

# **Automated Searching of Illicit Drug Logos and Imprints**

by

Ping Hei Ronnie Ng

A thesis submitted for the Degree of Doctor of  
Philosophy (Science)

2015

School of Chemistry and Forensic Science

Faculty of Science

University of Technology, Sydney

## **Certificate of Original Authorship**

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Ping Hei Ronnie Ng

# Acknowledgements

I would like to thank the University of Technology, Sydney for providing me this opportunity to carry out my own research in the forensic science discipline. In particular, the following people have given me a lot of guidance and help throughout the project. This project would not have been so successful without them.

1. Dr. Brian Reedy: He has given me a lot of assistance and help throughout this research. I enjoyed interpreting and discussing this project with him. His insights have inspired me with new ideas in numerous occasions. His patience and encouragement throughout this period means a lot to me. He has also dedicated a lot of his personal time for my project, especially the time spent in reading and giving me valuable comments on my literature review and my thesis.
2. Dr. Mark Tahtouh: He offered this interesting project to me. He provided me access to the AFP instruments (macroscopes, twin swan lights and ring lights etc), a genuine pill press and different kinds of dies for pressing and photographing my own tablets. The images that were processed in this project were therefore collected as they would have been in real forensic examination work.
3. Martin Lopatka: Thank you for your idea about matching side profiles of tablets and your suggestions for my MATLAB scripts.
4. The Australian Federal Police: Thank you for providing funding for my project.
5. Joshua Abraham: Thank you for your help in MATLAB programming and suggestions for various image processing methods, transformations and operations.
6. Mum, Dad and sis: I have unlimited support from all of you. I would not be here without you.

# Table of Contents

Chapter 1: Introduction .....	1
1.1 Background.....	2
1.2 Content-based Image Retrieval (CBIR).....	3
1.3 Low-level Feature Extraction .....	4
1.3.1 Image Pre-processing.....	5
1.3.2 Colour .....	7
1.3.3 Texture .....	11
1.3.4 Shape.....	15
1.3.5 Segmentation .....	23
1.3.6 Colour Layout .....	26
1.3.7 Others.....	29
1.4 Searching algorithms.....	31
1.4.1 Minkowski-type Metric .....	31
1.4.2 Normalised Inner Product.....	33
1.4.3 Quadratic Form Distance .....	33
1.4.4 Mahalanobis Distance .....	34
1.4.5 Kullback-Leibler Divergence/Distance .....	35
1.4.6 Others.....	35
1.5 Database for CBIR and Performance Evaluation.....	37
1.6 Drug Tablet Searching .....	38
1.7 The Semantic Gap .....	40
1.7.1 Object Ontology .....	40
1.7.2 Machine Learning.....	41

1.7.3	Relevance Feedback (RF) .....	43
1.7.4	Semantic Template .....	44
1.7.5	Web Image Retrieval.....	44
1.7.6	Image Classification.....	45
1.8	Scope and Aims .....	47
Chapter 2:	Logo extraction from existing AFP illicit drug tablet images .....	48
2.1	Method.....	49
2.2	Edge Detection .....	50
2.2.1	The “Canny” edge detection .....	50
2.2.2	Other edge detection filters.....	52
2.2.3	Smoothing images before edge detection by resizing.....	53
2.3	Phase Congruency .....	54
2.4	Conversion to BW images .....	57
2.4.1	Determination of threshold by user input .....	57
2.4.2	A tablet-specific BW conversion .....	59
2.5	Preliminary logo searching.....	61
2.6	Summary .....	61
Chapter 3:	New methods for capturing tablet images for improved logo extraction	63
3.1	Method.....	65
3.2	Extracting the logo from multiple images.....	69
3.2.1	Logo extraction by edge detection .....	70
3.2.2	Logo extraction by converting to BW images .....	71
3.3	Adjusting the height of the ring light relative to the tablet.....	73
3.3.1	Logo Extraction by edge detection .....	74

3.3.2	Logo extraction by converting to BW images .....	76
3.3.3	The final logo image for searching .....	78
3.4	Logo searching .....	80
3.4.1	Complex moment invariants .....	81
3.4.2	Zernike Moment Invariants .....	87
3.4.3	Complex vs Zernike Moment Invariants .....	93
3.5	Summary .....	93
Chapter 4:	Searching illicit drug tablet images using colour .....	95
4.1	Method .....	97
4.2	Colour searching in the HSV colour space .....	98
4.2.1	New photos .....	99
4.2.2	Existing AFP photos .....	103
4.3	Colour searching in the CIE L*a*b* colour space .....	105
4.3.1	New photos .....	105
4.3.2	Existing AFP photos .....	110
4.4	A preliminary test for searching using both logo shapes and colours .....	113
4.5	Summary .....	119
Chapter 5:	Implementation into the AFP ENIPID database .....	120
5.1	Changes to the logo extraction script .....	122
5.2	Invariants for Symmetric Objects .....	124
5.3	Supplementary feature: side profiles .....	130
5.3.1	Method .....	131
5.3.2	Derivatives .....	132
5.3.3	Moment Invariants .....	136

5.4	Compiling MATLAB scripts to executable files.....	139
5.5	Summary .....	140
Chapter 6:	Conclusion .....	141
References.....		145
Appendix .....		154

## List of Figures

Figure 2.1: Some examples of the photos obtained. ....	49
Figure 2.2: The three tablet greyscale images used for illustration. ....	50
Figure 2.3 (a)-(h): "Canny" edge detection with different sigma values. Row 1 (a)-(d): Sigma = 1, 2, 4, 8; Row 2 (e)-(h): Sigma = 16, 32, 64, 128. Refer to the Appendix for a larger, clearer version. ....	51
Figure 2.4 (a)-(g): "Canny" edge detection with sigma = 32 at different thresholds. Row 1 (a)-(c): Threshold = 0.1563 (auto), 0.1, 0.12; Row 2 (d)-(g): 0.14, 0.16, 0.18, 0.2. Refer to the Appendix for a larger, clearer version. ....	52
Figure 2.5 (a)-(d): An example of edge images using different filters. (a): The "Laplacian of Gaussian" filter; (b): The "Sobel" filter; (c): The "Prewitt" filter; (d): The "Roberts" filter. Refer to the Appendix for a larger, clearer version. ....	53
Figure 2.6 (a)-(e): Each row lists the edges detected using (a) the "Canny" method, (b) the "Laplacian of Gaussian" method, (c) the "Sobel" method, (d) the "Prewitt" method and (e) the "Roberts" method from the same resized photo. ....	54
Figure 2.7: Scaled images showing the relative strength of phase congruencies as the noise threshold point increased from 2 to 10, 20 and 30. The true threshold for (a) would be below 10 while that for (b) would be above 20. ....	55
Figure 2.8 (a) – (e): Five final edges extracted from their respective original image using phase congruency. ....	56
Figure 2.9: Two examples of converting two tablet photos into BW images using different thresholds. ....	58
Figure 2.10: More testing on the optimal threshold. ....	59
Figure 2.11: A typical image histogram (intensities at uint 8) with two peaks. ....	60



Figure 2.12: Thresholds at 30%, 45%, 60%, 75% and 90% away from point B to point C. .....	60
Figure 2.13: Some examples of converting BW images using our own threshold method. .....	61
Figure 3.1: A genuine pill press. ....	65
Figure 3.2: A close-up of the pill press. ....	66
Figure 3.3: The dies used to press pills. ....	66
Figure 3.4: An example of the tablets pressed by the dies. A blank tablet pressed using a blank die is not shown here. Row 1: “Yin and Yang”, “8-ball”, “apple”, “three- pointed star 1” (“Mercedes”) and “three-pointed star 2” (“peace”); Row 2: “Euro”, “butterfly”, “kangaroo”, “e” and “score” . ....	67
Figure 3.5: The six colours used to dye the tablets. ....	67
Figure 3.6: Four images of the same tablet with four different light locations. ....	69
Figure 3.7: Left four images: Images generated by the four images as shown in Figure 3.6 using minimum, 2 <sup>nd</sup> minimum, median and average pixel greyscale values respectively; Right image: Same tablet under uniform lighting. ....	69
Figure 3.8: An example of edges detected using the five images in the order same as Figure 3.7. Row 1: the “Canny” method; Row 2: the “Laplacian of Gaussian” method; Row 3: the “Sobel” method; Row 4: the “Prewitt” method; Row 5: the “Roberts” method. ....	70
Figure 3.9: An example of converted BW images by our own tablet-specific BW conversion method using the five images in the same order as Figure 3.7. ....	72
Figure 3.10: The new proposed imaging setup. ....	74
Figure 3.11: Only the light rays hitting the surface area are reflected. ....	74

Figure 3.12: Examples of images taken with the ring light lowered towards the tablets. Row 1: "blank", "score", "Yin and Yang", "8-ball" and "three-pointed star 1"; Row 2: "three-pointed star 2", "Euro", "e", "apple" and "kangaroo". .....	74
Figure 3.13: An example of edge detection results using different filters .....	75
Figure 3.14: An example of false positives generated from a granule of dye .....	76
Figure 3.15: Some examples of BW images segmenting the logos out of the tablet surfaces.....	76
Figure 3.16: Another example of logo extraction by converting to BW.....	78
Figure 3.17: The flow to the final logo BW image from (a) to (d).....	80
Figure 3.18: A logo where filling the black, zero-valued pixels inside is undesirable (left); the result of such filling (right). .....	80
Figure 3.19: An ideal search result generated by comparing the complex moment invariants using Euclidean distance.....	81
Figure 3.20: Searching a "three-pointed star" logo using Euclidean distance.....	84
Figure 3.21: Logo searching using Euclidean distance.....	85
Figure 3.22: Same logo searching as in Figure 3.21 but with standardised Euclidean distance used instead. ....	85
Figure 3.23: An example of a search for "8-ball" logos by Euclidean distance using Zernike moment invariants. ....	88
Figure 3.24: Searching "e" logos using Zernike moment invariants and the Standardised Euclidean distance. ....	91
Figure 4.1: An example of a pair of images of the same tablet under different lighting conditions. The left-handed image is used for logo extraction, while the right hand image is used to extract other features. ....	96

Figure 4.2: Sample colours of the new tablets we pressed. ....	97
Figure 4.3: Sample photos chosen within the AFP database.....	98
Figure 4.4: An example of all nine positive hits ranking top of the list using HSV values. .....	99
Figure 4.5: Searching Tartrazine Yellow tablets using Euclidean distance.....	99
Figure 4.6: Searching Sunset Yellow tablets using Euclidean distance.....	99
Figure 4.7: Searching Tartrazine Yellow tablets using Standardised Euclidean distance. .....	100
Figure 4.8: Searching Sunset Yellow tablets using Standardised Euclidean distance. .	100
Figure 4.9: Searching Tartrazine Yellow tablets using City Block distance. ....	101
Figure 4.10: Searching Sunset Yellow tablets using City Block distance. ....	101
Figure 4.11: Searching Green tablets using Normalised Inner Product.....	101
Figure 4.12: Searching Tartrazine Yellow tablets using the algorithm from Jeong et al. (99).....	102
Figure 4.13: Searching Sunset Yellow tablets using the algorithm from Jeong et al. (99). .....	102
Figure 4.14: An example of colour searching in the HSV colour space using Euclidean distance.....	104
Figure 4.15: Another example of colour searching in HSV colour space using Euclidean distance.....	104
Figure 4.16: An example of colour searching in the HSV colour space using Normalised Inner Product. ....	104

Figure 4.17: Another example of colour searching in HSV colour space using Normalised Inner Product. ....	104
Figure 4.18: An example of colour searching in the HSV colour space using the algorithm from Jeong et al. (99). ....	105
Figure 4.19: Another example of colour searching in HSV colour space using the algorithm from Jeong et al. (99). ....	105
Figure 4.20: An example of all nine positive hits ranking at the top of the list using L*a*b* values. ....	106
Figure 4.21: Searching Tartrazine Yellow tablets using Euclidean distance. ....	106
Figure 4.22: Searching Sunset Yellow tablets using Euclidean distance. ....	106
Figure 4.23: Searching Tartrazine Yellow tablets using Euclidean distance in double precision. ....	107
Figure 4.24: Searching Sunset Yellow tablets using Euclidean distance in double precision. ....	107
Figure 4.25: Searching Tartrazine Yellow tablets using Standardised Euclidean distance in double precision. ....	107
Figure 4.26: Searching Sunset Yellow tablets using Standardised Euclidean distance in double precision. ....	107
Figure 4.27: Searching Tartrazine Yellow tablets using City Block distance in double precision. ....	108
Figure 4.28: Searching Sunset Yellow tablets using City Block distance in double precision. ....	108
Figure 4.29: Searching Tartrazine Yellow tablets using Normalised Inner Product in double precision. ....	108

Figure 4.30: Searching Sunset Yellow tablets using Normalised Inner Product in double precision. ....	108
Figure 4.31: An example of colour searching in the $L^*a^*b^*$ colour space using Euclidean distance.....	110
Figure 4.32: Another example of colour searching in the $L^*a^*b^*$ colour space using Euclidean distance. ....	110
Figure 4.33: An example of colour searching in the $L^*a^*b^*$ colour space using City Block distance.....	111
Figure 4.34: Another example of colour searching in the $L^*a^*b^*$ colour space using City Block distance. ....	111
Figure 4.35: An example of colour searching in the $L^*a^*b^*$ colour space using Normalised Inner Product. ....	112
Figure 4.36: Another example of colour searching in the $L^*a^*b^*$ colour space using Normalised Inner Product. ....	112
Figure 4.37: Third example of colour searching in the $L^*a^*b^*$ colour space using Normalised Inner Product. ....	112
Figure 4.38: Fourth example of colour searching in the $L^*a^*b^*$ colour space using Normalised Inner Product. ....	112
Figure 4.39: Searching using both logo and colour features. ....	115
Figure 4.40: Another search using both logo and colour features. ....	115
Figure 4.41: Searching using both logo (normalised between 0 and 1) and colour features. ....	116
Figure 4.42: Another search using both logo (normalised between 0 and 1) and colour features.....	116

Figure 4.43: Searching using both logo (normalised between 0 and 1) and colour features. .....	117
Figure 4.44: Same query search as Figure 4.43 using both logo and colour features normalised by their standard deviations.....	117
Figure 4.45: Combined logo and colour searching using intersection (only three out of the four tablets with matching colour and logo appeared). All four would appear if the ranking cut-off was set to 15.....	117
Figure 4.46: Another combined logo and colour search using intersection (all four tablets with matching colour and logo appeared). ....	118
Figure 4.47: Combined logo and colour searching using union. Hit No. 10 and 19 appeared in the top 12 from the logo search, after all of the nine matching logos that were present in the database. ....	118
Figure 4.48: Another combined logo and colour searching using union. No. 10 and 18 were included for the same reason as in Figure 4.47. ....	118
Figure 5.1: Logo searching using <i>bwareaopen</i> to extract logos and ranking using the Zermike moment invariants and Euclidean distance. ....	124
Figure 5.2: The “Apple” logo which could be extracted using <i>bwlabel</i> but not by <i>bwareaopen</i> .....	124
Figure 5.3: Both "three-pointed star" logos will appear identical when rotated $120^\circ$ or $2\pi/3$ ( $N = 3$ ).....	124
Figure 5.4: Searching one of the “three-pointed star” logos using complex moment invariants for symmetric objects and Euclidean distance. ....	126
Figure 5.5: Searching the other “three-pointed star” logo using complex moment invariants for symmetric objects and Euclidean distance. ....	126

Figure 5.6: Searching one of the “three-pointed star” logos using complex moment invariants for symmetric objects and standardised Euclidean distance.....	126
Figure 5.7: Searching the other “three-pointed star” logo using complex moment invariants for symmetric objects and standardised Euclidean distance.....	126
Figure 5.8: Searching one of the “three-pointed star” logos using complex moment invariants for symmetric objects (not from <i>rotmi</i> ) and Euclidean distance. ....	127
Figure 5.9: Searching the other “three-pointed star” logo using complex moment invariants for symmetric objects (not from <i>rotmi</i> ) and Euclidean distance. ....	127
Figure 5.10: Searching one of the “three-pointed star” logos using complex moment invariants for symmetric objects (not from <i>rotmi</i> ) and standardised Euclidean distance.....	128
Figure 5.11: Searching the other “three-pointed star” logo using complex moment invariants for symmetric objects (not from <i>rotmi</i> ) and standardised Euclidean distance.....	128
Figure 5.12: The fifth image for a tablet. Each tablet has two side profiles, illustrated by the red line and the blue line. ....	130
Figure 5.13: Some examples of the side profile images in our database (total 30). ....	130
Figure 5.14: Defining orientations for discussions in this section. ....	131
Figure 5.15: The binary image of Figure 5.12 using the threshold of 0.1. ....	132
Figure 5.16: Figure 5.15 was rotated such that the major axis was parallel to the <i>y</i> -axis. ....	132
Figure 5.17: The location of the 20 points where the slopes were calculated.....	133
Figure 5.18: The location of the 20 points in another tablet.....	133

Figure 5.19: Searching one of the two side profiles of the query image. All other seven tablets with the same profiles appeared in the top 11 hits. ....	134
Figure 5.20: Searching one of the two side profiles of another query image. All other three tablets with the same profiles appeared in the top 11 hits. ....	134
Figure 5.21: Searching the other side profile of the query image in Figure 5.19. Only six out of the seven tablets with the same profiles appeared in the top 11 hits.....	134
Figure 5.22: Searching the other side profile of the query image in Figure 5.20. All other three tablets with the same profiles appeared in the top 11 hits. ....	134
Figure 5.23: The same search as in Figure 5.21 but the number of intervals was set to 25 instead of 20. Only six out of the seven tablets with the same profiles appeared in the top 11 hits.....	134
Figure 5.24: The same search as in Figure 5.21 again but the number of intervals was set to 50 instead of 20. Only six out of the seven tablets with the same profiles appeared in the top 11 hits. ....	134
Figure 5.25: Asymmetric side profiles could be very different depending on how the tablet was placed under the camera.....	135
Figure 5.26: Asymmetric side profiles A and B match if only their derivatives are collected clockwise or anti-clockwise. ....	135
Figure 5.27: Symmetric side profiles C and D are vectors of opposite signs if the derivatives are collected just from the top to the bottom (or vice-versa). ....	136
Figure 5.28: Asymmetric side profiles E and F have the same derivatives but different signs. ....	136
Figure 5.29: Searching the side of the query tablet using Zernike moment invariants. None of the seven tablets with the same profiles appeared in the top 11 hits. ....	138



- Figure 5.30: Searching the side of another query tablet using Zernike moment invariants. Only two out of the three tablets with the same profiles appeared at 8<sup>th</sup> and 9<sup>th</sup>.  
..... 138
- Figure 5.31: Searching the side of the query tablet using complex moment invariants (CMI1). Five of the seven tablets with the same profiles appeared at 1<sup>st</sup>-4<sup>th</sup> and 11<sup>th</sup>.  
..... 138
- Figure 5.32: Searching the side of another query tablet using complex moment invariants (CMI1). All other three tablets with the same profiles appeared at 1<sup>st</sup>, 2<sup>nd</sup> and 9<sup>th</sup>.  
..... 138
- Figure 5.33: Searching the side of the query tablet using complex moment invariants (CMI2). Two of the seven tablets with the same profiles appeared at 7<sup>th</sup> and 8<sup>th</sup>.  
..... 138
- Figure 5.34: Searching the side of another query tablet using complex moment invariants (CMI2). Only one out of the three tablets with the same profiles appeared at 2<sup>nd</sup>.  
..... 138

## List of Tables

Table 3.1: A break-down of the number of tablet logo images in the database for logo matching.....	68
Table 3.2: Rankings of the matching logos within the first 11 using complex moment invariants by different algorithms.....	86
Table 3.3: Rankings of the positive hits within the first 11 using different Zernike moment invariants searched by Euclidean distance.....	89
Table 3.4: Rankings of the positive hits within the first 11 using Zernike moment invariants with different algorithms.....	92
Table 3.5: Rankings of logos searched by Euclidean distance using different moment invariants.....	93
Table 4.1: Summary of the rankings of the positive hits within the top 11 using HSV colour space with different search algorithms.....	103
Table 4.2: Summary of the rankings of the positive hits within the first 11 using $L^*a^*b^*$ colour space by different algorithms.....	109
Table 4.3: A break-down of the number of tablet logo images in the database for logo and colour matching.....	114
Table 5.1: Rankings by Euclidean distance of the positive hits within the first 11 using different morphological operations.....	123
Table 5.2: Rankings of the positive hits within the first 11 using moment invariants for recognising symmetric logos normalised by different methods.....	129

## Abstract

This thesis describes solutions for the automation of illicit tablet image searching based on tablet logos, colour and edge profiles. An original aim was to develop or exploit algorithms for searching images in an existing Australian Federal Police (AFP) drug tablet database. However, the pre-existing images were found to be inadequate for use with accepted logo extraction methods, due to poor illumination and insufficient contrast between the tablet surface and the logos. Extracting a complete logo from them was difficult. New “high contrast” images were taken in a way such that the contrast between the tablet surface and the logo was maximised. This was achieved by lowering the ring light that was attached to the macroscope very close to the tablet, resulting in a very shallow angle of illumination.

An in-house algorithm was designed to select the most appropriate threshold for each of these “high contrast” images to be converted to black-and-white (BW) images. Morphological and logical operations were used to segment the logos from the tablets. The final images contained only the silhouettes of the logos.

The logo silhouettes were encoded by calculating the Zernike moment invariants with the Kintner method, followed by the normalisation approach. The best chance of finding matching logos in the database was by ranking these invariants using the Euclidean distance.

Matching illicit tablets by colour was also investigated. The best method found involved changing the RGB colour values of the tablet photos to the CIE  $L^*a^*b^*$  coordinates and then comparing these coordinates using the Normalised Inner Product.

The side profiles of tablets were suggested as an extra feature for comparison. Although there were limitations that remained to be solved (namely the symmetry problems for the positioning of tablets with asymmetric side profiles), results were promising when comparing the derivatives of symmetric side profiles using the Euclidean distance.

All of the algorithms developed or used in this project have been compiled into executable files such that any tablet database computer can run the scripts for searching

logos and side profiles. In addition, the scripts can provide measurements of a tablet's diameter and thickness in millimetres.