2007

Automatic annotation of digital photos

Wenbin Shao
University of Wollongong, wenbin@uow.edu.au

Follow this and additional works at: https://ro.uow.edu.au/theses

Recommended Citation

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au
NOTE

This online version of the thesis may have different page formatting and pagination from the paper copy held in the University of Wollongong Library.

UNIVERSITY OF WOLLONGONG

COPYRIGHT WARNING

You may print or download ONE copy of this document for the purpose of your own research or study. The University does not authorise you to copy, communicate or otherwise make available electronically to any other person any copyright material contained on this site. You are reminded of the following:

Copyright owners are entitled to take legal action against persons who infringe their copyright. A reproduction of material that is protected by copyright may be a copyright infringement. A court may impose penalties and award damages in relation to offences and infringements relating to copyright material. Higher penalties may apply, and higher damages may be awarded, for offences and infringements involving the conversion of material into digital or electronic form.
Automatic Annotation of Digital Photos

A thesis submitted in partial fulfilment of the requirements for the award of the degree

Master of Engineering by Research

from

UNIVERSITY OF WOLLONGONG

by

Wenbin Shao
Master of Engineering Studies

School of Electrical, Computer and Telecommunications Engineering

August 2007
Statement of Originality

I, Wenbin Shao, declare that this thesis, submitted in partial fulfilment of the requirements for the award of Master of Engineering - Research, in the School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, is wholly my own work unless otherwise referenced or acknowledged. The document has not been submitted for qualifications at any other academic institution.

Wenbin Shao
August 31, 2007
# Contents

Notation and Acronyms .......................... XVII  

Abstract ........................................... XXI  

Acknowledgments ................................... XXIII  

1 Introduction .................................... 1  
   1.1 Research objective .......................... 3  
   1.2 Thesis organisation ......................... 4  
   1.3 Contributions ................................ 5  
   1.4 Publications ................................ 7  

2 Literature review ............................... 9  
   2.1 Content-based image retrieval system .......... 10  
      2.1.1 Image contents ......................... 11  
      2.1.2 Image query ............................ 12  
      2.1.3 Semantic gap .......................... 13  
      2.1.4 CBIR applications ...................... 13  
   2.2 Low-level features for CBIR ................ 15
## Contents

2.2.1 Colour ......................................................... 15  
2.2.2 Texture ....................................................... 17  
2.2.3 Shape ......................................................... 18  

2.3 Automatic semantic annotation ................................. 19  
2.3.1 Classification of indoor versus outdoor images .......... 19  
2.3.2 Classification of cityscape versus landscape images .... 20  
2.3.3 Semantics-sensitive approach and linguistic indexing .... 20  
2.3.4 Classification of web images ................................ 21  
2.3.5 Frequent keyword mining .................................... 22  
2.3.6 Cross-media relevance model and model space approach . 22  
2.3.7 Subspace clustering and description logics ................. 23  
2.3.8 Region classification approach and salient objects ....... 24  
2.3.9 A Bayesian framework for image classification .......... 25  
2.3.10 Pairwise constrained clustering and semi-naïve Bayesian model ............................................. 25  

2.4 Similarity measure and indexing .............................. 26  

2.5 Interaction with users and system evaluation ................. 28  

2.6 Chapter summary ................................................. 30  

3 Visual features ................................................... 31  

3.1 Overview of MPEG-7 visual descriptors ....... 32  

3.2 MPEG-7 colour descriptors ................................... 34  
3.2.1 Dominant colour ............................................ 37  
3.2.2 Scalable colour .............................................. 38
3.2.3 Colour structure ........................................... 38
3.2.4 Colour layout .............................................. 39
3.3 MPEG-7 texture descriptors ................................. 40
   3.3.1 Homogeneous texture .................................... 40
   3.3.2 Texture browsing ......................................... 41
   3.3.3 Edge histogram ........................................... 41
3.4 MPEG-7 shape descriptors .................................... 42
   3.4.1 Region-based shape ....................................... 43
   3.4.2 Contour-based shape ..................................... 43
   3.4.3 Image segmentation methods ............................ 44
3.5 Proposed gradient direction histogram ....................... 48
   3.5.1 Gradient image calculation ............................... 49
   3.5.2 Normalization ............................................ 49
3.6 Chapter summary ............................................. 51

4  Pattern classification techniques ............................... 53
   4.1 Classifiers .................................................. 54
      4.1.1 Linear and quadratic classifiers ....................... 54
      4.1.2 k-nearest neighbours .................................. 55
      4.1.3 Bayes classifier ....................................... 55
      4.1.4 Neural networks ...................................... 56
   4.2 Support vector machines ................................... 57
      4.2.1 Mathematical background ............................ 58
      4.2.2 Kernel approach .................................... 60
4.2.3  Training parameters ........................................... 62

4.3  Multi-class support vector machines ................................. 64

4.3.1  One-versus-all SVMs ........................................... 65

4.3.2  Pair-wise SVMs .................................................. 66

4.3.3  Decision directed acyclic graph SVMs ......................... 68

4.3.4  Feature-pool multi-class SVMs ................................. 70

4.4  Chapter summary ................................................... 71

5  Two-class image classification ........................................ 72

5.1  The proposed approach ............................................. 73

5.2  Data collection ...................................................... 76

5.3  Visual feature extraction .......................................... 77

5.4  Experimental steps .................................................. 79

5.5  Two-class classification: landscape versus cityscape .......... 82

5.5.1  Analysis of visual descriptors ................................... 82

5.5.2  Improving the system ............................................ 83

5.5.3  Comparison with other techniques ............................... 85

5.6  Two-class classification for four categories ...................... 86

5.7  Chapter summary ................................................... 90

6  Multi-class image classification ....................................... 91

6.1  The proposed approach ............................................. 92

6.2  Multi-class annotation using SVMs ................................ 93

6.2.1  Using one-versus-all SVMs ..................................... 94

6.2.2  Using pair-wise SVMs with a single feature .................... 94
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.2.3</td>
<td>Using pair-wise SVMs with multiple features</td>
<td>95</td>
</tr>
<tr>
<td>6.2.4</td>
<td>Using decision directed acyclic graph SVMs</td>
<td>98</td>
</tr>
<tr>
<td>6.3</td>
<td>System performance under different conditions</td>
<td>100</td>
</tr>
<tr>
<td>6.3.1</td>
<td>Image cropping</td>
<td>100</td>
</tr>
<tr>
<td>6.3.2</td>
<td>Image resizing</td>
<td>103</td>
</tr>
<tr>
<td>6.3.3</td>
<td>Image rotation</td>
<td>104</td>
</tr>
<tr>
<td>6.4</td>
<td>Comparison with k-nearest neighbour classifiers</td>
<td>106</td>
</tr>
<tr>
<td>6.5</td>
<td>Comparison with neural networks</td>
<td>106</td>
</tr>
<tr>
<td>6.6</td>
<td>Chapter summary</td>
<td>108</td>
</tr>
<tr>
<td>7</td>
<td>Conclusion</td>
<td>109</td>
</tr>
<tr>
<td>7.1</td>
<td>Research summary</td>
<td>109</td>
</tr>
<tr>
<td>7.2</td>
<td>Conclusion</td>
<td>112</td>
</tr>
<tr>
<td>References</td>
<td></td>
<td>114</td>
</tr>
<tr>
<td>Appendices</td>
<td></td>
<td>126</td>
</tr>
<tr>
<td>A</td>
<td>Two-class SVM results</td>
<td>127</td>
</tr>
<tr>
<td>A.1</td>
<td>Using support vector machines</td>
<td>127</td>
</tr>
<tr>
<td>A.1.1</td>
<td>Landscape versus cityscape</td>
<td>127</td>
</tr>
<tr>
<td>A.1.2</td>
<td>Landscape versus vehicle</td>
<td>127</td>
</tr>
<tr>
<td>A.1.3</td>
<td>Landscape versus portrait</td>
<td>128</td>
</tr>
<tr>
<td>A.1.4</td>
<td>Cityscape versus vehicle</td>
<td>129</td>
</tr>
<tr>
<td>A.1.5</td>
<td>Cityscape versus portrait</td>
<td>130</td>
</tr>
<tr>
<td>A.1.6</td>
<td>Vehicle versus portrait</td>
<td>130</td>
</tr>
</tbody>
</table>
B Multi-class SVM results

B.1 Using one-versus-all SVMs .................................................. 132
B.2 Using pair-wise SVMs with a single feature ............................ 132
B.3 Using pair-wise SVMs with multiple features ......................... 133
B.4 Using DDAG SVMs ............................................................ 134
B.5 Using k-nearest neighbours ............................................... 134
  B.5.1 Using gradient direction histogram ............................... 135
  B.5.2 Using edge histogram ............................................... 136
  B.5.3 Using colour structure ............................................. 139
B.6 Using neural networks ..................................................... 141
List of Figures

1.1 Image representation pyramid. ........................................... 2
1.2 Proposed automatic annotation approach. ............................. 3
2.1 A typical content-based image retrieval system. ...................... 11
2.2 Semantic gap. ................................................................. 13
2.3 Three types of spatial colour histograms. ............................... 16
3.1 MPEG-7 visual descriptors. .................................................. 32
3.2 HSV colour space. ............................................................. 35
3.3 HMMD colour space. .......................................................... 36
3.4 128-cell HMMD quantization. .......................................... 38
3.5 Accumulation of colour structure histogram. .......................... 39
3.6 Frequency domain division layout for HTD. ............................ 40
3.7 Five types of edges .......................................................... 41
3.8 Definition of sub-image and image block ............................... 42
3.9 Watershed leads over-segmentation. ..................................... 45
3.10 Watershed segmentation procedure. .................................... 45
3.11 Two sample images for multiscale segmentation. .................... 46
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.12</td>
<td>Two segmentation results on wavelet level two.</td>
<td>46</td>
</tr>
<tr>
<td>3.13</td>
<td>Two segmentation results on wavelet level three.</td>
<td>47</td>
</tr>
<tr>
<td>3.14</td>
<td>Two segmentation results on wavelet level four.</td>
<td>47</td>
</tr>
<tr>
<td>3.15</td>
<td>Effects of different parameters in watershed segmentation.</td>
<td>48</td>
</tr>
<tr>
<td>3.16</td>
<td>Arrangement of a four-element feature vector.</td>
<td>50</td>
</tr>
<tr>
<td>3.17</td>
<td>Example of gradient direction images.</td>
<td>51</td>
</tr>
<tr>
<td>4.1</td>
<td>Neuron model.</td>
<td>56</td>
</tr>
<tr>
<td>4.2</td>
<td>SVM hyperplanes.</td>
<td>59</td>
</tr>
<tr>
<td>4.3</td>
<td>Mapping makes it possible find a nonlinear decision boundary for non-linear data.</td>
<td>60</td>
</tr>
<tr>
<td>4.4</td>
<td>Original data used for parameter effect test.</td>
<td>62</td>
</tr>
<tr>
<td>4.5</td>
<td>Effects of parameter $\gamma$ on the SVM decision boundaries.</td>
<td>63</td>
</tr>
<tr>
<td>4.6</td>
<td>Effects of parameter $C$ on the SVM decision boundaries.</td>
<td>64</td>
</tr>
<tr>
<td>4.7</td>
<td>One-versus-all SVM training phase.</td>
<td>65</td>
</tr>
<tr>
<td>4.8</td>
<td>Test phase of one-against-all SVMs and pair-wise SVMs.</td>
<td>66</td>
</tr>
<tr>
<td>4.9</td>
<td>Pair-wise SVM training phase.</td>
<td>67</td>
</tr>
<tr>
<td>4.10</td>
<td>A DDAG for four-class problems.</td>
<td>68</td>
</tr>
<tr>
<td>5.1</td>
<td>Proposed two-class image annotation system.</td>
<td>74</td>
</tr>
<tr>
<td>5.2</td>
<td>Five-fold cross validation.</td>
<td>75</td>
</tr>
<tr>
<td>5.3</td>
<td>Examples of landscape images in the dataset of 14400 images.</td>
<td>77</td>
</tr>
<tr>
<td>5.4</td>
<td>Examples of cityscape images in the dataset of 14400 images.</td>
<td>78</td>
</tr>
<tr>
<td>5.5</td>
<td>Examples of vehicle images in the dataset of 14400 images.</td>
<td>79</td>
</tr>
<tr>
<td>5.6</td>
<td>Examples of portrait images in the dataset of 14400 images.</td>
<td>80</td>
</tr>
</tbody>
</table>
5.7 Comparison of the visual features in landscape versus cityscape image classification task, on a test set of 3000 images. 82
5.8 The scale scheme of feature combination. All the data are scaled along the horizontal direction. 84
6.1 Proposed multi-class image annotation system. 93
6.2 Optimized DDAG structure. 99
6.3 The overall classification rates of different multi-class SVMs. 100
6.4 Image cropping parameters. 101
6.5 The overall classification rates when the input images are cropped. 102
A.1 Comparison of the visual descriptors in landscape versus cityscape image classification task, on a test set of 3000 images. 128
A.2 Comparison of the visual descriptors in landscape versus vehicle image classification task, on a test set of 3000 images. 128
A.3 Comparison of the visual descriptors in landscape versus portrait image classification task, on a test set of 3000 images. 129
A.4 Comparison of the visual descriptors in cityscape versus vehicle image classification task, on a test set of 3000 images. 129
A.5 Comparison of the visual descriptors in cityscape versus portrait image classification task, on a test set of 3000 images. 130
A.6 Comparison of the visual descriptors in vehicle versus portrait image classification task, on a test set of 3000 images. 131
List of Tables

2.1 Application areas of CBIR. ........................................... 14
2.2 Summary of articles on automatic annotation. .................. 27
2.3 Classification performance. ........................................... 29

3.1 HSV uniform quantization. ........................................... 37
3.2 Computing time of watershed and normalized cuts (in seconds). 48
3.3 The gradient direction histogram vectors for example images. .... 52

5.1 Database summary. .................................................... 76
5.2 Classification rates of the visual features on test set using SVMs, in landscape versus cityscape problem. .............................. 83
5.3 Classification rates of the k-NN classifier and the EDH feature. ... 85
5.4 Classification rates of two-class SVMs for different visual features, estimated using five-fold cross validation on training sets. .... 87
5.5 Mahalanobis distance between the training set and test set for different visual features. .................................................. 88
5.6 Classification rates of two-class SVMs for different visual features on test sets. ......................................................... 89
<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1</td>
<td>Salient feature summary for six two-class classifiers.</td>
<td>92</td>
</tr>
<tr>
<td>6.2</td>
<td>Classification rates for the one-versus-all SVM method, on the test</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td>set of four classes. The features used are gradient direction</td>
<td></td>
</tr>
<tr>
<td></td>
<td>histogram and edge direction.</td>
<td></td>
</tr>
<tr>
<td>6.3</td>
<td>Confusion matrix of pair-wise SVMs with majority voting, on the</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>test set of four classes.</td>
<td></td>
</tr>
<tr>
<td>6.4</td>
<td>Confusion matrix of pair-wise SVMs with confidence score voting,</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>on the test set of four classes.</td>
<td></td>
</tr>
<tr>
<td>6.5</td>
<td>Feature combination strategies for pair-wise SVMs.</td>
<td>96</td>
</tr>
<tr>
<td>6.6</td>
<td>Confusion matrix of multi-feature pair-wise SVMs with majority</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td>voting, on the test set of four classes (strategy A).</td>
<td></td>
</tr>
<tr>
<td>6.7</td>
<td>Confusion matrix of multi-feature pair-wise SVMs with confidence</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td>score voting, on the test set of four classes (strategy A).</td>
<td></td>
</tr>
<tr>
<td>6.8</td>
<td>Confusion matrix of multi-feature pair-wise SVM with majority</td>
<td>98</td>
</tr>
<tr>
<td></td>
<td>voting, on the test set of four classes (strategy B).</td>
<td></td>
</tr>
<tr>
<td>6.9</td>
<td>Confusion matrix of multi-feature pair-wise SVMs with confidence</td>
<td>98</td>
</tr>
<tr>
<td></td>
<td>score voting, on the test set of four classes (strategy B).</td>
<td></td>
</tr>
<tr>
<td>6.10</td>
<td>Confusion matrix of DDAG SVMs, on the test set of four classes.</td>
<td>98</td>
</tr>
<tr>
<td>6.11</td>
<td>Confusion matrix of optimized DDAG SVMs, on the test set of four</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td>classes.</td>
<td></td>
</tr>
<tr>
<td>6.12</td>
<td>The details for five image cropping tests.</td>
<td>100</td>
</tr>
<tr>
<td>6.13</td>
<td>Confusion matrix of pair-wise SVMs with majority voting, on the</td>
<td>103</td>
</tr>
<tr>
<td></td>
<td>resized image test set of four classes (80% of its original size).</td>
<td></td>
</tr>
</tbody>
</table>
6.14 Confusion matrix of pair-wise SVMs with confidence score voting, on the resized image test set of four classes (80% of its original size). 103

6.15 Confusion matrix of pair-wise SVMs with majority voting, on the resized image test set of four classes (50% of its original size). 104

6.16 Confusion matrix of pair-wise SVMs with confidence score voting, on the resized image test set of four classes (50% of its original size). 104

6.17 Confusion matrix of pair-wise SVMs with majority voting, on the resized image test set of four classes (150% of its original size). 105

6.18 Confusion matrix of pair-wise SVMs with confidence score voting, on the resized image test set of four classes (150% of its original size). 105

6.19 Confusion matrix of pair-wise SVMs with majority voting, on the rotated image test set of four classes (90°). 106

6.20 Confusion matrix of pair-wise SVMs with confidence score voting, on the rotated image test set of four classes (90°). 106

6.21 Confusion matrix of the \(k\)-NN classifier using the proposed GDH feature. 107

6.22 Confusion matrix of the \(k\)-NN classifier using MPEG-7 edge histogram. 107

6.23 Confusion matrix of neural network using gradient direction histogram. 107

7.1 Comparison of SVMs, \(k\)-NN and neural networks. 112
B.1 Classification rates for the one-versus-all SVM method, on the test set of four classes. The features used are gradient direction histogram and edge direction.

B.2 Confusion matrix of pair-wise SVMs with majority voting, on the test set of four classes.

B.3 Confusion matrix of pair-wise SVMs with confidence score voting, on the test set of four classes.

B.4 Confusion matrix of multi-feature pair-wise SVMs with majority voting, on the test set of four classes (strategy A).

B.5 Confusion matrix of multi-feature pair-wise SVMs with confidence score voting, on the test set of four classes (strategy A).

B.6 Confusion matrix of multi-feature pair-wise SVMs with majority voting, on the test set of four classes (strategy B).

B.7 Confusion matrix of multi-feature pair-wise SVMs with confidence score voting, on the test set of four classes (strategy B).

B.8 Confusion matrix of DDAG SVMs, on the test set of four classes.

B.9 Confusion matrix of optimized DDAG SVMs, on the test set of four classes.

B.10 Confusion matrix of the $k$-NN classifier using the proposed GDH feature and $k = 1$, on the test set of four classes.

B.11 Confusion matrix of the $k$-NN classifier using the proposed GDH feature and $k = 3$, on the test set of four classes.

B.12 Confusion matrix of the $k$-NN classifier using the proposed GDH feature and $k = 5$, on the test set of four classes.
B.13 Confusion matrix of the $k$-NN classifier using the proposed GDH feature and $k = 7$, on the test set of four classes. . . . . . . . . . . . 135

B.14 Confusion matrix of the $k$-NN classifier using the proposed GDH feature and $k = 9$, on the test set of four classes. . . . . . . . . . . . 136

B.15 Confusion matrix of the $k$-NN classifier using the proposed GDH feature and $k = 11$, on the test set of four classes. . . . . . . . . . . . 136

B.16 Confusion matrix of the $k$-NN classifier using the proposed GDH feature and $k = 13$, on the test set of four classes. . . . . . . . . . . . 136

B.17 Confusion matrix of the $k$-NN classifier using the proposed GDH feature and $k = 15$, on the test set of four classes. . . . . . . . . . . . 136

B.18 Confusion matrix of the $k$-NN classifier using the MPEG-7 edge histogram and $k = 1$, on the test set of four classes. . . . . . . . . . . . 137

B.19 Confusion matrix of the $k$-NN classifier using the MPEG-7 edge histogram and $k = 3$, on the test set of four classes. . . . . . . . . . . . 137

B.20 Confusion matrix of the $k$-NN classifier using the MPEG-7 edge histogram and $k = 5$, on the test set of four classes. . . . . . . . . . . . 137

B.21 Confusion matrix of the $k$-NN classifier using the MPEG-7 edge histogram and $k = 7$, on the test set of four classes. . . . . . . . . . . . 137

B.22 Confusion matrix of the $k$-NN classifier using the MPEG-7 edge histogram and $k = 9$, on the test set of four classes. . . . . . . . . . . . 137

B.23 Confusion matrix of the $k$-NN classifier using the MPEG-7 edge histogram and $k = 11$, on the test set of four classes. . . . . . . . . . . . 138

B.24 Confusion matrix of the $k$-NN classifier using the MPEG-7 edge histogram and $k = 13$, on the test set of four classes. . . . . . . . . . . . 138
B.25 Confusion matrix of the $k$-NN classifier using the MPEG-7 edge histogram and $k = 15$, on the test set of four classes. .......................... 138

B.26 Confusion matrix of the $k$-NN classifier using the MPEG-7 colour structure and $k = 1$, on the test set of four classes. .......................... 139

B.27 Confusion matrix of the $k$-NN classifier using the MPEG-7 colour structure and $k = 3$, on the test set of four classes. .......................... 139

B.28 Confusion matrix of the $k$-NN classifier using the MPEG-7 colour structure and $k = 5$, on the test set of four classes. .......................... 139

B.29 Confusion matrix of the $k$-NN classifier using the MPEG-7 colour structure and $k = 7$, on the test set of four classes. .......................... 139

B.30 Confusion matrix of the $k$-NN classifier using the MPEG-7 colour structure and $k = 9$, on the test set of four classes. .......................... 140

B.31 Confusion matrix of the $k$-NN classifier using the MPEG-7 colour structure and $k = 11$, on the test set of four classes. .......................... 140

B.32 Confusion matrix of the $k$-NN classifier using the MPEG-7 colour structure and $k = 13$, on the test set of four classes. .......................... 140

B.33 Confusion matrix of the $k$-NN classifier using the MPEG-7 colour structure and $k = 15$, on the test set of four classes. .......................... 140

B.34 Confusion matrix of neural network using gradient direction histogram. ................................................................. 141
Notation and Acronyms

Notation

\( \alpha_i \)  Lagrange multiplier

\( A^T \)  Transpose of matrix A

\( c_{ij} \)  Support vector machine classifier trained from the \( i \)-th class and \( j \)-th class

\( |d| \)  Absolute value of \( d \)

\( D_{ij} \)  Decision function corresponding to \( c_{ij} \)

\( \epsilon_i \)  Slack variable

\( K(x, y) \)  Kernel

\( P(x|\omega) \)  Class-conditional probability density for \( x \) conditioned by \( \omega \)

\( \langle w \cdot x \rangle \)  Dot product between \( w \) and \( x \)

\( ||w|| \)  Euclidean norm of vector \( w \)

\( x \)  A feature vector, \( x = [x_1, x_2, \ldots, x_n]^T \)

\( y_i \)  class label, +1 or -1
Acronyms

**ADAG**  Adaptive directed acyclic graph

**ANMRR**  Average normalized modified retrieval rank

**ARTMAP**  A class of neural networks based on adaptive resonance theory

**CBIR**  Content based image retrieval

**CCV**  Colour coherence vector

**CL**  MPEG-7 colour layout

**CMRM**  Cross-media relevance model

**CR**  Classification rate

**CS**  MPEG-7 colour structure

**CSS**  Curvature Scale-Space

**DC**  MPEG-7 dominant colour

**DCT**  Discrete cosine transform

**DDAG**  Decision directed acyclic graph

**DFT**  Discrete Fourier transform

**DL**  Description Logics

**EDH**  Edge direction histogram
EH  MPEG-7 edge histogram

EM  Expectation maximization

GDH  Gradient direction histogram

HMMD, HSV, LUV, RGB, YCbCr  Colour spaces

HT  MPEG-7 homogeneous texture

HMMD, HSV, LUV, RGB, YCbCr  Colour spaces

k-NN  k-nearest neighbours

LOO  Leave-one-out

LOOCV  Leave-one-out cross validation

MHMM  Multi-resolution hidden Markov model

MPEG  Moving Picture Experts Group

PWC  Pair-wise coupling

SC  MPEG-7 scalable colour

SNB  Semi-naive Bayesian model

SNP  Summation of negative probability

SVM  Support vector machine

VC  Vapnik-Chervonenkis

XM  MPEG-7 eXperimentation Model
In this thesis, the term SVM refers to two-class classification problems. The terms pair-wise SVM and one-versus-all SVM refer to multi-class classification problems.
Abstract

Content-based image retrieval searches for an image by using a set of visual features that characterize the image content. This technique has been used in many areas, such as geographical information processing, space science, biomedical image processing, target recognition in military applications and bioinformatics. Many approaches have been proposed to reduce the gap between the low-level visual features and high-level contents. In this thesis, a multi-class automatic annotation system is developed to bridge the semantic gap. Given an image, the proposed system will automatically generate keywords corresponding to the image contents. The system is evaluated using a large image database consisting of over 16000 images collected from various online repositories.

The proposed multi-class annotation system is based on salient features and support vector machines (SVMs). A new feature called gradient direction histogram is proposed for image classification. Instead of relying on a single feature, the SVMs in our system can automatically select the most suitable features from a pool of six MPEG-7 visual descriptors and the proposed gradient direction histogram. Multi-class SVMs are constructed using two-class SVMs in different combinations.
We have examined several multi-class support vector machines including one-versus-all SVMs, pair-wise SVMs and decision directed acyclic graph SVMs. The results confirm that the pair-wise and decision directed acyclic graph SVMs are suitable for multi-class applications. In pair-wise SVMs, we propose a voting scheme named confidence score voting. Our results show that, compared to majority voting, confidence score voting improves the classification accuracy. Combining salient features leads to a significant improvement in the classification rate.

The proposed system is compared to \( k \)-nearest neighbours and neural networks using the same dataset. The results show that the proposed system outperforms these two classifiers in the four-class classification problem. The research project also investigates the system performance when the input image is cropped, resized or rotated.
Acknowledgments

I would like to express my gratitude to my Parents and Sisters, who have supported me during my studies and research projects.

I also want to thank my principal supervisor, Associate Professor Golshah Naghdy, for all of her guidance, counsel, and technical support.

Special thanks also go to my co-supervisor Dr. Son Lam Phung for all his time, assistance, knowledge and provision of the image data used in my research project.

Moreover, I gratefully acknowledge the ongoing support of the staff of the School of Electrical, Computer and Telecommunications Engineering for giving me personal and professional support during my studies at the University of Wollongong.

Finally thanks to my fellow students and friends, who have helped me during my study at the University.