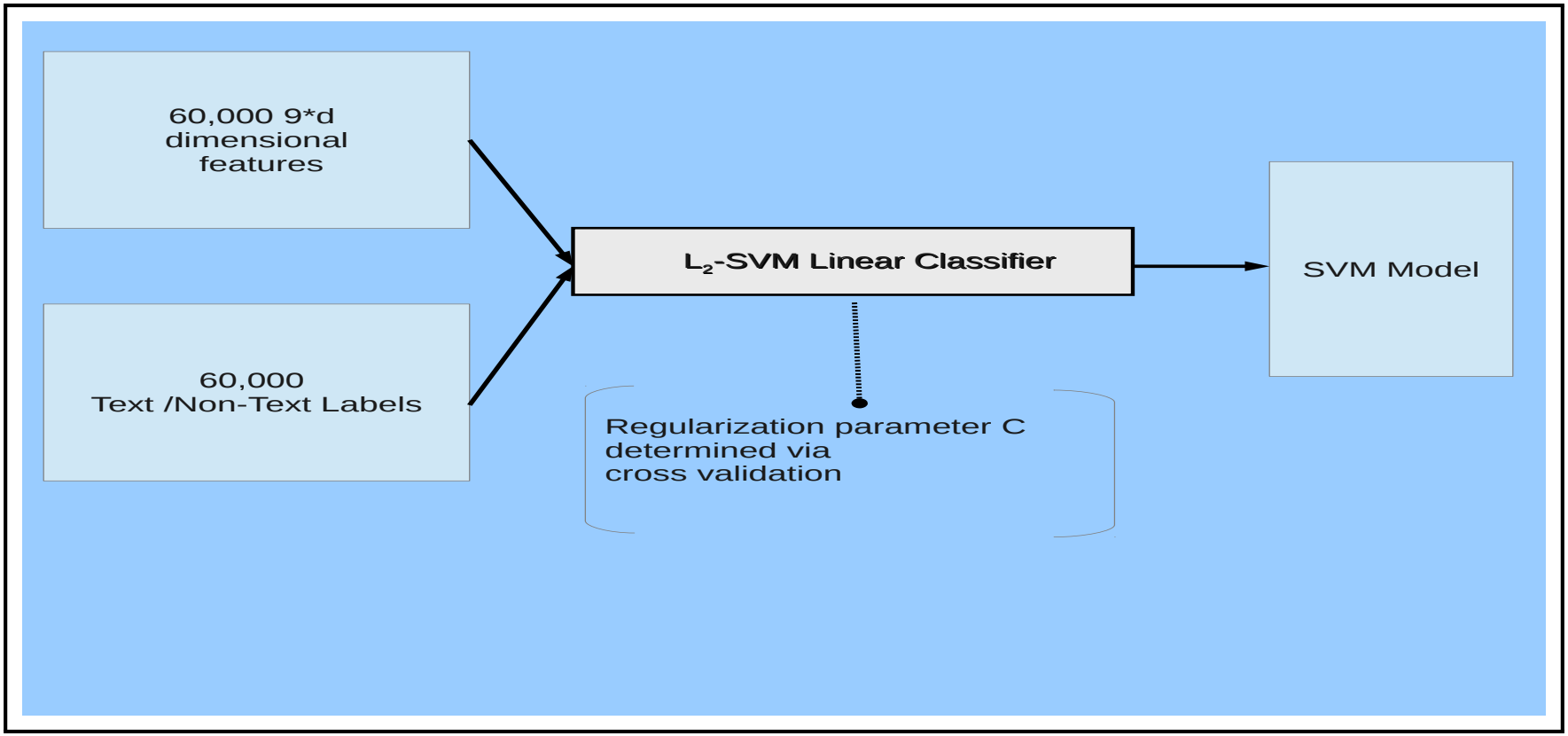
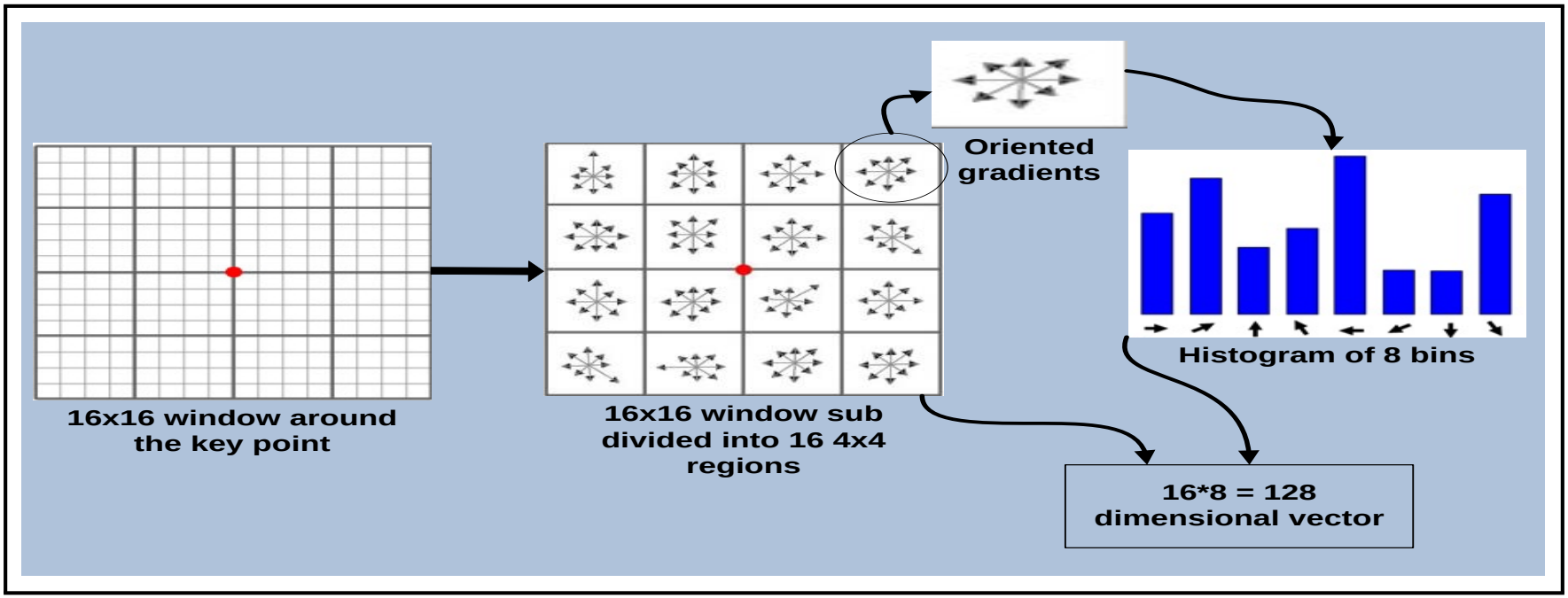


Image Credits: *ICDAR 2003*

[15] A. Coates, B. Carpenter, C. Case, S. Satheesh, B. Suresh, T. Wang, D. J. Wu und A. Y. Ng: *Text detection and character recognition in scene images with unsupervised feature learning*. In: *Document Analysis and Recognition (ICDAR), 2011 International Conference on*, S. 440–445. IEEE, 2011

Key Point Representation - SIFT Descriptor [16]



- 16x16 region around the key point is considered
- Each 4x4 sub region, histogram of oriented gradients is calculated with 8 bins each
- All the values from these histograms (16*8=128) form the 128 element descriptor

[16] D. G. Lowe: *Distinctive image features from scale-invariant keypoints*. International journal of computer vision, 60(2):91–110, 2004

SVM Classifier Training [17]

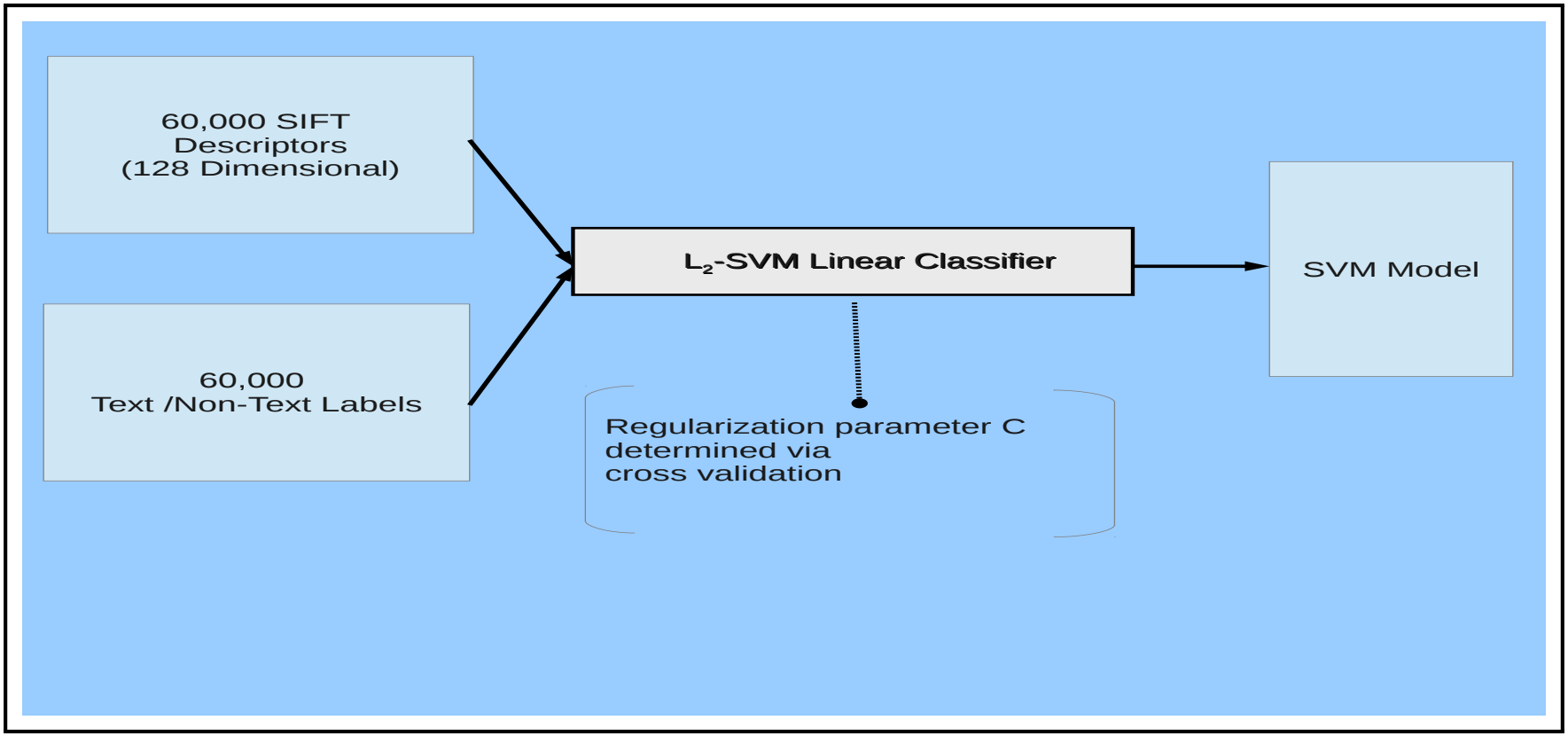


Image Credits: *ICDAR 2003*

[17] A. Coates, B. Carpenter, C. Case, S. Satheesh, B. Suresh, T. Wang, D. J. Wu und A. Y. Ng: *Text detection and character recognition in scene images with unsupervised feature learning*. In: *Document Analysis and Recognition (ICDAR), 2011 International Conference on*, S. 440–445. IEEE, 2011

Evaluation Method

- Both the text detectors are evaluated in similar fashion

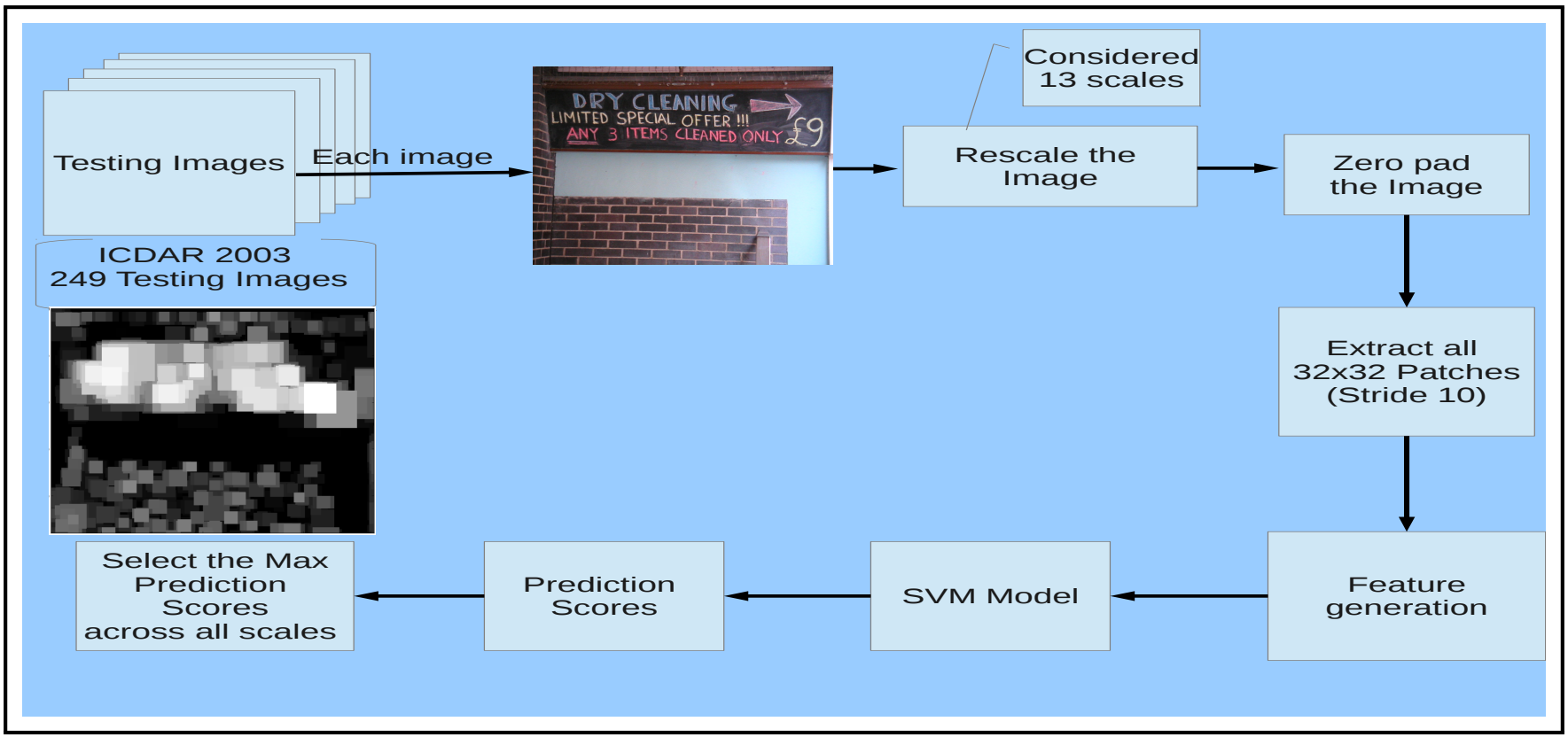
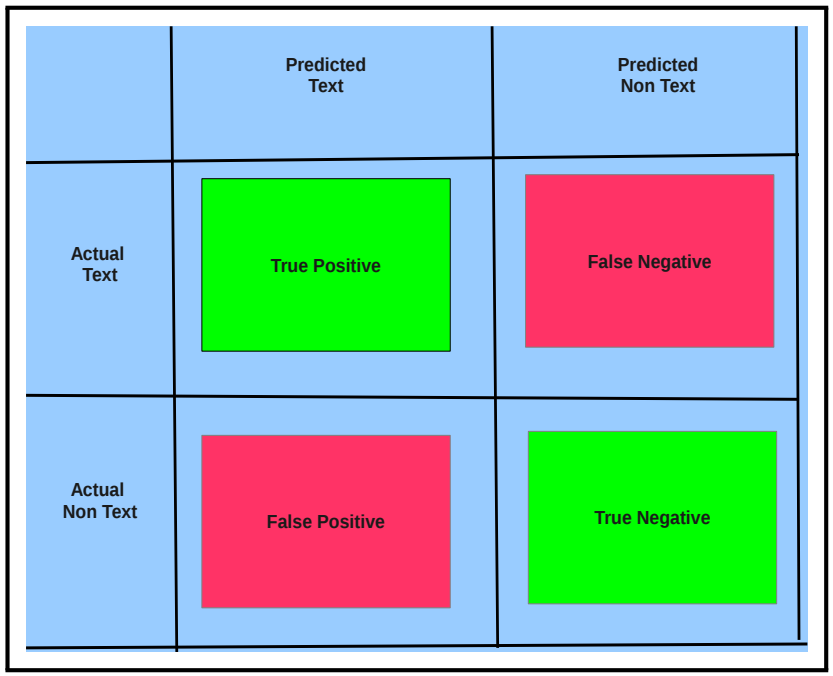


Image Credits: *ICDAR 2003*

Evaluation Metric

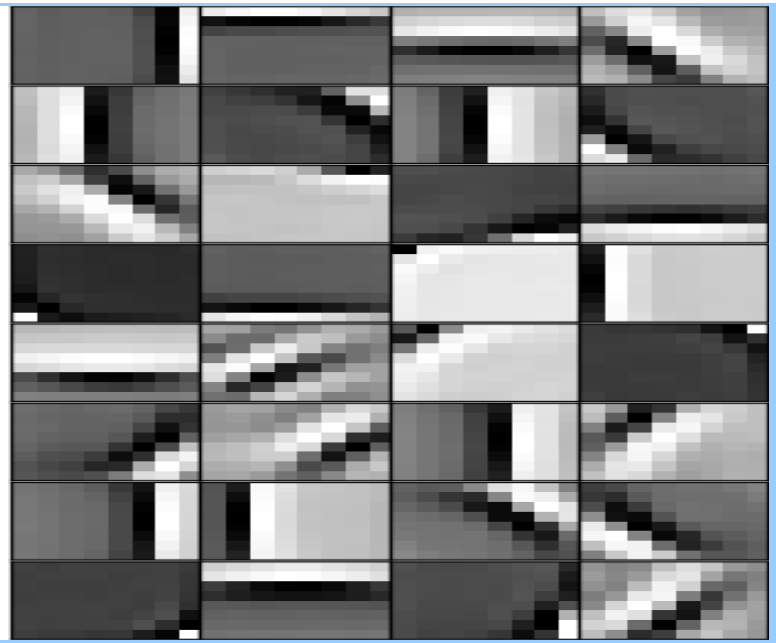
- Fixed thresholds considered across all test images
- For each test image, at each threshold level, True Positive, False Positive and False Negative values were calculated
- For whole dataset, at each threshold level Accumulate True Positive, False Positive and False Negative values and then calculate Precision and Recall



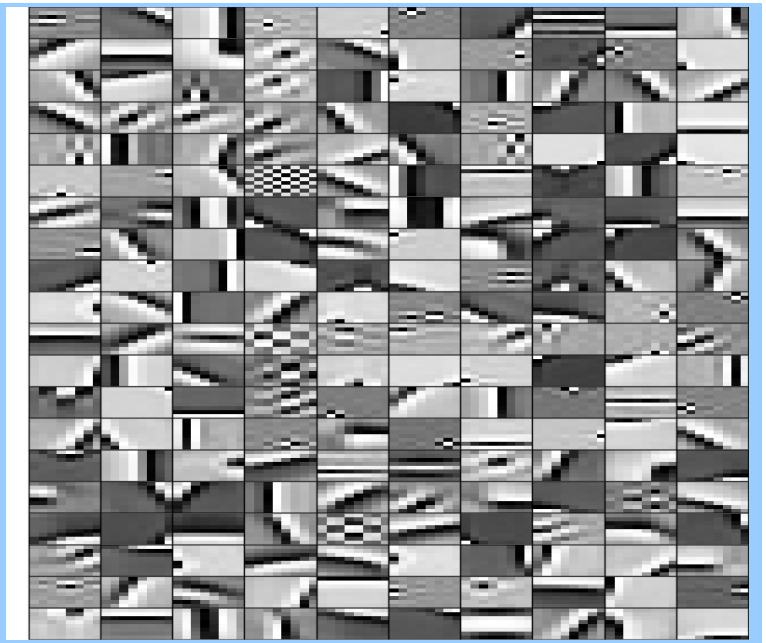
$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negative}}$$

Visual Vocabulary Comparison



Visual Vocabulatory (32 Centeroids)

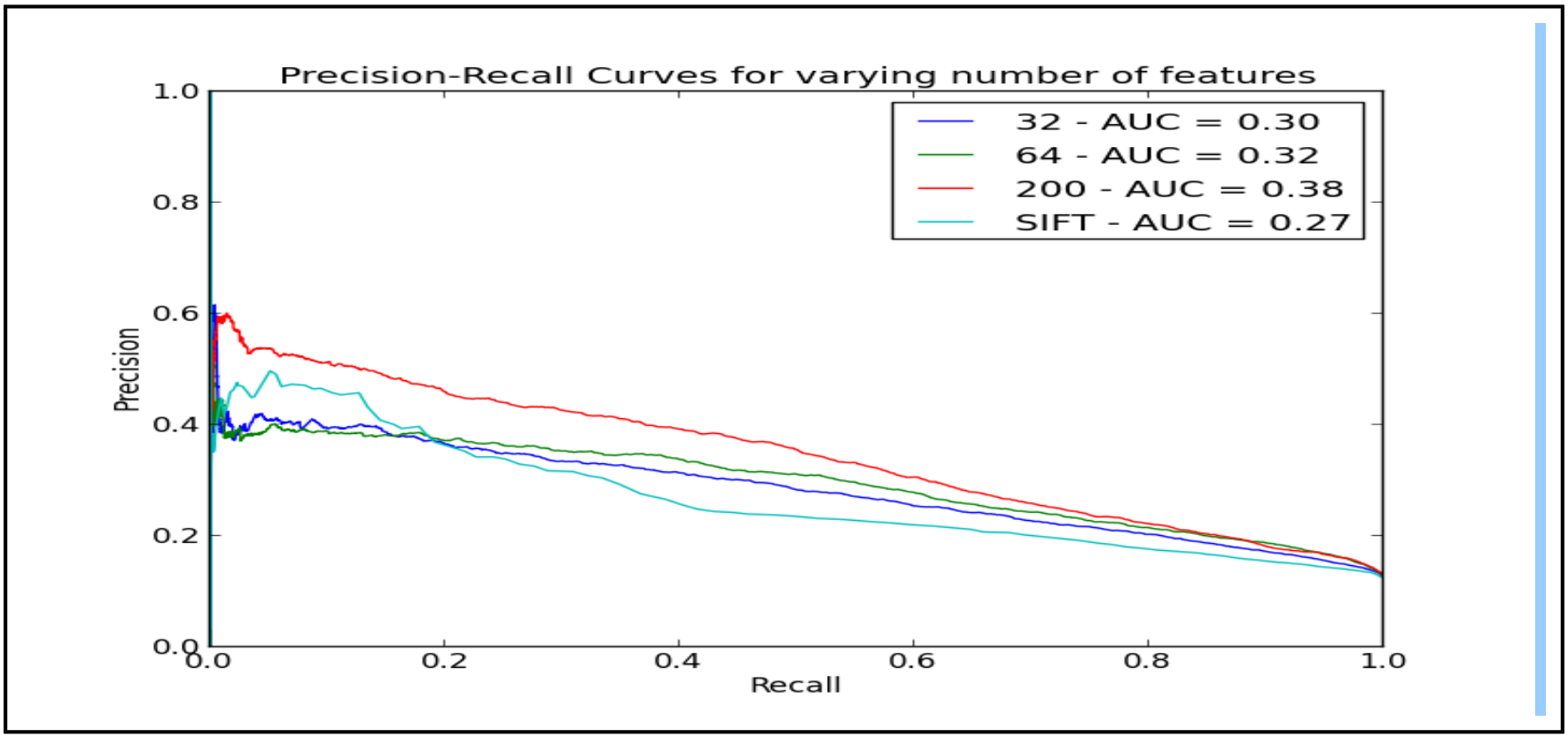


Visual Vocabulatory (200 Centeroids)

Effect of Varying Number of Centeroids

Number of Centeroids	Classification Accuracy	Area Under PR Curve
32	83.25	0.30
64	84.60	0.32
200	87.52	0.38
SIFT Descriptor	78.7	0.27

Effect of Varying Number of Centeroids



Text Detection Results

Text Detection Results

Maximally Stable Extremal Regions (MSER) [18]

Characteristics of MSER

- Connected components with similar intensity values bounded by contrasting backgrounds

Image Credits: *Matas et al.*

[18] J. Matas, O. Chum, M. Urban und T. Pajdla: *Robust wide-baseline stereo from maximally stable extremal regions*. Image and vision computing, 22(10):761–767, 2004

Maximally Stable Extremal Regions (MSER) [19]

- Virtually unchanged over a range of thresholds

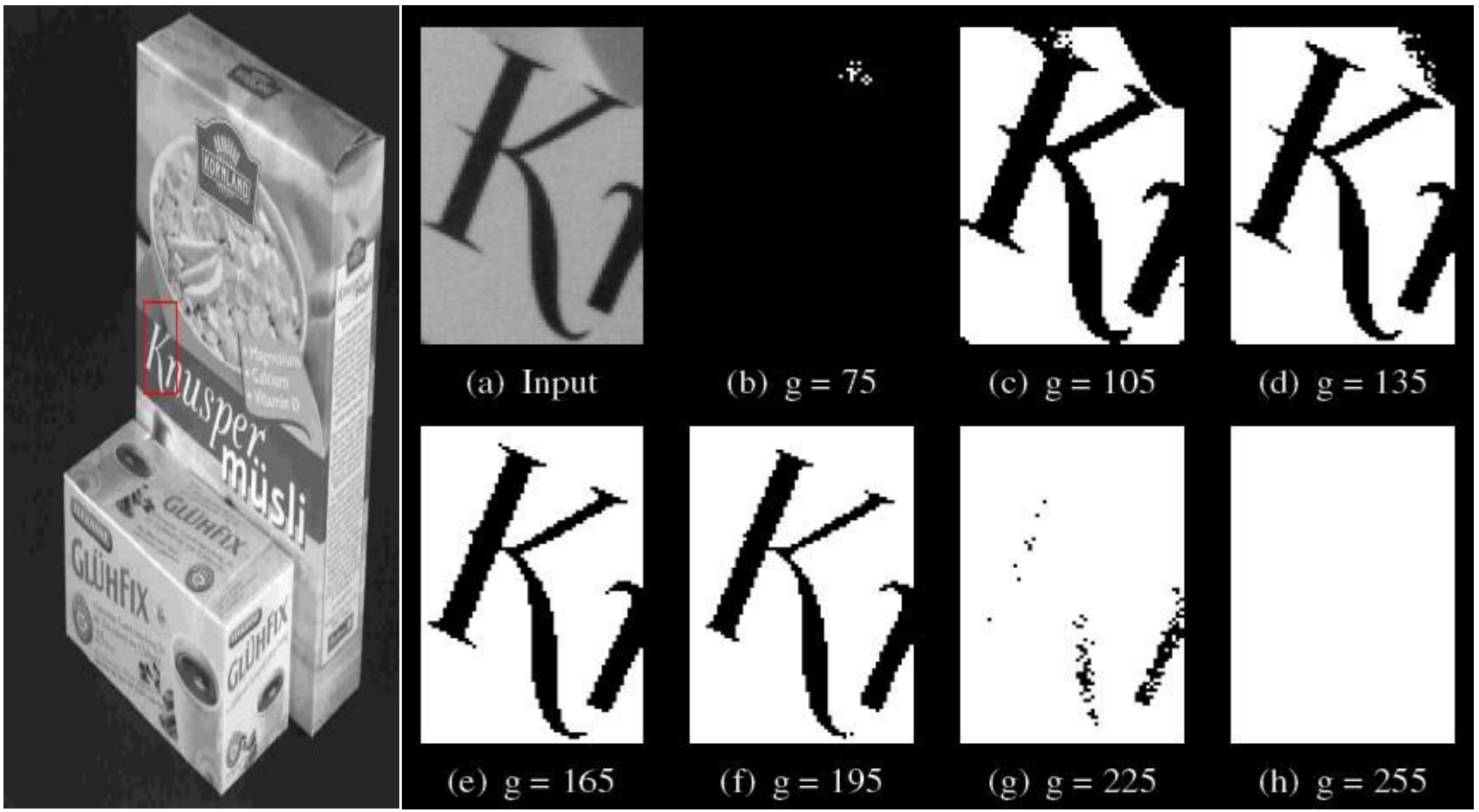


Image Credits: *Matas et al.*

[19] J. Matas, O. Chum, M. Urban und T. Pajdla: *Robust wide-baseline stereo from maximally stable extremal regions*. Image and vision computing, 22(10):761–767, 2004

Why Maximally Stable Extremal Regions (MSER) ?

- MSER detector performs well on images containing homogeneous regions with distinctive boundaries [20]
- Hence an MSER detector can provide relevant interest regions for text detection

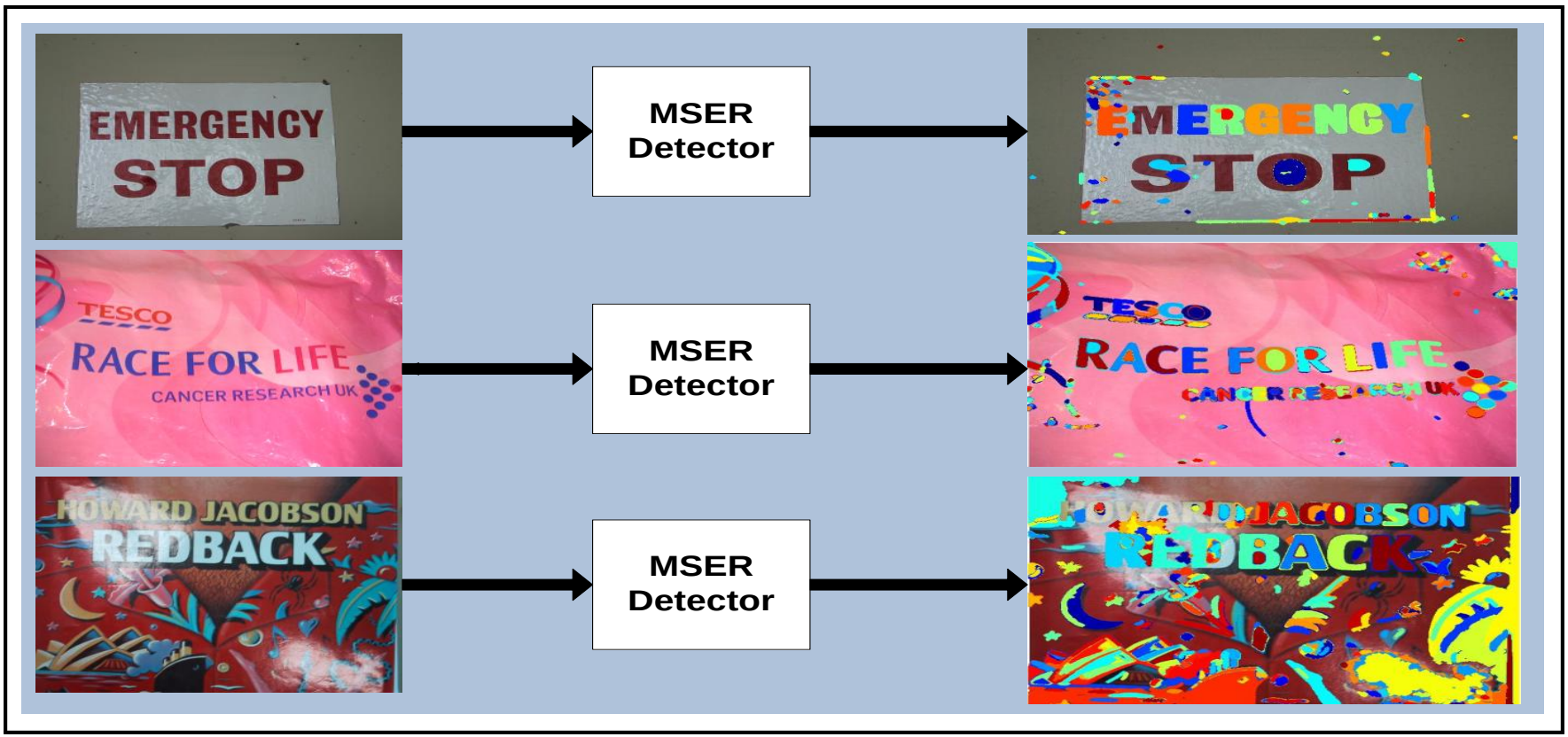


Image Credits: *ICDAR2003*

[20] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir und L. Van Gool: *A comparison of affine region detectors*. International journal of computer vision, 65(1-2):43–72, 2005

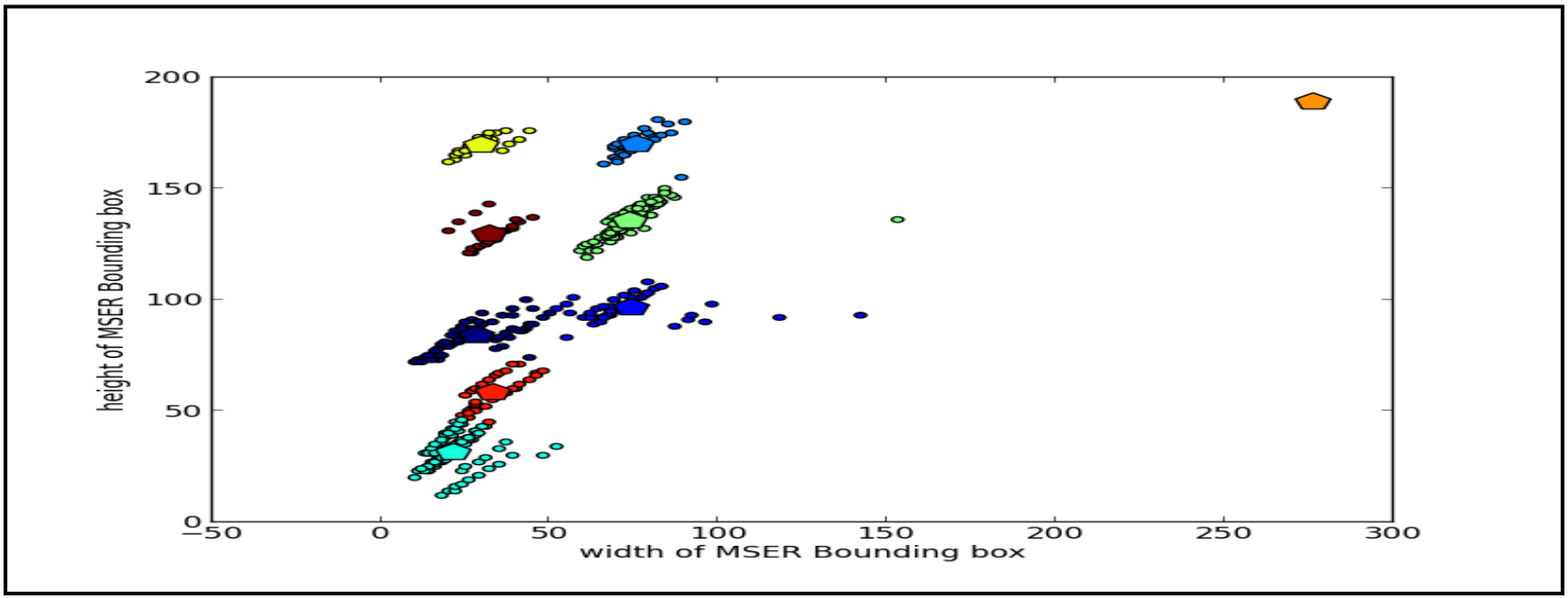
Scale detection using Maximally Stable Extremal Regions (MSER)

- Open CV MSER detector

Image Credits: *ICDAR2003*

Scale detection using Maximally Stable Extremal Regions (MSER)

- Cluster (width of bounding box, height of bounding box) using Lloyds algorithm
- Run the clustering for Number of centeroids = 1 till Number of centeroids = 13
- Select the best cluster using GAP statistics [21]



[21] R. Tibshirani, G. Walther und T. Hastie: *Estimating the number of clusters in a data set via the gap statistic*. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 63(2):411–423, 2001

Questions?


Thank You