

A COMPARATIVE STUDY OF ALGORITHMS FOR AUTOMATIC SEGMENTATION OF DERMOSCOPIIC IMAGES

Bachelor's Thesis



Cristina González Gonzalo

Bachelor's Degree in Audiovisual System Engineering
Universidad Carlos III de Madrid

Tutor: Miguel Ángel Fernández Torres

September 2016

Para mi abuela.

Acknowledgements

I would like to thank my parents, who have unconditionally supported me along these four years, from the very beginning until the completion of this project, during tough times with their wise advice and their encouraging words, as well as at moments of joy, with the most sincere of the smiles. Thank you for always encouraging me to keep pursuing my goals and teaching me that putting effort and passion into them leads to the best reward.

I wish to thank my sister because she is a constant source of inspiration and support. Thanks for guiding me towards and through this engineering world. She once mentioned that it is full of mountains to be climbed. Now I can say there is no better feeling than climbing them together.

I am also grateful to my family for their continuous caring, especially to my grandfather. He has never run out of words to express his support and confidence in me: "Try the best you can, because the best you can will always be good enough".

I am thankful to the person who has changed my world and lived by my side the last two years of this immense experience. We created the perfect team to overcome bad times and celebrate the good ones.

I would like to thank my friends inside and outside this University for those shared experiences that are already meant to never be forgotten.

Finally, I would like to express my gratitude to my tutor, Miguel Ángel, for his continuous monitoring throughout the different phases of the project, as well as to the Multimedia Processing Group of Universidad Carlos III de Madrid, for their welcoming reception and giving me the chance to develop this study within their walls.

Abstract

Melanoma is the most common as well as the most dangerous type of skin cancer. Nevertheless, it can be effectively treated if detected early. Dermoscopy is one of the major non-invasive imaging techniques for the diagnosis of skin lesions. The computer-aided diagnosis based on the processing of dermoscopic images aims to reduce the subjectivity and time-consuming analysis related to traditional diagnosis. The first step of automatic diagnosis is image segmentation.

In this project, the implementation and evaluation of several methods were proposed for the automatic segmentation of lesion regions in dermoscopic images, along with the corresponding implemented phases for image preprocessing and postprocessing. The developed algorithms include methods based on different state of the art techniques. The main groups of techniques which have been selected to be studied and implemented are thresholding-based methods, region-based methods, segmentation based on deformable models, as well as a new proposed approach based on the bag-of-words model. The implemented methods incorporate modifications for a better adaptation to features associated with dermoscopic images.

Each implemented method was applied to a database constituted by 724 dermoscopic images. The output of the automatic segmentation procedure for each image was compared with the corresponding manual segmentation in order to evaluate the performance. The comparison between algorithms was carried out regarding the obtained evaluation metrics.

The best results were achieved by the combination of region-based segmentation based on the multi-region adaptation of the k-means algorithm and the subsequent application of the Chan-Vese deformable model for the lesion contour refinement.

Index Terms — Dermoscopy, melanoma, automatic segmentation, skin, lesion, evaluation, comparison.

Index

Index of figures.....	viii
Index of diagrams	xi
Index of tables.....	xii
Index of examples	xiii
1. Introduction	1
1.1 Motivation and goals.....	1
1.2 Document structure	2
2. State of the art	4
2.1 Dermoscopy	4
2.2 Regulatory framework	4
2.3 Automatic segmentation techniques applied to dermoscopic images.....	5
2.3.1 Threshold-based methods.....	7
2.3.1.1 Global thresholding.....	7
2.3.1.2 Local thresholding	10
2.3.1.3 Multithresholding methods	10
2.3.2 Region-based methods.....	10
2.3.2.1 Clustering-based segmentation	11
2.3.3 Edge-based methods.....	12
2.3.4 Segmentation based on deformable models	14
2.3.4.1 Parametric deformable models	14
2.3.4.1.1 Traditional snakes	15
2.3.4.1.2 Gradient vector flow snakes	16
2.3.4.2 Geometric deformable models	19
2.3.4.2.1 Chan-Vese method.....	20
2.3.5 Segmentation based on the bag-of-words model.....	20
2.3.5.1 Detection of keypoints	21
2.3.5.2 Description of keypoints	22
2.3.5.2.1 SIFT descriptor.....	22
2.3.5.2.2 Colour-SIFT descriptors	23
2.3.5.3 Bag-of-words model.....	25
2.3.5.4 Adaptation to superpixel level	26
2.3.5.4.1 SLIC superpixels.....	26
2.3.5.4.2 Histogram intersection.....	27
2.4 Preprocessing of dermoscopic images.....	27
2.5 Postprocessing of dermoscopic images	28
3. System design.....	31
3.1 Problem definition	31
3.2 Limitations.....	31
3.3 Programming tools.....	33
3.4 Implementation.....	34
3.4.1 Preprocessing	34

3.4.1.1 Morphological preprocessing.....	34
3.4.1.2 Filtering.....	35
3.4.2 Automatic segmentation.....	36
3.4.2.1 Threshold-based methods.....	36
3.4.2.1.1 Traditional Otsu method.....	36
3.4.2.1.2 Local Otsu method.....	38
➤ Local Otsu on grid.....	38
➤ Local Otsu on concentric rectangular partitions.....	39
3.4.2.1.3 Multithresholding Otsu method.....	40
➤ Method to determine the number of regions within a dermoscopic image based on the analysis of local maxima in the luminance histogram.....	41
3.4.2.1.4 Alternatives of design.....	43
3.4.2.2 Region-based methods.....	43
3.4.2.2.1 Two-region k-means clustering.....	43
3.4.2.2.2 Multi-region k-means clustering.....	45
➤ Method to address illumination inconsistency in dermoscopic images.....	46
3.4.2.2.3 Alternatives of design.....	50
3.4.2.3 Segmentation based on deformable models.....	50
3.4.2.3.1 Gradient vector flow snakes.....	50
3.4.2.3.2 Chan-Vese method.....	54
3.4.2.3.3 Alternatives of design.....	55
3.4.2.4 Segmentation based on the bag-of-words model.....	57
3.4.2.4.1 Detection of keypoints.....	58
3.4.2.4.2 Description of keypoints.....	59
3.4.2.4.3 Bag-of-words model: generation of visual vocabulary.....	60
3.4.2.4.4 Matching of keypoints and visual words.....	61
3.4.2.4.5 Superpixel-level approach.....	62
3.4.2.4.6 Alternatives of design.....	64
3.4.3 Postprocessing.....	66
3.4.3.1 Morphological postprocessing.....	67
3.4.3.2 Expansion of lesion region.....	68
4. System evaluation.....	70
4.1 Database.....	70
4.2 Evaluation measures.....	70
4.3 Experiments and results.....	72
4.3.1 Threshold-based methods.....	72
4.3.1.1 Traditional Otsu method.....	72
4.3.1.2 Local Otsu method.....	75
4.3.1.3 Multithresholding Otsu method.....	76
4.3.2 Region-based methods.....	79
4.3.2.1 Two-region k-means clustering.....	79
4.3.2.1.1 Radial distance and <i>R, G, B</i> components.....	79
4.3.2.1.2 Radial distance and <i>L, a, b</i> components.....	80

4.3.2.1.3 Radial distance and S, V components	81
4.3.2.2 Multi-region k-means clustering	83
4.3.3 Segmentation based on deformable models	87
4.3.3.1 Gradient vector flow snakes.....	87
4.3.3.2 Chan-Vese method.....	89
4.3.4 Segmentation based on the bag-of-words model.....	92
5. Conclusions	101
6. Future work.....	104
7. Planning and methodology.....	105
8. Socio-economic background.....	107
8.1 Budget	107
8.1.1 Material costs	107
8.1.2 Personnel costs.....	107
8.1.3 Total project cost.....	108
8.2 Socio-economic environment	108
References	110
Text references.....	110
Figure references	114
Regulatory references	116
Appendix A: Further studies on state of the art	117
A.1 Colour spaces	117
A.2 Region-growing techniques	118
A.3 Hierarchical clustering	119
A.4 Edge detection operators	121
A.4.1 Gradient operators.....	121
A.4.2 Second derivative operators	121
A.4.3 Laplacian of Gaussian	122
A.4.4 Gaussian edge detection.....	122
A.5 Modifications of GVF algorithm.....	123
A.6 Active shape models	124
A.7 Modifications of SIFT descriptor	124
Appendix B: Cross-validation procedure	125
Appendix C: Summary.....	126
C.1 Introduction	126
C.2 Preprocessing and postprocessing of dermoscopic images	127
C.3 Automatic segmentation of dermoscopic images	128
C.3.1 Threshold-based methods.....	128
C.3.2 Region-based methods.....	130
C.3.3 Segmentation based on deformable models	133
C.3.4 Segmentation based on the bag-of-words model.....	134
C.4 Conclusions and future work	136

Index of figures

Figure 1: Dermatoscope [I].....	4
Figure 2: Image thresholded with Otsu method [II].....	9
Figure 3: One-dimensional edge profile [IV]	12
Figure 4: MR image of heart left ventricle and its potential energy function [V].....	15
Figure 5: Effect of α on the elasticity of the curve	15
Figure 6: Traditional snake [VI]	18
Figure 7: GVF snake [VI]	18
Figure 8: Example of embedding a curve as a level set. From left to right: original curve, level set function where the curve is embedded as the zero level set, height map of the level set function with its zero level set in black [V]	20
Figure 9: Overview of method based on extraction of local features and bag-of-words model [VII]	21
Figure 10: SIFT keypoint descriptor [VIII]	23
Figure 11: Procedure adapted to colour description [IX].....	24
Figure 12: Example of dilation [XIII]	29
Figure 13: Example of erosion [XIII]	29
Figure 14: Example of opening [XIII]	29
Figure 15: Example of closing [XIII]	30
Figure 16: Inconstant illumination [I]	32
Figure 17: Bubbles of fluid [I].....	32
Figure 18: Distortion [I].....	32
Figure 19: Thin hairs [XIV].....	32
Figure 20: Thick hairs [I].....	32
Figure 21: Variegated colouring [I]	32
Figure 22: Blood vessels [I].....	32
Figure 23: Irregular border [XIV]	32
Figure 24: Fuzzy border [I].....	32
Figure 25: Low contrast [XIV].....	32
Figure 26: Regression [I]	32
Figure 27: Multiple lesions [I].....	32

Figure 28: Traditional Otsu method. From left to right: original image, preprocessed luminance image, luminance histogram with threshold value situated on top (0.604).....	37
Figure 29: Traditional Otsu method. From left to right: original image, preprocessed luminance image, luminance histogram with threshold value situated on top (0.467).....	37
Figure 30: Pillbox and Gaussian filters [XVI].....	38
Figure 31: Local Otsu on grid. On top: original image (left), preprocessed luminance image (center), manual mask (right). On bottom: binary mask generated by local Otsu on grid (left), mask after Gaussian filtering (center), mask after pillbox filtering (right).	39
Figure 32: Local Otsu on concentric partitions. From left to right: binary masks of external (top), middle (center) and inner (bottom) partitions, areas of image under each partition, thresholded partitions.	39
Figure 33: Local Otsu on concentric partitions. On top: original image (left) and preprocessed luminance image (right). On bottom: binary mask before (left) and after smoothing (right)....	40
Figure 34: Multithresholding Otsu method. On top: original image (left), preprocessed luminance image (center), luminance histogram with number of required thresholds above: 3 (right). On bottom: manual mask (left), segmented regions (center), binary mask after merging of regions and complement computing (right).....	42
Figure 35: Multi-region k-means. On top: original image (left), preprocessed luminance image (center), luminance histogram with number of determined regions above, 3 (right). On bottom: manual mask (left), mask before region merging (center), mask after region merging and complement (right).	46
Figure 36: Multi-region k-means. From left to right: original image, manual mask, automatic mask before image inconsistency processing.....	47
Figure 37: Multi-region k-means. On top: original image (left), preprocessed luminance image (center), luminance histogram (number of determined regions: 2) (right). On bottom: manual mask (left), mask previous to focus processing (center), automatic mask after one focus processing (right).....	49
Figure 38: Multi-region k-means. On top: original image (left), preprocessed luminance image and manual mask (center), luminance histogram (number of determined regions: 3) (right). On bottom: mask before region merging and focus processing (left), mask after first (center) and second (right) focus processing.....	49
Figure 39: GVF snakes. From left to right: original image and edge map.....	51
Figure 40: GVF snakes. Normalized gradient vector field.	52
Figure 41: GVF snakes. On top: horizontal and vertical components of gradient. On bottom: horizontal and vertical components of gradient vector flow.	52
Figure 42: GVF initial curve. Original image without black frames and initial curve on blue.	53

Figure 43: Validation of internal forces parameters. Example of accuracy evolution regarding the modification of the elasticity and rigidity parameters.	54
Figure 44: Chan-Vese method. From left to right: original image, multi-region k-means segmentation, Chan-Vese input mask	55
Figure 45: GVF snakes. Colour components analysis to define edge map.	56
Figure 46: Extraction of keypoints. From left to right: manual mask with Harris Laplace detected keypoints, manual mask with dense sampling detected keypoints. Lesion keypoints are shown on red and detected skin keypoints on yellow.	59
Figure 47: Description of keypoints. Random selection of 50 keypoints detected by Harris Laplace detector and described by SIFT descriptor.	60
Figure 48: Matching of keypoints and visual words. On top: original image (left) and manual mask (right). On bottom: generated binary mask after matching (left) and mask after the pillbox filter is applied (right).	61
Figure 49: Generation of superpixels. On top: original reference image without black frames (left), generated superpixels (center), superpixel separation between lesion and skin (right). On bottom: manual mask (left), appearance of the mask generated from superpixels division (right).....	63
Figure 50: Detection of keypoints. From left to right: detected frames increasing from top to bottom the peak threshold (0, 10, 20, 30), detected frames increasing the edge threshold from top to bottom (3.5, 5, 7.5, 10). [XVII]	64
Figure 51: Variations in superpixel shape regarding the value of the regularizer parameter [XVIII].	66
Figure 52: Disk-shaped structural element [XIX].....	67
Figure 53: Morphological parameters validation. Example of accuracy evolution regarding the modification of the radius of opening and closing structural elements.	68
Figure 54: Expansion of lesion region.	69
Figure 55: Representation of RGB colour space [X]	117
Figure 56: Representation of HSV colour space [XI]	117
Figure 57: Representation of Lab colour space [XII]	118
Figure 58: Histogram of image and the obtained dendrogram [III]	119

Index of diagrams

Diagram 1: Overview of system design	34
Diagram 2: Preprocessing.	34
Diagram 3: Traditional Otsu method.	36
Diagram 4: Local Otsu method.....	38
Diagram 5: Multithresholding Otsu method. Determination of number of regions in Diagram 6.	40
Diagram 6: Method to determine number of regions within a dermoscopic image.....	41
Diagram 7: Two-region k-means clustering.	43
Diagram 8: Multi-region k-means clustering. Determination of number of regions in Diagram 6. Checking of illumination inconsistency in Diagram 9.....	45
Diagram 9: Method to address illumination inconsistency in dermoscopic images. Focus analysis in Diagram 10. Focus processing in Diagram 11.....	47
Diagram 10: Focus analysis.	48
Diagram 11: Focus processing.....	48
Diagram 12: Gradient vector flow snakes.....	51
Diagram 13: Chan-Vese method.	54
Diagram 14: Segmentation method based on the bag-of-words model.	58
Diagram 15: Superpixel-level approach.	62
Diagram 16: Postprocessing.....	66
Diagram 17: Methodology per implemented method.....	106

Index of tables

Table 1: Evaluation concepts [XV].....	70
Table 2: Evaluation of traditional Otsu method.....	75
Table 3: Evaluation of local Otsu method.....	76
Table 4: Evaluation of multithresholding Otsu method.....	78
Table 5: Average evaluation measures for each implemented thresholding-based method. ...	78
Table 6: Evaluation of two-region k-means clustering (radial distance and R, G, B components).	80
Table 7: Evaluation of two-region k-means clustering (radial distance and L, a, b components).	81
Table 8: Evaluation of two-region k-means clustering (radial distance and S, V components)..	83
Table 9: Evaluation of multi-region k-means clustering.	86
Table 10: Average evaluation measures for each implemented clustering-based method.	87
Table 11: Evaluation of GVF snakes	89
Table 12: Evaluation of Chan-Vese method on B component.	91
Table 13: Chan-Vese method on B component and reduced points on initial curve.	92
Table 14: Average evaluation measures for each implemented segmentation method based on deformable models.	92
Table 15: Evaluation of segmentation method based on the extraction of keypoints by dense sampling, the description of keypoints by SIFT and the matching of keypoints to a vocabulary generated by k-means algorithm with 500 iterations and constituted by 32 lesion words and 32 skin words.	99
Table 16: Average evaluation measures for each implemented method.....	101
Table 17: Material costs (hardware)	107
Table 18: Material costs (software)	107
Table 19: Personnel costs.....	108
Table 20: Total project cost.....	108

Index of examples

Example 1: Traditional Otsu method. Compact lesion. From left to right: original image, manual mask, mask before postprocessing, final mask.....	73
Example 2: Traditional Otsu method. Compact lesion with skin texture. From left to right: original image, manual mask, mask before postprocessing, final mask.	73
Example 3: Traditional Otsu method. Expanded lesion. From left to right: original image, manual mask, mask before postprocessing, final mask.....	73
Example 4: Traditional Otsu method. Colour-variegated lesion. From left to right: original image, manual mask, mask before postprocessing, final mask.	73
Example 5: Traditional Otsu method. Lesion with inner regression areas. From left to right: original image, manual mask, mask before postprocessing, final mask.	73
Example 6: Traditional Otsu method. Low-contrast between lesion and skin. From left to right: original image, manual mask, mask before postprocessing, final mask.	74
Example 7: Traditional Otsu method. Low contrast between lesion and skin. From left to right: original image, manual mask, mask before postprocessing, final mask.	74
Example 8: Traditional Otsu method. Presence of hairs. From left to right: original image, manual mask, mask before postprocessing, final mask.....	74
Example 9: Traditional Otsu method. Presence of traces of liquid. From left to right: original image, manual mask, mask before postprocessing, final mask.	74
Example 10: Traditional Otsu method. Illumination inconsistency. From left to right: original image, manual mask, mask before postprocessing, final mask.	75
Example 11: Local Otsu method. Unstable thresholds causes by skin patterns. From left to right: original image, manual mask, mask before postprocessing, final mask.	76
Example 12: Local Otsu method. Colour-variegated, expanded lesion. From left to right: original image, manual mask, mask before postprocessing, final mask.	76
Example 13: Multithresholding Otsu method. Expanded lesion. From left to right: original image, manual mask, segmented regions, final mask.	77
Example 14: Multithresholding Otsu method. Colour-variegated lesion. From left to right: original image, manual mask, segmented regions, final mask.	77
Example 15: Multithresholding Otsu method. Low contrast between lesion and skin. On top: original image (left), manual mask (center) luminance histogram (right). On bottom: segmented regions (left), final mask (right).....	77
Example 16: Multithresholding Otsu method. Presence of liquid traces. From left to right: original image, manual mask, segmented regions, final mask.	78

Example 17: Multithresholding Otsu method. Illumination inconsistency. From left to right: original image, manual mask, segmented regions, final mask.	78
Example 18: Two-region k-means clustering (radial distance, R, G, B). Low contrast between lesion and skin. From left to right: original image, manual mask, mask before postprocessing, final mask.	79
Example 19: Two-region k-means clustering (radial distance, R, G, B). Colour-variegated lesion. From left to right: original image, manual mask, mask before postprocessing, final mask.	79
Example 20: Two-region k-means clustering (radial distance, R, G, B). Illumination inconsistency. From left to right: original image, manual mask, mask before postprocessing, final mask.	80
Example 21: Two-region k-means clustering (radial distance, L, a, b). Low contrast between lesion and skin. From left to right: original image, manual mask, mask before postprocessing, final mask.	80
Example 22: Two-region k-means clustering (radial distance, L, a, b). Presence of hair. From left to right: original image, manual mask, mask before postprocessing, final mask.	81
Example 23: Two-region k-means clustering (radial distance, L, a, b). Colour-variegated lesion. From left to right: original image, manual mask, mask before postprocessing, final mask.	81
Example 24: Two-region k-means clustering (radial distance, S, V). Low contrast between lesion and skin. From left to right: original image, manual mask, mask before postprocessing, final mask.	82
Example 25: Two-region k-means clustering (radial distance, S, V). Presence of hair. From left to right: original image, manual mask, mask before postprocessing, final mask.	82
Example 26: Two--region k-means clustering (radial distance, S, V). Well-segmented, colour-variegated lesion. From left to right: original image, manual mask, mask before postprocessing, final mask.	82
Example 27: Two--region k-means clustering (radial distance, S, V). Poorly-segmented, colour-variegated lesion. From left to right: original image, manual mask, mask before postprocessing, final mask.	82
Example 28: Two--region k-means clustering (radial distance, S, V). Illumination inconsistency. From left to right: original image, manual mask, mask before postprocessing, final mask.	83
Example 29: Multi-region k-means clustering. Low contrast between lesion and skin. On top: original image (left), manual mask (center) luminance histogram (right). On bottom: segmented regions (left), final mask (right).	84
Example 30: Multi-region k-means clustering. Colour-variegated lesion. On top: original image (left), manual mask (center) luminance histogram (right). On bottom: segmented regions (left), final mask (right).	84

Example 31: Multi-region k-means clustering. Regression areas inside the lesion. On top: original image (left), manual mask (center) luminance histogram (right). On bottom: segmented regions (left), final mask (right).....	85
Example 32: Multi-region k-means clustering. Presence of blue-white veil and traces of liquid. On top: original image (left), manual mask (center) luminance histogram (right). On bottom: segmented regions (left), final mask (right).....	85
Example 33: Multi-region k-means clustering. Fragmented lesion. On top: original image (left), manual mask (center) luminance histogram (right). On bottom: segmented regions (left), final mask (right).	85
Example 34: Multi-region k-means clustering. Addressed illumination inconsistency problem. . On top: original image (left), manual mask (center) luminance histogram (right). On bottom: segmented regions (left), final mask (right).....	86
Example 35: GVF snakes. Final curve outside the lesion region. From left to right: original image, manual mask, final mask.	87
Example 36: GVF snakes. Final curve inside the lesion region. From left to right: original image, manual mask, final mask.	88
Example 37: GVF snakes. Comparative of initial (blue) and final (red) snake, after 100 iterations.	88
Example 38: GVF snakes. Centered lesion that has been under-segmented. From left to right: original image, manual mask, final mask.	88
Example 39: GVF snakes. Expanded lesion has been under-segmented. From left to right: original image, manual mask, final mask.	88
Example 40: Chan-Vese method. Adjustment of final curve to real boundaries. On top: original image (left), B component (center), manual mask (right). On bottom: initial mask (left), mask before postprocessing (center), final mask (left).....	90
Example 41: Chan-Vese method. Final contour not completely adjusted. On top: original image (left), B component (center), manual mask (right). On bottom: initial mask (left), mask before postprocessing (center), final mask (left).	90
Example 42: Chan-Vese method. Final contour well adjusted. On top: original image (left), B component (center), manual mask (right). On bottom: initial mask (left), mask before postprocessing (center), final mask (left).	91
Example 43: Harris Laplace detection at the presence of traces of liquid.....	93
Example 44: Segmentation based on visual vocabulary and histogram intersection at superpixel-level.	94
Example 45: Segmentation based on dense sampling, SIFT, 32/32 vocabulary. Compact lesion. On top: original image (left), manual mask (right). On bottom: mask after matching (left), mask after filtering (center), final mask (right).	95

Example 46: Segmentation based on dense sampling, SIFT, 32/32 vocabulary. Compact lesion that requires postprocessing to fill holes. On top: original image (left), manual mask (right). On bottom: mask after matching (left), mask after filtering (center), final mask (right). 95

Example 47: Segmentation based on dense sampling, SIFT, 32/32 vocabulary. Compact lesion that requires postprocessing to fill the hole inside the lesion region. On top: original image (left), manual mask (right). On bottom: mask after matching (left), mask after filtering (center), final mask (right). 95

Example 48: Segmentation based on dense sampling, SIFT, 32/32 vocabulary. Compact, colour-variegated lesion. On top: original image (left), manual mask (right). On bottom: mask after matching (left), mask after filtering (center), final mask (right). 96

Example 49: Segmentation based on dense sampling, SIFT, 32/32 vocabulary. Compact lesion with areas of low pigmentation. On top: original image (left), manual mask (right). On bottom: mask after matching (left), mask after filtering (center), final mask (right). 96

Example 50: Segmentation based on dense sampling, SIFT, 32/32 vocabulary. Expanded lesion. On top: original image (left), manual mask (right). On bottom: mask after matching (left), mask after filtering (center), final mask (right). 97

Example 51: Segmentation based on dense sampling, SIFT, 32/32 vocabulary. Low contrast between lesion and skin. On top: original image (left), manual mask (right). On bottom: mask after matching (left), mask after filtering (center), final mask (right). 97

Example 52: Segmentation based on dense sampling, SIFT, 32/32 vocabulary. Colour-variegated lesion. Improvement in connectivity when the number of iterations of k-means clustering is incremented. On top: original image (left), manual mask (center), mask after matching (right). On center (100 iterations): mask after filtering (left), final mask (right). On bottom (500 iterations): mask after filtering (left), final mask (right). 98

Example 53: Segmentation based on dense sampling, SIFT, 32/32 vocabulary. Presence of traces of liquid. Improvement in connectivity when the number of iterations of k-means clustering is incremented. On top: original image (left), manual mask (center), mask after matching (right). On center (100 iterations): mask after filtering (left), final mask (right). On bottom (500 iterations): mask after filtering (left), final mask (right). 99

1. Introduction

1.1 Motivation and goals

Melanoma is the most frequent and one of the most lethal types of skin cancer, but it is also among the most curable cancers when it is early identified. In order to improve its detection and increase survival rates for malignant melanoma, dermoscopy is integrated in the diagnosis procedure. It is a non-invasive imaging technique that allows an enhanced visualization of skin lesions, so that significant patterns that cannot be detected to the naked eye become recognisable.

Nevertheless, the traditional diagnosis procedure bases its conclusions exclusively on observation and certain visual rules regarding lesion features such as colour or asymmetry. Two main handicaps are associated with this procedure. Firstly, the diagnosis highly depends on the related complications and subjectivity of the exploration, as well as the expertise of the dermatologist. Therefore, different conclusions from the same lesion may be reached if different professionals carry out the study. Secondly, errors in diagnosis lead to the application of intrusive measures in order to surgically remove lesions wrongly classified as malignant.

The computer-aided diagnosis of skin lesions aims to support conventional diagnosis and reduce its drawbacks. Regarding the first handicap, the computerized analysis of dermoscopic images provides objective and globalised diagnosis, i.e., a uniform protocol to classify the lesion as benign or malign based on a system trained through reference databases, independent of the medical center or the health professional responsible for the diagnosis. As for the second handicap, an automatic procedure allows to accurately check the presence or absence of lesion characteristics and contribute the diagnosis with an additional conclusion, which may avoid certain invasive surgical interventions such as the extraction of tissue for biopsy. Therefore, it is a time-saving technique, since it allows physicians to focus on potentially dangerous lesions while discarding harmless cases. Additionally, since the diagnosis is based on the processing of dermoscopic images taken from the patient, it facilitates remote medical treatment, which is especially beneficial in rural areas or places that are difficult to access, by means of teledermatology techniques.

The first step of automatic diagnosis is image segmentation. It supposes a phase of paramount importance because the subsequent steps are based on the accuracy of the division between lesion and skin regions. The extraction of lesion features must be applied exclusively over the lesion region in order to obtain only truly relevant characteristics to represent significant patterns related to melanoma conditions. The more exact the segmentation of the lesion region, the less the confounding information for the posterior classification algorithm.

Diverse tools and segmentation techniques are being developed over the last years in order to help the prevention of malignant melanoma. The segmentation of skin lesions is a wide field with a scattered and growing number of proposed models. In our study, different methods

from this broad state of the art were selected in order to be analysed, implemented and compared, along with introduced modifications and also proposed innovative techniques.

This study aims to contribute to the research of optimal segmentation techniques for dermoscopic images. This area, within the automation of medical procedures, is currently under major development and progressively moving forward, since it may help cover three current healthcare necessities as previously mentioned: increase of efficiency, gratification for both involved patients and health professionals and telemedicine.

1.2 Document structure

The document is divided in different sections dedicated to specific aspects regarding our study:

- **Section 1:** Introduction. This section presents the motivation behind the development of the project and the goals within the area of analysis as well as the objectives of the study.
- **Section 2:** State of the art. Firstly, an overview of the dermoscopy technique and its regulatory framework is included. Then, the section focuses on the automatic segmentation of dermoscopic images, along with the required preprocessing and postprocessing stages. An exhaustive analysis of different segmentation techniques in the specialised literature is carried out regarding their category. Several methods were selected in order to be implemented, whose formulation, functioning and application to dermoscopic images are described.
- **Section 3:** System design. In the first place, the problem to be solved is outlined, including the associated limitations and the programming tools that have been applied to implement the proposed solutions. Once the problem is exposed, the proposed image preprocessing and postprocessing phases as well as the implementation of the selected segmentation methods are described. For each method the corresponding introduced modifications for a better adaptation to dermoscopic images are explained. Additionally, each category of segmentation techniques includes the considered alternatives of design.
- **Section 4:** System evaluation. In this section, the database used for the analysis of the implemented methods is described, as well as the evaluation metrics that were chosen to evaluate the performance of the resulting segmentations. Then, the evaluation of each implemented method is presented and compared with other methods inside or outside its segmentation category, along with diverse examples of its functioning and performance statistics.
- **Section 5:** Conclusions. In this section, the performance of the implemented methods and several aspects regarding the obtained outcomes are interpreted. After a global comparison, the best achieved results are determined.

- **Section 6:** Future work. The main segmentation factors to be improved and those lines of work that offer potential modifications are identified. Several proposed lines of investigation regarding improvement of the system are also included.
- **Section 7:** Planning and methodology. The main aspects related to the organization of the study development as well as the methodology of implementation are described.
- **Section 8:** Socio-economic background. Firstly, the budget of the project, regarding both material and personnel costs, is presented. Finally, the socio-economic environment around the study is described.

2. State of the art

2.1 Dermoscopy

Dermoscopy [1], also known as epiluminescence microscopy, is a non-invasive diagnostic technique based on optical magnification and either liquid immersion and low-angle-of-incidence lighting or cross-polarized lighting that allows an easy visualization of pigmented skin lesions. This diagnostic tool improves the recognition of morphological structures not visible by the naked eye or with conventional clinical images.

The technique consists in placing oil, alcohol or water on the skin lesion. This fluid eliminates surface reflection and allows a better visualization of pigmented structures within the epidermis, the dermoepidermal junction and the superficial dermis. A specific device is used for the inspection of the lesion, which is normally a dermatoscope or a digital imaging system. The magnification of these instruments range from 6x to 100x, although the widely used 10x magnification dermatoscope enables a sufficient assessment of pigmented skin lesions in daily routine [1].



Figure 1: Dermatoscope [1]

Previous studies demonstrated that dermoscopy improves accuracy in diagnosis. Dermoscopy allows the identification of numerous morphological features, such as atypical pigment networks, dots, globules, streaks, blue-white areas and blotches. This reduces screening errors and facilitates the differentiation between difficult lesions [2]. Dermoscopy is reported to have increased up to 10%-27% the sensitivity of clinical diagnosis accomplished by the naked eye [3], but it is important to take into account that formal training among dermatologists is essential for an efficient application of this technique.

2.2 Regulatory framework

In the context of dermoscopy, there are not legal requirements for the implementation of a computer-based diagnosis procedure. However, there exists a regulatory framework around the dermoscopic images obtained from patients by means of the dermoscopy technique. These are the images used by the dermatologists in order to study the primal appearance and the consequent evolution of the dermal lesions. Simultaneously, these images are employed in the implementation of automatic diagnosis systems along all the phases, including the segmentation step, studied in this project.

When a patient starts a dermatology treatment, pictures of his/her lesions might be required for the purpose of chronological control as a comparative measure, further study or even teledermatology service, an increasing technique for patients in rural areas. Therefore, the patient must be given a document by the dermatologist that contains all the relevant information related to the informed consent, i.e., the patient must allow the taking and storage of images and, at the same time, his/her privacy and the protection of personal data must be assured.

Regarding the taking of images, it is necessary to ensure the image rights of the patient, since an image may contain information about his/her identity and an unauthorised use may violate his/her right to privacy. The Spanish Constitution guarantees “the right to honour, to personal and family privacy and the own image” in Section 18.1 [A]. The Spanish Organic Law 1/1982 of civil protection, right to honour, to personal and family privacy and the own image, was approved according to the previous section in the Constitution [B].

The general data protection must be also secured, as the main personal data from the patient is contained within the documentation, such as name, ID number or place of residence. This aspect is regulated by the current Spanish Organic Law 15/1999 of protection of personal data [C].

The use of digital images has allowed dermatologists an easier storage and visualization of dermoscopic images. When a physician gains the patient’s informed consent, he/she is allowed to manipulate the obtained images, but has several obligations, such as encrypted storage or secured transfer of data. Additionally, if any image processing is applied to the original picture, the modifications must be stored without replacement of the initial version. Besides, dermatologists must clearly inform the patient about what the process of image manipulation involves, what the purposes are, where and for how long the images will be stored and who will have access to them [D].

2.3 Automatic segmentation techniques applied to dermoscopic images

Image segmentation [4] is the procedure of dividing an image into multiple segments, assigning a label to every pixel, so that pixels with the same label share certain visual features, such as colour, intensity or texture. The output is a set of regions that cover the entire image. Adjacent regions are fundamentally different with respect to the same characteristics. Essentially, the main goal of segmentation is to acquire a simplified representation of an image, easier to analyse.

The standard approach in dermoscopic image analysis has three steps: 1) image segmentation, 2) feature extraction and selection, 3) lesion classification [5]. The first step of the process, i.e. the segmentation of the lesion, consists in the classification of all points in the image as part of the lesion or simply part of the surrounding non-lesional skin [6]. It is crucial for image analysis due to two main requirements. First, the border structure serves as relevant information for diagnosis, since many clinical features, such as asymmetry, border irregularity or abrupt border

cutoff are studied directly from the boundaries. Second, the extraction of other important clinical features, such as atypical pigment networks, globules, blue-white areas or regression structures, critically depends on the accuracy of the lesion segmentation [2].

At the clinical diagnosis of melanoma, dermatologists usually search visually for localized patterns, colours, atypical structures, as well as analysing the shape and the border of the lesion. One example of medical procedure is the ABCD rule (Asymmetry, Border, Colour and Differential structures). An alternative approach is based on the 7-point checklist and weighs the presence or absence of seven dermoscopic criteria associated with melanomas: atypical pigment network and vascular pattern, irregular streaks, dots and pigmentation, regression structures and blue-whitish veil [7].

It may be observed that this procedure is based fundamentally on the visual recognition of skin patterns and the expertise of dermatologists and may lack of objective analysis and diagnosis in some cases. Providing this diagnosis with objective measures in order to support and facilitate the dermatologists' decisions is the main purpose of implementing an efficient automatic process for the analysis of dermoscopic images. Image segmentation, as the first step of the process, is expected to satisfy the mentioned requirements, so that the entire system can achieve an accurate diagnosis. The aim is an automatically applied segmentation phase, so that every lesion that enters the diagnosis procedure can be rapidly isolated from the non-lesional skin areas without human interaction. Therefore, time is saved for the study of the inner lesion and its boundaries are objectively defined.

Since the late 1990s numerous methods have been developed for automated segmentation of dermoscopic images. Within this project several algorithms among the main groups of techniques have been selected to be studied and applied. The following division for the selected segmentation techniques is proposed [2] [4]:

1. **Threshold-based segmentation.** These algorithms determine one or more proper threshold values to divide the pixels of an image into several classes and separate objects from the background.
2. **Region-based segmentation.** These methods segment an image into regions that are similar according to a set of predefined criteria.
3. **Edge-based segmentation.** These methods involve the detection of edges, important local changes in the intensity of an image, between the regions using edge operators.
4. **Segmentation based on deformable models.** These algorithms involve the detection of object contours using curve evolution techniques.
5. **Segmentation based on the bag-of-words model.** These experimental segmentation methods, already applied in the extraction of image features and image classification, are based on the detection and description of keypoints within the image and the posterior generation of a visual codebook.

Many of the methods included in this section, regardless of the segmentation techniques they are based on, take advantage of the properties that different colour spaces may offer. For instance, the implementation in one specific colour space can lead to an improvement in performance. By applying colour transformations in order to execute the same procedure in

different colour systems for a posterior comparison, the most suitable colour space for a particular segmentation algorithm can be found. The description of the **colour spaces** which have been used in our study can be found in Appendix A.1.

2.3.1 Threshold-based methods

The selection of thresholds is an important task in computer vision and detection systems. A threshold set too high may result in missed detections, while a threshold set too low may result in false positives. A fixed threshold may not perform appropriately if the properties among scenes change. Therefore, it is necessary to adapt a threshold dynamically at different scenarios in order to address these limitations [7].

Let N be the set of natural numbers, (x,y) be the spatial coordinate of a digitized image, and $G=\{0, 1, \dots, l-1\}$ be a set of positive integers representing gray levels. Then, an image function can be defined as the following mapping: $N \times N \rightarrow G$. The gray level of a pixel with coordinate (x,y) is denoted as $f(x,y)$. Let $t \in G$ be a threshold and $B = \{b_0, b_1\}$ be a pair of binary gray levels and $b_0, b_1 \in G$. The result of thresholding an image function at gray level t is a binary image function $f_t: N \times N \rightarrow B$, such that [9]:

$$f_t(x,y) = \begin{cases} b_0, & \text{if } f(x,y) < t \\ b_1, & \text{if } f(x,y) \geq t \end{cases}$$

Regarding dermoscopic image analysis, thresholding methods achieve good results when there is defined contrast between lesion and skin, thus the corresponding image histogram is bimodal, but do not usually work accurately when the modes of two regions overlap [5].

2.3.1.1 Global thresholding

A global thresholding technique is one that thresholds the entire image with a single threshold value [9]. Within this technique two thresholding methods can be differentiated: point-dependent techniques and region-dependent techniques.

A thresholding method is point-dependent if the threshold is determined from the gray level of each pixel. Otsu method [10], minimum error method [11], p-tile method [9] and entropic methods [12] are among the most used techniques within this group.

On the other hand, if the threshold is determined from the local property in the neighbourhood of each pixel, then the method is region-dependent. Several methods belong to this technique, such as histogram transformation methods, methods based on second-order gray level statistics or relaxation methods [9].

This project has focused thresholding-based segmentation on point-dependent techniques, specifically the **Otsu method** [10], which is one of the most used threshold-based methods in

computer vision. It is based on unsupervised automatic threshold selection and maximizes the separability of the resultant classes in gray levels.

In ideal cases the histogram of the image has a deep and sharp valley between two of the peaks representing objects and background, so that the threshold can be set at the bottom of the valley. However, for most real pictures it is difficult to detect valley bottom precisely, especially in cases when the valley is flat and broad, there is prominent noise in the image, or the two peaks are extremely unequal in height.

Let $L=\{1, 2, \dots, L\}$ be the number of gray levels in an image and $N=n_1 + n_2 + n_L$ the total number of pixels. The gray-level histogram is normalized and regarded as a probability distribution:

$$p_i = n_i/N, \quad p_i \geq 0, \quad \sum_{i=1}^L p_i = 1$$

The pixels of the image are partitioned into two classes, one for the background (C_0) and one for the objects (C_1), or vice versa, by a threshold k : C_0 denotes pixels with levels $[1, \dots, k]$ and C_1 denotes pixels with levels $[k+1, \dots, L]$. The probabilities of class occurrence are given by:

$$\omega_0 = \Pr(C_0) = \sum_{i=1}^k p_i = \omega(k)$$

$$\omega_1 = \Pr(C_1) = \sum_{i=k+1}^L p_i = 1 - \omega(k)$$

And the class mean levels are given by:

$$\mu_0 = \sum_{i=1}^k i \Pr(i|C_0) = \sum_{i=1}^k i p_i / \omega_0 = \mu(k) / \omega(k)$$

$$\mu_1 = \sum_{i=k+1}^L i \Pr(i|C_1) = \sum_{i=k+1}^L i p_i / \omega_1 = \frac{\mu_T - \mu(k)}{1 - \omega(k)}$$

The zeroth- and first-order cumulative moments of the histogram up to the k th level are given, respectively, by:

$$\omega(k) = \sum_{i=1}^k p_i \quad \mu(k) = \sum_{i=1}^k i p_i$$

While the total mean level of the original picture is:

$$\mu_T = \mu(L) = \sum_{i=1}^L i p_i$$

The following relation can then verified for any value of k :

$$\omega_0 \mu_0 + \omega_1 \mu_1 = \mu_T, \quad \omega_0 + \omega_1 = 1$$

The class variances are given by:

$$\sigma_0^2 = \sum_{i=1}^k (i - \mu_0)^2 \Pr(i|C_0) = \sum_{i=1}^k (i - \mu_0)^2 p_i / \omega_0$$

$$\sigma_1^2 = \sum_{i=k+1}^L (i - \mu_1)^2 \Pr(i|C_1) = \sum_{i=k+1}^L (i - \mu_1)^2 p_i / \omega_1$$

Let σ_W^2 be the within-class or intra-class variance (the variance inside the same class), σ_B^2 the between-class variance or inter-class variance (the variance between the different classes) and σ_T^2 the total variance of levels. The three variances are respectively given by:

$$\sigma_W^2 = \omega_0 \sigma_0^2 + \omega_1 \sigma_1^2$$

$$\sigma_B^2 = \omega_0 (\mu_0 - \mu_T)^2 + \omega_1 (\mu_1 - \mu_T)^2 = \omega_0 \omega_1 (\mu_1 - \mu_0)^2$$

$$\sigma_T^2 = \sum_{i=1}^L (i - \mu_T)^2 p_i$$

Then the following discriminant criterion measures can be determined in order to evaluate the quality of the threshold at level k , i.e., the threshold that gives the best separation of classes in gray levels:

$$\lambda = \sigma_B^2 / \sigma_W^2, \quad \kappa = \sigma_T^2 / \sigma_W^2, \quad \eta = \sigma_B^2 / \sigma_T^2$$

Since η is the simplest one of the three criterion functions, it is selected as the criterion measure to evaluate the threshold. The optimal threshold k^* is the one that maximizes η , or equivalently maximizes σ_B^2 .

$$\sigma_B^2(k^*) = \max_{1 \leq k < L} \sigma_B^2(k)$$

The maximum value, η^* , can be then used as a measure to evaluate the separability of classes or the bimodality of the histogram. This is a significant measure, for it is invariant under affine transformations of the gray-level scale. It is determined within the range:

$$0 \leq \eta^* \leq 1$$

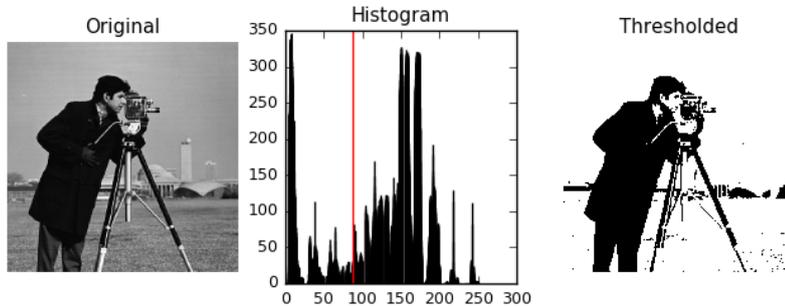


Figure 2: Image thresholded with Otsu method [11]

2.3.1.2 Local thresholding

In local thresholding, the image is partitioned into smaller subimages and a threshold is computed for each subimage. The result is a thresholded image with gray level discontinuities at the boundaries of each two different subimages. A smoothing algorithm must be applied in order to reduce the discontinuities. A point-dependent method or a region-dependent method may be applied to determine the threshold of each subimage [9]. Otsu method is one of the models that can be adapted into local thresholding by dividing the original image in squares belonging to a grid that covers the entire image or concentric radial rings, and then applying the algorithm on each of those partitions.

2.3.1.3 Multithresholding methods

Many global thresholding methods can be modified in order to compute not only one threshold for the entire image, but several thresholds, so that the segmentation of different regions is more complete and exact. This is a useful tool at dermoscopic image segmentation, since most lesions contain several colour-variegated areas and a clear individual threshold at the image histogram cannot be observed. The differentiation and segmentation of each of these areas by multiple thresholds can facilitate an exact isolation of the lesion from the non-lesional skin.

Otsu method extension to multithresholding [10] is straightforward due to the discriminant criterion η . For example, in the case of three-thresholding classes segmentation, two thresholds are assumed between the minimum gray level and the maximum one (L): $1 \leq k_1 < k_2 < L$. They separate the three classes: C_0 for $[1, \dots, k_1]$, C_1 for $[k_1 + 1, \dots, k_2]$, and C_2 for $[k_2 + 1, \dots, L]$. The criterion measure η , equivalent to σ_B^2 , is then a function of two variables (k_1 and k_2) and two optimal thresholds (k_1^* and k_2^*) are determined by maximizing σ_B^2 :

$$\sigma_B^2(k_1^*, k_2^*) = \max_{1 \leq k_1 < k_2 < L} \sigma_B^2(k_1, k_2)$$

The selected thresholds generally become less conclusive when the number of classes o be separated increases. This is because the criterion measure, σ_B^2 , has been defined in one-dimensional scale and may lose its meaning as the number of classes is incremented. It should also be noticed that the expression of σ_B^2 and the maximization procedure become more complex.

2.3.2 Region-based methods

Region-based methods involve the distribution of pixels into several homogenous regions according to common image properties [13], which consist of:

- Intensity values from the original image or computed values based on image operators.
- Textures or patterns that are exclusive to each type of region.
- Spectral profiles that provide multidimensional image data.

Region-based methods can facilitate the segmentation of those dermoscopic images in which lesions contain several areas with different colour and texture, since every region can be segmented separately and merged together afterwards. Nevertheless, these approaches can have difficulties in these cases too, that is, when the lesion as well as the skin region are too textured or have colour-variegated areas, leading to an oversegmentation problem, where an exceeding number of subregions are detected and the merging processing becomes more complicated.

There exist two fundamental types of region-based segmentation techniques: clustering-based segmentation and region-growing techniques. The region-based algorithms implemented in this project are based on clustering-based segmentation. Therefore, a deeper analysis of these methods is included in the next section. Nevertheless, in Appendix A.2 an overview of region-growing techniques can be found.

2.3.2.1 Clustering-based segmentation

A cluster is a collection of objects which are similar among them and are dissimilar to the objects that belong to other clusters. Therefore, the goal of clustering-based methods is to determine the intrinsic grouping in a set of unlabeled data [14]. Clustering problems are present in many different applications, such as vector quantization and pattern recognition. The approach of what constitutes a good cluster depends on the application and there are many methods for finding clusters according to various criteria [15].

Clustering techniques can be divided into two categories: partitional clustering and hierarchical clustering. Partitional clustering segmentation approaches have been the ones selected for implementation in this project and they are described below. However, an overview of the application of hierarchical clustering is included in Appendix A.3.

Partitional clustering decomposes a data set into a set of disjoint clusters. Given a data set of N points, a partitioning method creates K ($N \geq K$) partitions of the data, with each partition representing a cluster, by satisfying the following requirements [16]:

- 1) Each group contains at least one point.
- 2) Each point belongs to exactly one group, although in fuzzy partitioning methods, which are not covered in this project, a point can belong to more than one group.

Many partitional clustering formulations attempt to minimize an objective function. Among these formulations the most widely used is **k-means clustering**, which is covered in our study. Given a set on n data points in real d -dimensional space, \mathbb{R}^d , and an integer c , the aim of the algorithm is to determine a set of c points in \mathbb{R}^d , so that the distance from each data point to its nearest center is minimized [15], i.e., a squared error function is minimized:

$$J(V) = \sum_{i=1}^c \sum_{j=1}^{c_i} (\|x_i - v_j\|)^2$$

where $\|x_i - v_j\|$ is the Euclidean distance between a data point and the cluster center, c_i is the number of data points in the i^{th} cluster and c is the number of cluster centers [13].

The minimization of the objective function is carried out by an iterative procedure that comprises the following steps [13]:

1. A number of c points are placed into the space represented by the objects that are being clustered. These points are the initial group of centroids.
2. Each object is assigned to the group with the closest centroid.
3. When all objects have been assigned a centroid, the positions of the c centroids are recalculated.
4. The steps 2 and 3 are repeated until the centroids do not change their position, i.e., until a convergence condition is reached. The algorithm will eventually converge to a point that is a local minimum, but the result does not have to be necessarily a global minimum [15].

K-means clustering process may be hindered by image noise and outliers with the data, but it is relatively robust, fast and efficient. It requires a priori specification of the number of clusters, that is, it is necessary to introduce the number of final regions into which the objects will be grouped. This can be tricky when it comes to dermoscopic image segmentation, as every lesion can have its own pattern for the number of differential lesional areas. The algorithm is also significantly sensitive to a random initialization of the centroids. In dermoscopic images it is important to determine certain fixed initialization for the centroids, so that the process of separating lesion and skin areas can be facilitated from the beginning and the clusters are created taking into account these initial features.

2.3.3 Edge-based methods

Edge-based segmentation is based on edge detection algorithms, whose aim is to identify those points in the image where there are discontinuities in brightness in order to find the boundaries of objects within the image.

Discontinuities in image intensity can be either *step* discontinuities, where the image intensity abruptly changes from one value on one side of the discontinuity to a different value on the opposite side, or *line* discontinuities, where the image intensity abruptly changes its value but returns to the starting value within short distance. However, these types of sharp borders are rare in real images because of low-frequency components and the smoothing of sensing devices: *step* edges become *ramp* edges and *line* edges become *roof* edges [21].

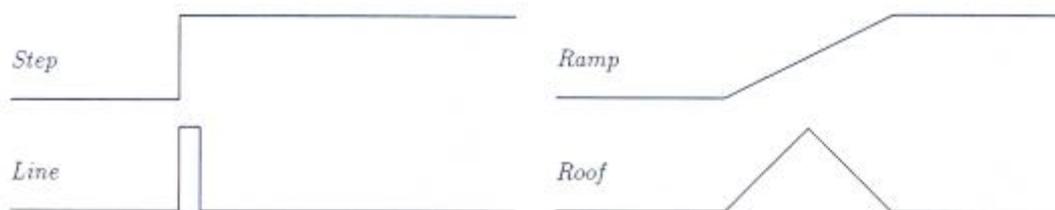


Figure 3: One-dimensional edge profile [IV]

It is difficult to develop an edge detection operator that reliably finds edges and is immune to noise. Several steps are necessary in order to achieve an accurate detection [21]:

1. Filtering: since some edge detection algorithms are susceptible to noise and other handicaps in discrete computations, a filter is commonly applied to improve their performance. However, there must be a trade-off between edge strength and noise reduction. If the level of accuracy is too high, noise will cause the detection of numerous false edges. On the other hand, if too much filtering is applied to the reduction of noise, the accuracy might be reduced because many relevant edges might not be detected [19].
2. Enhancement: the detection of edges is facilitated by determining changes of intensity in the neighbourhood of a point. Enhancement emphasizes pixels where there is a significant local change of intensity and is usually performed by computing the gradient magnitude.
3. Detection: many points in an image satisfy the criteria to be considered as part of an edge, but only those points with strong edge content are desired. Some method should be used to determine which points are edge points. Frequently, thresholding provides the criterion used for detection.

Many edge detection algorithms include a fourth step:

4. Localization: the location of the edge and its orientation can be estimated with subpixel resolution if required.

Some of the most used edge detection operators are classified and described in Appendix A.4. The description of each detector is based on the exhaustive analysis included in [21].

Dermoscopic image segmentation presents a set of difficulties to edge-based methods due to their criteria of detection. Edge-based approaches perform poorly when the boundaries of the lesion are not well defined and the transition between skin and lesion is smooth. In these situations the resultant edges contain gaps and the contour of the lesion may leak through them. Another difficulty for these methods is the presence in the image of some elements such as hair, specular reflections and irregularities in the skin texture. In these cases it is common the appearance of spurious edge points that do not belong to the lesion boundary [5].

Although some edge-based techniques present advanced implementations that aim to resolve some handicaps of these algorithms, such as the sensitivity to noise and the selection of a unique, fixed threshold, the edge detectors are still affected by the properties of dermoscopic images. For this reason, they are usually combined with other more specific and developed methods, which exploit the advantages of edge-based algorithms and include them as a phase of the entire procedure. It is the case of the traditional active contours model, which is studied in the next section.

2.3.4 Segmentation based on deformable models

Deformable models, also referred to as snakes or active contours, are curves or surfaces defined within an image domain that can move under the influence of internal forces, which are defined within the curve or surface itself, and external forces, which are computed from the image data. The internal forces keep the model smooth during the deformation process, whereas the external forces move the model toward an object boundary or other desired features within the image. By constraining extracted boundaries to be smooth and incorporating a priori information about the shape of the object to be segmented, deformable models allow integrating boundary elements into a consistent mathematical description [22].

Active contour algorithms are often very sensitive to initialization. The initial snake points can be manually defined by an operator or automatically determined. In order to make the segmentation process fully automated, an effective automatic snake initialization method is required [23].

There are two types of deformable models: parametric deformable models and geometric deformable models.

2.3.4.1 Parametric deformable models

Parametric deformable models represent curves explicitly in their parametric forms during deformation. This representation allows direct interaction with the model and can lead to a compact representation for fast real-time implementation [22]. Parametric active contours synthesize parametric curves within an image domain and allow them to move toward desired features, usually edges [24].

There are two different types of formulations for parametric deformable models: energy minimization formulation and dynamic force formulation. The two formulations lead to similar results, but the first one has the advantage that its solution satisfies a minimum principle, whereas the second one has the flexibility of allowing the use of more general types of external forces [22]. The parametric deformable models studied in this project are based on the energy minimization formulation.

The premise of the energy minimization formulation is to find a parameterized curve that minimizes the weighted sum of internal energy and potential energy. The internal energy specifies the tension or the smoothness of the contour. The potential energy is defined over the image domain and typically has local minima at image intensity edges occurring at object boundaries, which can be observed in the following figure [22].

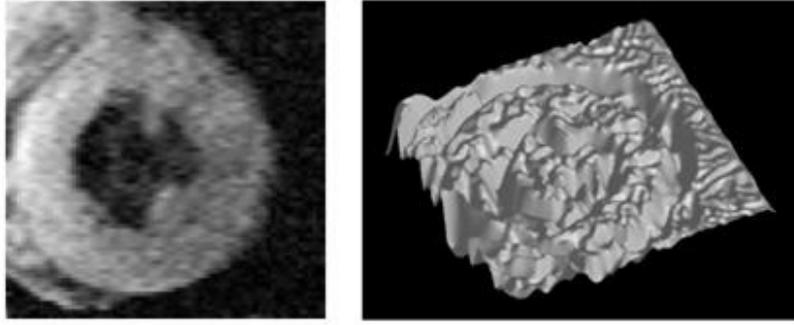


Figure 4: MR image of heart left ventricle and its potential energy function [V]

Typically, the curves are drawn toward the edges by potential forces. Additional forces, such as pressure forces or gradient vector flow, which is studied later, comprise the external forces. There are also internal forces designed to hold the curve together (elasticity forces) and to keep it from bending too much (bending forces) [24]. In order to find the object boundary, parametric curves are initialized within the image domain and are forced to move toward the potential energy minima under the influence of both of these forces [22].

2.3.4.1.1 Traditional snakes

As described in [24], a traditional snake is a curve $\mathbf{x}(s) = [x(s), y(s)]$, $s \in [0, 1]$, that moves through the spatial domain of an image to minimize the following energy functional:

$$E = \int_0^1 \frac{1}{2} [\alpha |x'(s)|^2 + \beta |x''(s)|^2] + E_{ext}(x(s)) ds$$

where $\mathbf{x}'(s)$ and $\mathbf{x}''(s)$ denote the first and second derivatives of $\mathbf{x}(s)$ with respect to s . The parameters α and β are weighting parameters that control the tension and the rigidity of the snake, respectively: α controls the penalty for internal elasticity and bigger values mean less stretching; bigger values of β mean the curve is harder to bend (if $\beta=0$ the snake becomes second order discontinuous and develops a corner).

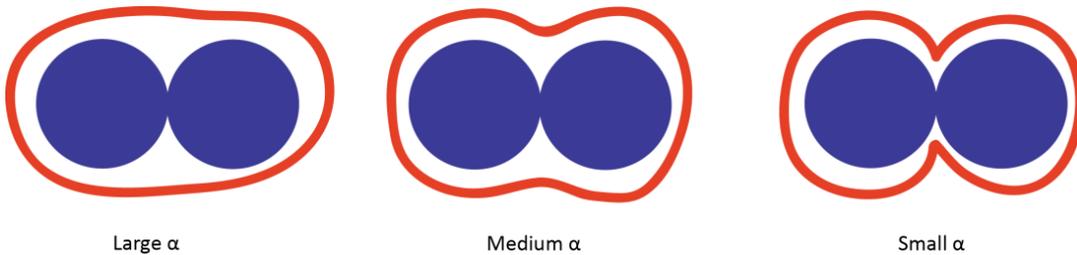


Figure 5: Effect of α on the elasticity of the curve

The external energy function E_{ext} is derived from the image so that it reaches its smaller values at the features of interest, such as boundaries. Given a gray-level image $I(x,y)$, typical external energies designed to lead an active contour toward *step* edges are:

$$E_{ext}^{(1)}(x, y) = -|\nabla I(x, y)|^2$$

$$E_{ext}^{(2)}(x, y) = -|\nabla [G_\sigma(x, y) * I(x, y)]|^2$$

where $G_\sigma(x,y)$ is a two-dimensional Gaussian function with standard deviation σ and ∇ is the gradient operator. Larger values of σ will cause the boundaries to become blurry, but they are sometimes necessary in order to increase the capture range of the active contour.

A snake that minimizes E must satisfy the following Euler equation:

$$\alpha x''(s) - \beta x''''(s) - \nabla E_{ext} = 0$$

This can be interpreted as a force balance equation:

$$\mathbf{F}_{int} + \mathbf{F}_{ext}^{(p)} = 0$$

$$\mathbf{F}_{int} = \alpha x''(s) - \beta x''''(s)$$

$$\mathbf{F}_{ext}^{(p)} = -\nabla E_{ext}$$

The internal force \mathbf{F}_{int} controls the stretching and bending and the external potential force $\mathbf{F}_{ext}^{(p)}$ pulls the snake toward the desired image edges.

To find a solution to the Euler equation, the snake is made dynamic by treating \mathbf{x} as function of time t as well as s and computing iteratively until it stabilizes and a numerical solution is achieved.

$$x_t(s, t) = \alpha x''(s, t) - \beta x''''(s, t) - \nabla E_{ext}$$

A key problem with traditional snake formulation is its limited capture range. The magnitude of the external forces decreases rapidly away from the object boundary. As a consequence, the initialization of the contour must be close to the original boundary or else it likely converges to the wrong result. Additionally, prior knowledge of whether to shrink or expand the snake is needed. When σ is increased the range is augmented, but the boundary localization becomes less accurate and distinct. Traditional snakes also show difficulties at adapting to boundary concavities.

2.3.4.1.2 Gradient vector flow snakes

The analysis included in [24] describes the gradient vector flow (GVF) as a class of external forces. These forces are dense vector fields derived from images by minimizing a certain energy functional in a variational framework. The minimization is achieved by solving a pair of decoupled linear partial differential equations that diffuses the gradient of an edge map in regions distant from the boundary. The amount of diffusion adapts according to the strength of the edges to avoid the distortion of object boundaries [22].

GVF snakes have advantages over traditional snakes. They are less sensitive to initialization and have the ability to move into boundary concavities. Besides, a GVF snake does not require prior knowledge whether to shrink or expand toward the boundary. These snakes also have a larger capture range, so the curve can be initialised farther away from the boundary.

The previously computed force balance condition is used as starting point:

$$\mathbf{F}_{int} + \mathbf{F}_{ext}^{(p)} = 0$$

The first step is the definition of an edge map $f(x, y)$, derived from the image $I(x, y)$, having the property of being larger near the image edges. Any gray-level or binary edge map can be used, for example:

$$f(x, y) = -E_{ext}^{(i)}(x, y)$$

The edge map satisfies three general properties. First, the gradient of an edge map (∇f) has vectors pointing toward the edges, which are normal to the edges at the edges points. Second, these vectors have normally large magnitudes only in the vicinity of the edges. Third, in homogeneous regions, where $I(x, y)$ is nearly constant, ∇f is nearly null.

These properties have consequences on the behaviour of a traditional snake. The first one is a highly desirable property, since a snake initialized close to the edge will converge to a stable configuration near the edge. But the second and the third properties are undesirable. Because of the second property, the capture range will be very small, and homogeneous regions will have no external forces due to the third property. The GVF snakes aim to keep the desirable property of the gradients near the edges and at the same time extend the gradient map farther away from the edges and into homogeneous regions using a computational diffusion process. This way, the snake will be able to move toward object boundaries and boundary concavities even from homogeneous regions.

A new static external force, the gradient vector flow (GVF) is then defined: $\mathbf{F}_{ext}^{(g)} = \mathbf{v}(x, y) = [u(x, y), v(x, y)]$. It minimizes the following energy functional:

$$\varepsilon = \int \int \mu(u_x^2 + u_y^2 + v_x^2 + v_y^2) + |\nabla f|^2 |v - \nabla f|^2 dx dy$$

When ∇f is small, the energy is dominated by the sum of the squares of the partial derivatives of the vector field, yielding a slow varying field, i.e., the result is smooth in the absence of data. When ∇f is large, the second term dominates the integrand and is minimized by setting $\mathbf{v} = \nabla f$. Therefore, \mathbf{v} is kept nearly equal to the gradient of the edge map when it is large, but the field is forced to be slowly-varying in homogeneous regions.

The parameter μ regularizes the trade-off between the first and the second term of the integrand. It should be set according to the amount of noise present in the image: it should be increased when there is a high amount of noise.

To obtain the corresponding dynamic snake equation, the potential force term from traditional snakes ($-\nabla E_{ext}$) is replaced by the gradient vector field, $\mathbf{v}(x, y)$:

$$x_t(s, t) = \alpha x''(s, t) - \beta x''''(s, t) + v$$

It is solved numerically by discretization and iteration, in identical manner as the traditional snake.

A comparison between traditional potential forces and GVF external forces, as well as between the convergence of a traditional snake and the convergence of a GVF snake, is shown in the next figures.

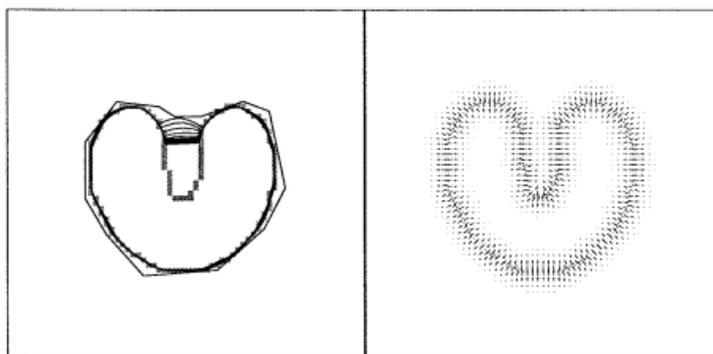


Figure 6: Traditional snake [VI]

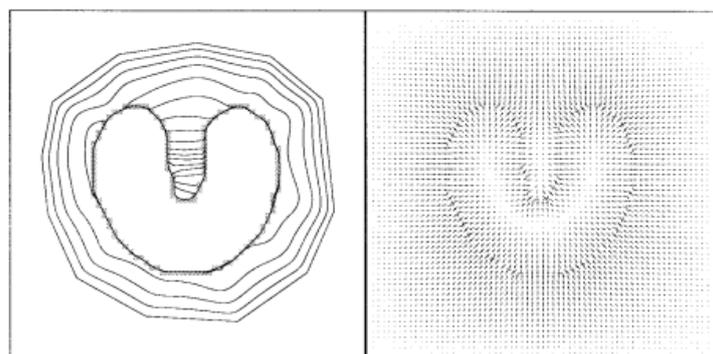


Figure 7: GVF snake [VI]

GVF snakes generate fairly good results even when dealing with image noise and concave edges. However, during the diffusion of the external force field, the conflict among forces happens. Thus, in object boundaries with deep concavities the GVF snakes still find difficulties [25]. GVF active contours still show substantial noise sensitivity and certain inability to capture and preserve weak edges. They have also a high computational cost [26]. In the related literature, several methods aim to improve the results of this already advanced algorithm. The description of some them can be found in Appendix A.5.

Regarding the efficiency of parametric deformable models, it may be observed that they depend on numerous parameters, such as μ , α and β . It is necessary to validate those parameters and choose their optimal values for a particular image, as well as understand the interaction among them. Even more parameters can be found in implementations of the algorithm. It is the case of the viscosity of the curve (γ), which controls the velocity and the maximum distance that the curve can be deformed in each iteration: the bigger its value, the harder and slower the deformation. Another extra parameter is the weight of the external forces with respect the internal forces (κ): the bigger its value, the stronger the force toward the image edges. This aspect makes the implementation and adaptation of these models complex, since dermoscopic images have their own intricate properties regarding image texture and lesion localization, among others.

2.3.4.2 Geometric deformable models

As proposed in [22], geometric deformable models are based on the theory of curve evolution and the level set method. Curves are evolved only using geometric measures, resulting in an evolution that is independent of parameterization, so they can be represented implicitly as a level set of a higher-dimensional scalar function. The parameterization is computed only after the deformation is completed. As in parametric deformable models, the image data is employed for evolution of the curves, in order to recover object boundaries.

The purpose of curve evolution theory is to study the deformation of curves using only geometric measures such as the unit normal and curvatures as opposed to the quantities that depend on parameters such as derivative of an arbitrary parameterized curve. Let $\mathbf{X}(s, t) = [X(s, t), Y(s, t)]$ be a moving curve, where s is any parameterization and t is the time. N denotes its inward unit normal and κ denotes its curvature. The evolution of the curve along its normal direction is given by the following partial differential equation:

$$\frac{\partial \mathbf{X}}{\partial t} = V(\kappa)N$$

where $V(\kappa)$ is called speed function, since it determines the speed of the curve evolution.

There are two main curve deformations: curvature deformation and constant deformation. Curvature deformation smoothes the curve and has a similar effect to the use of the elastic internal force in parametric deformable models. On the other hand, constant deformation determines the pressure force and direction of deformation, and plays the same role as the pressure force in parametric deformable models.

The basic idea of geometric deformable models is to associate the speed of deformation, using curvature and/or constant deformation, with the image data, so that the curve evolves until it stops at object boundaries. This evolution process is implemented using the level set method.

In the level set method a curve is represented implicitly as a level set of a 2D scalar function, called level set function, which is defined on the same domain as the image. The level set is defined as the set of points that have the same function value. The purpose of the level set function is to provide an implicit representation of the evolving curve. Instead of tracking a curve through time, the level set method evolves a curve by updating the level set function at fixed coordinates through time. An initial function must be constructed such that its zero level set corresponds to the position of the initial contour. Given a level set function $\phi(x, y, t)$ with the contour $\mathbf{X}(s, t)$ as its zero level set, then:

$$\phi[\mathbf{X}(s, t), t] = 0$$

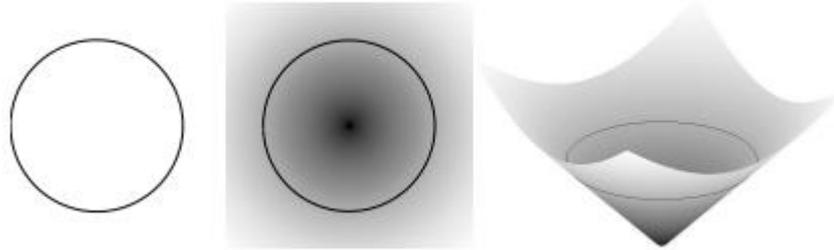


Figure 8: Example of embedding a curve as a level set. From left to right: original curve, level set function where the curve is embedded as the zero level set, height map of the level set function with its zero level set in black [V]

2.3.4.2.1 Chan-Vese method

The Chan-Vese method proposed in [28], also referred to as active contours model without edges, is based on the level set formulation and offers some advantages over the parametric deformable methods. The model is supposed to detect boundaries not necessarily defined by gradient. In the level set formulation, the problem becomes a “mean-curvature flow”-like evolving the active contour, which is supposed to stop on the desired boundary. However, the stopping criterion does not depend on the gradient of the image, but is instead related to a particular segmentation of the image.

The classical snakes rely on the edge function to stop the curve evolution, so the curve may pass through the boundary. If the image is noisy, then the smoothing filtering has to be strong, but it smoothes the edges too. In the Chan-Vese method the stopping term is based on Mumford-Shah segmentation techniques: it detects objects with very smooth boundaries or even with discontinuous boundaries. In addition, the model detects automatically interior contours and the initial curve can be placed anywhere in the image. The contour is also unbiased, i.e., free to either shrink or expand based on the image features [29].

There exist certain segmentation methods related to deformable models that propose an alternative approach and are used in medical applications; therefore, they may also be adapted to the processing of dermoscopic images. One example is active shape models, which are described in Appendix A.6.

2.3.5 Segmentation based on the bag-of-words model

The extraction of local features and bag-of-words model are two widespread methods in computer vision. Together combined they constitute a powerful tool to learn and represent image contents for a posterior procedure of pattern recognition or image classification. Due to their proper functioning in these two fields, it was decided to implement an experimental segmentation method for this project. It is considered to be experimental due to the lack of literature specified on the application of the extraction of local descriptors and bag-words

model to image segmentation. But, since they are promising techniques to be adapted into automatic segmentation processes, they were selected to be tested for dermoscopic image segmentation.

The aim of the method is, as for the rest of segmentation techniques adapted to dermoscopic images, to separate the lesion areas from the non-lesional skin. This method is based on the generation of visual codebooks from a reference set of dermoscopic images. These visual vocabularies are created by applying clustering techniques to a collection of previously extracted keypoints, from both the lesion and skin regions.

The implementation process consists in a set of steps, all of them based on the algorithms and models that are described in this section. An overview of the whole process is showed in the next figure.

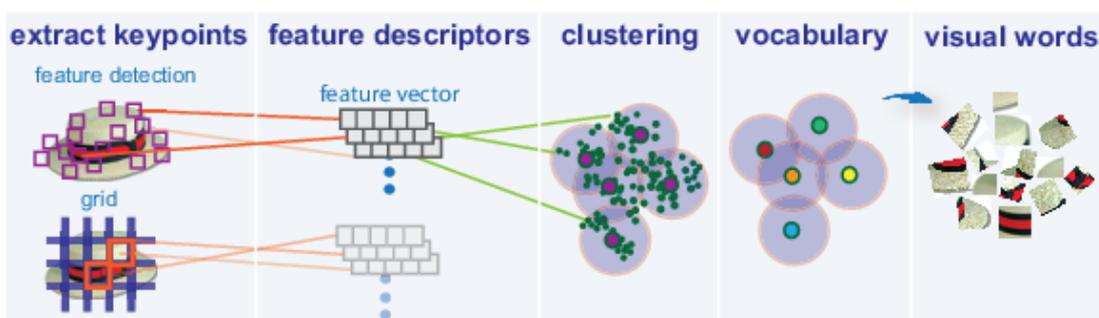


Figure 9: Overview of method based on extraction of local features and bag-of-words model [VII]

2.3.5.1 Detection of keypoints

Local features within an image refer to a pattern or distinct structure, such as a point, edge or small image patch. Examples of local features are blobs, corners and edge pixels. They are usually associated with areas in the image that differ from its immediate surrounding by texture, colour or intensity. It is not important what the local feature represents, but what it makes it distinct from its surroundings. Therefore, feature detection consists in selecting regions of an image that have unique content, points of interest, i.e., keypoints, that can be used for further processing [31].

A keypoint detector relies on gradient-based intensity variation approaches in order to detect good local features. The aim of a detector is to extract distinctive invariant features that can be used to perform reliable matching between different views of an object or scene. Consequently, an optimal keypoint must be robust to changes in viewing conditions such as noise, scale, rotation or illumination, so that it allows repeatable detections. A keypoint has also a unique location assigned to it, so that changes in viewing conditions do not affect its location. Additionally, the neighbourhood around the feature center must vary enough to accomplish a reliable comparison between features [31] [32].

There are two major stages for keypoints detection. In order to achieve invariance, it is necessary to start by constructing a scale space representation for the image, which is done by convolving the image with a set of Gaussian kernels $G_{\sigma_D}(x,y)$, each one with a different scale σ_D . Then, the keypoints, as well as their characteristic scale are computed using the information of

the scale-space [33]. Once a keypoint candidate has been found by comparing a pixel to its neighbours, the next step is to perform a detailed fit to the nearby data for location, scale and ratio principal curvatures. This information allows points to be rejected due to their poor localization along the edge or because their low contrast, which makes them too sensitive to noise. The third step is orientation assignment. By assigning a consistent orientation to each keypoint based on local image properties, the keypoint can be represented relative to this orientation and be invariant to image rotation. An orientation histogram is formed from the gradient operators of sample points within a region around the keypoint [32].

Literature offers a variety of keypoints detectors. An interesting comparison can be found in [34]. The following covariant feature detectors are among the most applied ones [35]:

- Difference of Gaussian (DoG) detector uses the local extrema trace of the multiscale Laplacian operator to detect features in scale and space.
- Hessian detector uses the local extrema of the multiscale determinant of the Hessian operator.
- Hessian Laplace detector does the same as the previous operator for localization in space, but uses the multiscale Laplacian for localization in scale.
- Harris Laplace detector uses the multiscale Harris cornerness measure instead of the determinant of Hessian for localization in space, and is otherwise identical to the previous operator.
- Hessian multiscale detector detects features spatially at multiple scales by using the multiscale of Hessian operator, but does not attempt to estimate their scale.
- Harris multiscale detector is like the previous one, but uses the multiscale Harris measure instead.

An alternative to detectors consists in assuming that keypoints are equally spaced and are the nodes of a regular grid placed on the image [33]. This way a keypoint is extracted every certain number of pixels. An advantage of this technique is that a constant sampling of features in the image is assured, without relying on the performance of a detector. The main disadvantage is that there is no discarded keypoints due to their irrelevance, so not invariant keypoints can also be taken into account.

2.3.5.2 Description of keypoints

The previous detection operations have assigned an image location, scale and orientation to each keypoint and provide invariance. The next step is to compute a descriptor for the local image region that is highly distinctive and invariant to possible variations, such as changes in illuminations and viewpoint.

2.3.5.2.1 SIFT descriptor

The SIFT (Scale-invariant feature transform) descriptor proposed in [32] is computed in the luminance space of the image and describes the shape of the region around the keypoint by using histograms of the gradient.

First, the image gradient magnitudes and orientations are sampled around the keypoint location, using the scale of the keypoint to select the level of Gaussian blur for the image, i.e., they are weighted by a Gaussian window, which is represented by a circle on the center diagram of the following figure. The purpose of this Gaussian window is to avoid sudden changes in the descriptor with small changes in the position of the window and to give less emphasis to gradients that are far from the center of the descriptor.

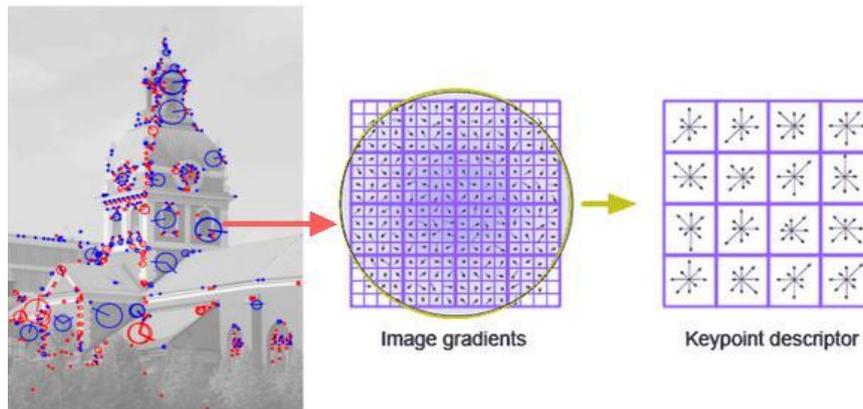


Figure 10: SIFT keypoint descriptor [VIII]

The keypoint descriptor allows for significant shift in gradient positions, by accumulating the samples into orientation histograms summarizing the contents over 4x4 sample regions. Each orientation histogram shows eight directions, with the length of each arrow corresponding to the magnitude of that histogram entry. This can be observed on the right diagram of the previous figure.

Therefore, a 4x4 array of orientation histograms with 8 orientations bins in each results in a 128 (4x4x8) element feature vector for each keypoint descriptor.

SIFT keypoint descriptors are highly distinctive and allow a single feature to find its correct match with good probability in a large database of features. However, in an overfull image, many features from the background will not have any correct match in the database, giving rise to many false matches in addition to the correct ones [32]. That could be problematic for dermoscopic image segmentation, since one single image can contain a great variety of textures and illumination changes.

In Appendix A.7 an additional modification of SIFT descriptor regarding medical image analysis can be found.

2.3.5.2.2 Colour-SIFT descriptors

The standard SIFT descriptor is computed in the luminance image. This can be a drawback, since it ignores the colour features of the image, which can provide discriminative information [33]. Dermoscopic images contain valuable information in colour space that helps identify the type of lesion and its diagnosis. Also the differences in colour between lesion and skin areas are notable in most images. Therefore, the application of colour descriptors is a potential tool to obtain a reliable characterization of lesion and skin regions.

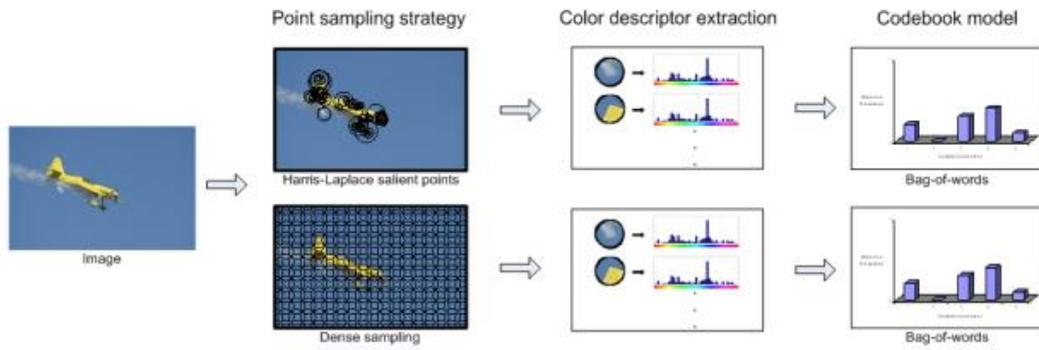


Figure 11: Procedure adapted to colour description [IX]

Three types of colour descriptors are differentiated due to their nature and invariant properties [37]:

- Colour histograms, which do not contain local spatial information and are inherently pixel-based.
- Colour moments, which contain local photometrical and spatial information derived from pixel values.
- Colour-SIFT descriptors, which contain local spatial information and are derivative-based.

Colour-SIFT descriptors are the ones selected in this project to be adapted into melanoma segmentation. Their purpose is to achieve the invariance that the standard SIFT descriptor lacks. The SIFT descriptor is not invariant to light colour changes, since the intensity channel is a combination of the R , G and B channels. The following colour-SIFT descriptors suppose a colour extension of the previously studied SIFT descriptor [37]:

- **Hue-SIFT** descriptor consists in a concatenation of the hue histogram descriptor with the SIFT descriptor. In the HSV colour space, the hue becomes unstable near the gray axis. The usage of a weighed hue histogram addresses this instability. Because the bins of the hue histogram are independent, the periodicity of the hue channel is achieved. The Hue-SIFT descriptor is scale-invariant and shift-invariant.
- **rgSIFT**. Descriptors are added for the r and g chromaticity components of the normalized RGB colour model, which is scale-invariant. In the normalized RGB colour model, the chromaticity components r and g describe the colour information in the image (b is redundant):

$$\begin{pmatrix} r \\ g \\ b \end{pmatrix} = \begin{pmatrix} \frac{R}{R+G+B} \\ \frac{G}{R+G+B} \\ \frac{B}{R+G+B} \end{pmatrix}$$

- **OpponentSIFT**, which is tested in our study, describes all the channels in the opponent colour space by using SIFT descriptors:

$$\begin{pmatrix} O1 \\ O2 \\ O3 \end{pmatrix} = \begin{pmatrix} \frac{R-G}{\sqrt{2}} \\ \frac{R+G-2B}{\sqrt{6}} \\ \frac{R+G+B}{\sqrt{3}} \end{pmatrix}$$

The channel O_3 represents the intensity information, while the other two channels describe the colour information in the image. These other two channels contain also some intensity information, but due to the normalization of the SIFT descriptor, they are invariant to changes in light intensity.

2.3.5.3 Bag-of-words model

Once the keypoints are described, the bag-of-words model can be applied, in order to generate a visual codebook based on the features extracted from the reference set of images. When the model is adapted to dermoscopic segmentation, the codebook contains words that represent the lesion areas and words that represent the skin areas.

The codebook approach allows then classification by describing an image as a bag of features, where image features, typically SIFT, are represented by discrete visual prototypes, which are previously defined in the given vocabulary [38]. In the case of dermoscopic images, the aim is to differentiate the lesion areas from the skin areas, according to the similarities of the features from the entering image with the words in the vocabulary.

A codebook can be obtained by following two different approaches: an annotation approach or a data-driven approach. It is the latter the one studied for this project. A data-driven approach applies an aggressive vector quantization on the space of the extracted features by using k-means clustering [38].

As previously studied, the classic k-means clustering algorithm finds cluster centroids that minimize the distance between data points and the nearest centroid. Thus, k-means can be viewed as a way of constructing a dictionary $D \in R^{n \times k}$ of k vectors so that a data vector $x^{(i)} \in R^n$, $i=1, \dots, m$, can be mapped to a code vector that minimizes the error in reconstruction [39]. Therefore, each centroid is called a visual word and the set can be seen as a visual dictionary. This visual dictionary is used then to analyse each entering image, compare its patch features with the centroids and assign them to the closest one, i.e., the centroid that minimizes the Euclidean distance [33].

This hard assignment of words in the vocabulary to image feature vectors is one drawback of the codebook approach. The hard assignment results in two issues: codeword uncertainty and codeword plausibility. Codeword uncertainty refers to the problem of selecting the correct codeword out of two or more relevant candidates. The codebook approach merely selects the best representing codeword, ignoring the relevance of other candidates. Codebook plausibility denotes the problem of selecting a codeword without a real suitable candidate in the vocabulary. The codebook approach assigns the best fitting codeword, regardless the fact that this codeword is not a proper representative [38].

The hard assignment of codewords to image features ignores codeword uncertainty and may label image features by non-representative codewords as a consequence. This problem can be addressed by weighing several close codewords and not only taking into account the nearest one. In the literature several methods that propose a softer assignment can be found. One example is the kernel density estimation proposed in [38], which shows that allowing a degree of ambiguity in the assignment of codewords the categorization performance is improved.

Another example is the Best-Bin-First (BBF) algorithm proposed in [40], which only considers matches in which the nearest neighbour is less than 0.8 times the distance to the second-nearest neighbour and therefore there is no need to exactly solve the most difficult cases in which many neighbours are at very similar distances.

2.3.5.4 Adaptation to superpixel level

An approach of the described segmentation method based on the extraction of local features and the bag-of-words model adapted to superpixel level is also proposed in this project. This section includes an introduction to superpixels technique and histogram intersection as assignment method.

2.3.5.4.1 SLIC superpixels

Superpixels group pixels into perceptually meaningful atomic regions, which can replace the rigid structure of the pixel grid. They capture image redundancy, provide a convenient approximation to compute image features and reduce the complexity of processing tasks. There are different methods to generate superpixels, but the following properties are desirable in all of them: superpixels should adhere well to image boundaries, they should be fast to compute and memory efficient and, when applied to segmentation procedures, they should both increase speed and quality of results [41].

The simple linear iterative clustering (SLIC) proposed in [42] and described in [43] is one of the most popular algorithms. SLIC is a simple and efficient method to decompose an image in visually homogeneous regions. It is based on a spatially localized version of k-means clustering. Each pixel is associated to a feature vector:

$$\Psi(x, y) = \begin{bmatrix} \lambda x \\ \lambda y \\ I(x, y) \end{bmatrix}$$

Then k-means algorithm is run of these. The coefficient λ balances the spatial and appearance components of the feature vectors, imposing a degree of spatial regularization to the extracted regions. This coefficient is regulated by two parameters: the nominal size of the regions (region size) and the strength of the spatial regularization (regularizer), which sets the trade-off between clustering appearance and spatial regularization.

$$\lambda = \frac{\text{regularizer}}{\text{region size}}$$

The image is first divided into a grid, whose tiles have a size equivalent to the previous parameter region size. $M \times N$ is the dimension of each tile in the grid. The center of each grid tile is then used to initialize k-means algorithm. In order to avoid placing the centers on image discontinuities, they are moved in a 3x3 neighbourhood to minimize the edge strength. Then the regions are obtained by running k-means clustering from the obtained centers:

$$C = \{\Psi(x_i, y_j), i = 0, 1, \dots, M - 1 \quad j = 0, 1, \dots, N - 1\}$$

K-means uses the standard Lloyd algorithm alternating the assignment of pixels to their closest centers and the re-estimation of the centers as the average of the corresponding feature vectors of the pixels assigned to them. The only difference when compared to the standard k-means algorithm is that each pixel can only be assigned to the center of one of the 2x2 neighbouring tiles adjacent to the pixel. This guarantees a simplification of processing cost during each k-means iteration.

After k-means has converged, SLIC eliminates any connected region whose area is less than an established minimum number of pixels. This is done by merging regions to neighbour ones.

2.3.5.4.2 Histogram intersection

In the adaptation to superpixel level of the segmentation method based on local features and bag-of-words-model, there is also a reference set of images. Superpixels are extracted from each image of the set, as well as for every entering image to be segmented. For every superpixel, a histogram is generated. The histogram counts the frequency of occurrence of each visual word contained in the codebook previously generated. This histogram is then considered as the new feature vector that characterizes one superpixel.

The histogram intersection method is applied to measure the similarity between histograms, in order to assign to each superpixel of the image to be segmented one superpixel from the reference set. The following operation is performed between one entering superpixel and each superpixel of the reference set, in order to find the most similar one:

$$d(\mathbf{x}, \mathbf{y}) = \sum_i \min(x_i, y_i)$$

where \mathbf{x} and \mathbf{y} are histograms and x_i and y_i are the i -th bin of each of both histograms [33].

2.4 Preprocessing of dermoscopic images

The preprocessing step before the application of segmentation methods is an important aspect for an effective recognition and examination of pigmented skin lesions. Dermoscopic images often contain external artefacts, such as black frames, ink markings, rulers or traces of fluid used for dermoscopic observation, as well as intrinsic cutaneous features, such as skin lines, blood vessels, hair or changes in skin texture. Additionally, images also contain features related to variations in illumination, such as reflections or shadows. All these elements complicate the border detection procedure and may result in a loss of accuracy and increase in computational time [2].

The most straightforward way to remove these artefacts or reduce their effect on lesion segmentation is to smooth the image using a general-purpose filter such as mean, median or Gaussian filter [2].

The application of a Gaussian smoothing operation is a common option for the preprocessing of dermoscopic images, since it allows a balance between blurring of undesirable artefacts and conservation of edges. An important aspect that should be taken into account is the scalar space. The amount of blurring depends on the value of σ , the standard deviation used in the filter implementation. A larger σ results in better noise filtering but at the same time important edge information is lost. If a small filter is used, there is likely to be more noise due to the insufficient averaging. In general, small filters result in too many noise points and large filters result in dislocation of edges and even false edges. The exact size of the filter cannot be determined without knowing the size of the image and the location of objects in the image [21].

An alternative strategy for artefact removal is to use a specialized method for each artefact type. For instance, some specific methods for fluid bubbles removal or hair removal are proposed in the related literature. They are usually based on colour space transformation, illumination correction, contrast enhancement or morphological operations [2].

2.5 Postprocessing of dermoscopic images

In order to obtain the final lesion border, the segmentation output must be postprocessed. According to the selected segmentation method, a precise sequence of postprocessing operations should be applied. Nevertheless, certain operations seem to be generally helpful, regardless of the implemented segmentation method [2].

- **Region merging.** The segmentation procedure ideally produces two regions: the lesion and the background skin. However, since these regions are rarely homogeneous, segmentation methods usually divide them in multiple detected subregions. In order to obtain a single lesion object, subregions that belong to the lesion should be first identified and then merged. The background skin region can be estimated from the colour and location of the corners of the image, so that the remaining subregions can be merged into one lesion region.
- **Islands and holes removal.** Islands (small isolated regions tagged as lesion in the label image) and holes (small regions inside the main lesion region that have been tagged as skin) can be eliminated using morphological operators, which are implemented using a structural element, as described in [46].
A structural element can have any shape and size. Its center is placed on every pixel of the processed image and the morphological operation is applied over all the points situated under the structural element.
The dilation process fills spaces where the structural element does not fit, such as small holes and “bays”. If the structural element intersects with an object when placed on a pixel, the value of that pixel turns positive, that is, it is added to the lesion mask.

$$\delta_C(A) = A \oplus C = \{x | (\hat{C})_x \cap A \neq \emptyset\}$$

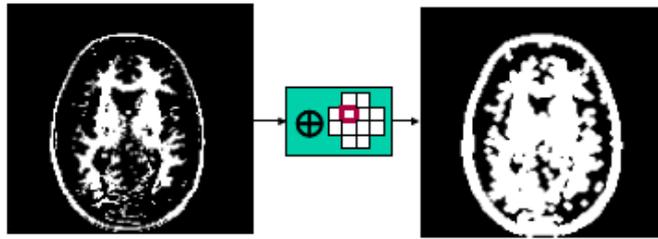


Figure 12: Example of dilation [XIII]

The erosion process eliminates resulting small elements from the segmentation, such as islands and protuberances, which are considered irrelevant for the final lesion region. If the structural element does not fit in the object when placed on a pixel, the value of that pixel turns negative, that is, it is eliminated from the lesion region.

$$\varepsilon_C(A) = A \ominus C = \{X | C_x \subset A\}$$

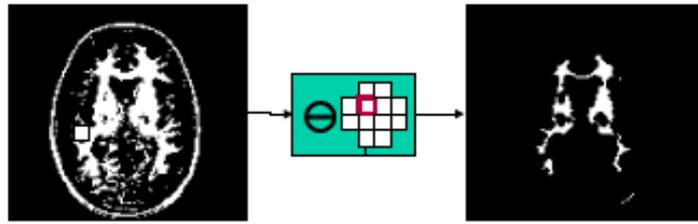


Figure 13: Example of erosion [XIII]

As it may be observed, dilation expands the components of an image and erosion shrinks them. There are two other important morphological operations based on the previous ones: opening and closing. These operations combined allow recovering the original size of the processed mask but the eliminated parts cannot be restored.

Opening is the combination of applying an erosion operation followed by a dilation operation, both with the same structural element. This process generally smooths the contour of an object, breaks narrow connectors and eliminates thin protuberances.

$$\gamma_B(A) = A \circ B = (A \ominus B) \oplus B$$

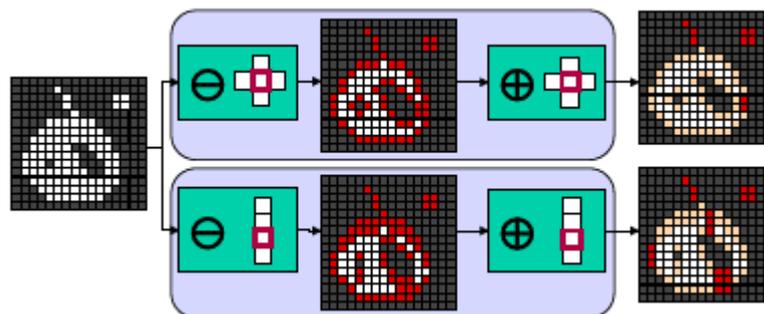


Figure 14: Example of opening [XIII]

Closing is the combination of applying first a dilation operator and then an erosion operator, both with the same structural element. This operator also tends to smooth sections of contours but, as opposed to opening, it generally fuses narrow breaks and long thin gulfs, eliminates small holes and fills in the contour.

$$\varphi_B(A) = A \cdot B = (A \oplus B) \ominus B = \varepsilon_B(\delta_B(A))$$

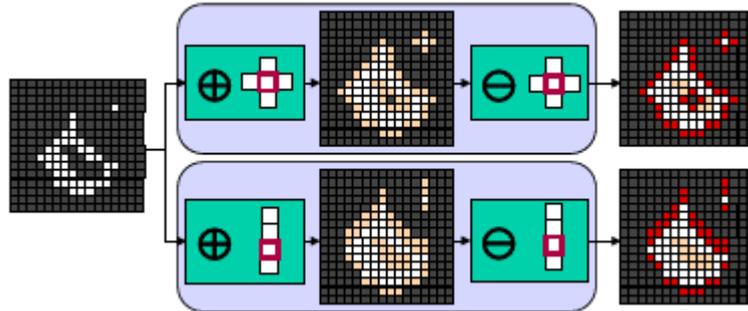


Figure 15: Example of closing [XIII]

- **Border smoothing.** Most segmentation methods produce regions with irregular borders. More natural borders can be obtained by using a variety of operations, such as major filtering, morphological filtering or curve fitting.
- **Border expansion.** It has been observed that automatic segmentation methods generate borders which are mostly contained within the borders determined manually by the dermatologists. This is because automatic procedures tend to find the sharpest pigment change, whereas the dermatologists choose the outmost detectable pigment. The discrepancy between the two borders can be reduced by expanding the computer-detected border using morphological filtering, Euclidean distance transform or iterative region growing. The latter cannot only be implemented at pixel level, but also at superpixel level by taking into account a colour consistency criterion between lesion and skin regions, as proposed in [47].

3. System design

3.1 Problem definition

This project aims to support the computer-aided diagnosis of malign melanoma at its very first step: image segmentation. The objective is to implement a system that achieves the automatic segmentation of dermoscopic images, in order to separate the lesion region from the non-lesional areas in the image. An accurate segmentation of the lesion from the background skin facilitates the posterior phases in diagnosis, since the extraction of lesion features relies on the shape of the lesion and must be applied exclusively inside the lesion perimeter, so as not to extract wrong or irrelevant characteristics that may lead to an incorrect classification of the lesion.

In order to implement an efficient automatic segmentation procedure, several segmentation methods have been studied, selected and implemented in this project. These methods allow obtaining a binary mask per image that isolates the lesion region. For their correct adaptation to dermoscopic images properties, several modifications have been also introduced in most methods, as well as specific processing for the original images and the resulting binary masks.

The performance of the different implemented methods has been analysed regarding the manual masks, which are considered as ground truth and denote the real boundaries for each lesion. A comparative study of the obtained results has been carried out, with the purpose of finding the most adequate method among the implemented ones.

3.2 Limitations

Dermoscopic images suppose a great challenge for segmentation algorithms, since there is a wide variety of lesions regarding size, shape, boundaries, colour, along with several types of skin and textures. The objective is to find an algorithm which functions correctly with all the images that are introduced in the system. The diversity among lesions might complicate the generalization of the segmentation procedure.

As a result of this variety, external and intrinsic features of skin lesions hinder the automatic delineation of lesion boundaries. External artefacts might be due to the device used to obtain the image, such as distortion, black frames or changes in illumination, or the objects used at dermoscopy observation, such as rulers or traces of fluid. Intrinsic cutaneous features include artefacts such as hair, variegated colouring, blood vessels, irregular and fuzzy borders, low contrast between lesion and background skin, diverse skin patterns such as regression or even multiple lesions. Several examples of external and internal features in dermoscopic images are showed in the following figures.



Figure 16: Inconstant illumination [I]



Figure 17: Bubbles of fluid [I]



Figure 18: Distortion [I]



Figure 19: Thin hairs [XIV]



Figure 20: Thick hairs [I]



Figure 21: Variegated colouring [I]

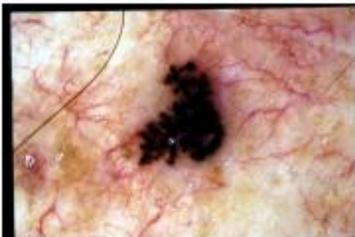


Figure 22: Blood vessels [I]



Figure 23: Irregular border [XIV]



Figure 24: Fuzzy border [I]



Figure 25: Low contrast [XIV]



Figure 26: Regression [I]



Figure 27: Multiple lesions [I]

On the other hand, the decision of potential segmentation methods to be implemented is difficult because of the scattered literature. Due to the interdisciplinary nature of the field, the existing literature is scattered among numerous medical and engineering journals [2]. There are numerous scientific publications which propose a large range of different algorithms that aim to achieve the efficient segmentation of dermoscopic images.

Besides, the study, comparison and selection between methods for the posterior implementation must take into account several factors, such as technical limitations or computational time. The complexity of an algorithm is an issue that must be considered when choosing a segmentation method. It may have been observed that some of the previously studied algorithms, such as those based on deformable contours, depend on a high number of

parameters whose values need to be determined. In general, the more the number of parameters, the harder the determination of the optimal ones [2]. Algorithms may also have limitations when their accuracy heavily depends on the colour space in which they are applied. It is unlikely that the same colour space is optimal for different methods or even for images that have been acquired by different systems [48].

Finally, it is remarkable the importance of manual masks in the analysis of the performance of automatic segmentation methods. These masks are used to measure the accuracy of every automatic mask generated by the system. It has been demonstrated in [49] that a single dermatologist, even one who is experienced in dermoscopy, cannot be used as an absolute reference for evaluating border detection accuracy. Manual borders are not completely precise, since inter-dermatologist borders and even intra-dermatologist borders may show significant disagreement. In the literature, some methods are proposed in order to accomplish an objective criterion that takes into account variations in the ground-truth images, as the one proposed in [50].

In this project only one manual mask per lesion generated by the same person has been used as ground truth to analyse the performance of the implemented methods. Therefore, it is appropriate to underline that the evaluation phase of the automatic masks is based on one single criterion of ground truth.

3.3 Programming tools

Matlab [51] has been the programming language and environment chosen for the implementation of every segmentation method included in this project. Matlab's Image Processing Toolbox and contains useful and powerful resources that facilitate the analysis and visualization of images, as well as the development of algorithms. Some of the implemented segmentation algorithms are included among the functions within the toolbox and only certain modifications in order to adapt them to dermoscopic images have been required. It is the case of threshold-based methods, edge-based methods or the standard implementation of the k-means algorithm, used in clustering-based methods or the creation of the visual vocabulary for the bag-of-words model, as it has been previously remarked.

Other more complex algorithms require external resources for their implementation. Although also programmed for Matlab, the original code for the implementation of methods based on deformable contours [52] has been utilized and modified for its better adaptation to dermoscopic images properties. On the other hand, the external library VLFeat [53] has been also included among the programming tools due to its robust techniques for the detection of keypoints, the description of keypoints based on SIFT descriptor and the posterior matching phase with the visual vocabulary. This external library has also been selected for the generation of SLIC superpixels. Another external library, ColorDescriptor [54], is the one employed to extract descriptors from colour space and not only from the luminance image.

Additionally, department servers have been requested for those processes that have required excessive computational load, such as the validation of parameters for optimal morphological operators, instead of executing them exclusively on the local computer.

3.4 Implementation

The design of the implemented system follows the conventional structure of the automatic segmentation phase inside the computer-aided diagnosis of dermoscopic images. Three major steps constitute the system:

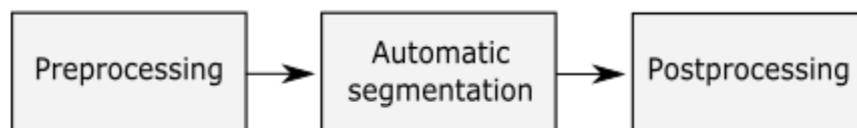


Diagram 1: Overview of system design

The first step consists in a preprocessing procedure, in order to reduce the impact of external artefacts within the dermoscopic image as well as intrinsic cutaneous features in the segmentation phase. The next step is the application of one of the selected automatic segmentation algorithms to separate the lesion region from the non-lesional skin areas. The automatic generated binary mask enters afterwards the postprocessing step, where the borders are smoothed and undesirable spurious lesion pixels or holes inside the lesion region are eliminated. The outcome of the system is an automatic generated mask that can be evaluated regarding the manual mask considered as ground truth, so that the performance of the applied segmentation algorithm can be analysed.

3.4.1 Preprocessing

As it has previously showed in Section 2.4, the preprocessing of dermoscopic images is a required procedure in order to facilitate the posterior segmentation of the image. In this project the preprocessing of dermoscopic images within the database is based on two procedures:

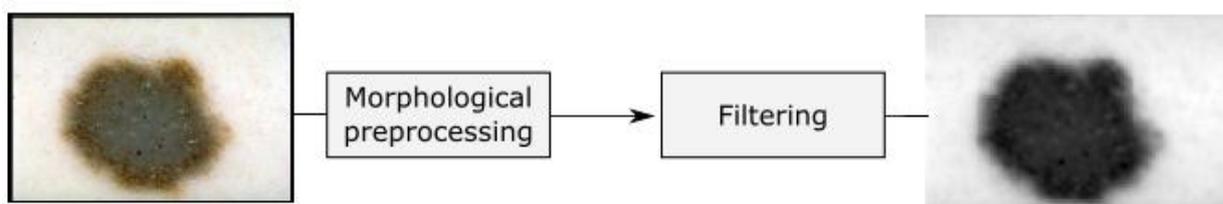


Diagram 2: Preprocessing.

3.4.1.1 Morphological preprocessing

The morphological procedure aims to eliminate a common external artefact along the images within the database: black frames. These black frames around the areas of analysis are due to

the device and the procedure that were used for taking the dermoscopic images. In order to avoid potential influence of the black borders along the performance of the system, it is necessary to eliminate them before any other procedure is applied to the images. Black frames consist in a dark and artificial-coloured area which contrasts with both the skin and lesion regions. Therefore, the segmentation algorithms may deviate from helpful image features and interpret the contrast areas around the black borders as relevant. These problematic areas contain strong edges, so the snakes from deformable models or the keypoints detector from the local features extraction may be biased towards the surrounding points. Besides, black frames suppose dark pixels and it is reflected in the image histogram, so thresholded-based methods as well as clustering-based techniques may be also affected.

The objective of the morphological procedure is to create a bounding-box that only contains the lesion and skin areas. After global and pixel-level based analysis of images within the database, it was empirically determined that the percentage of black frame at left and right as well as at top and bottom edges of the image is less than the 5% of the horizontal and vertical dimensions of the image, respectively.

The original image is first transformed into gray-scale space and the rows and columns of pixels contained within the 5% of the image edges are then analysed. The mean of luminance for each of these columns and rows is computed. Since the pixels are located on the image edges, they are supposed to be part of skin areas or, in case the lesion is expanded, not to have artificial-black colour. Therefore, if the luminance mean is under a heuristically determined threshold (a value of 0.3 within a normalized luminance histogram) then the column or row is considered as black frame and all of its pixels put to 0.

Additionally, black frames may not be compact due to image quality. This means there are cases in which rows or columns that do not satisfy the low luminance criterion are located between rows or columns that satisfy it. Therefore, if adjacent rows or columns are to be considered as black frames, the middle row or column also will.

Finally, the bounding-box is generated by extracting all rows and columns that have been considered as black frames and have value of 0. The luminance image with no black borders is then stored, as well as the dimensions of the final bounding box, in order to eliminate black frames in other colour components required by segmentation algorithms that do not operate on luminance images.

3.4.1.2 Filtering

The presence within the image of certain intrinsic cutaneous features, such as hair or irregular textures on the skin areas, may cause difficulties for the segmentation process. The application of a Gaussian smoothing operation is a common option, since a balance between blurring of undesirable features and conservation of edges can be achieved.

In this project the filtering step is common to all of the applied segmentation methods, regardless of the use of gray-scale images or different colour components. It is the scalar space, i.e., the amount of blurring, what can vary from one method to another. The selection

of the size of the Gaussian filter and the value of σ , i.e., the standard deviation used in the filter implementation, depends on which aspect the method bases the segmentation.

The most applied values are 15x15 pixels for the vertical and horizontal dimensions of the 2D Gaussian filter, and 15 also for the value of the standard deviation. It functions as an averaging filter rather than a Gaussian filter, since the kernel size is relatively small for the value of σ [27]. However, a large filter may be detrimental to methods based on deformable models, since they base part of the computations on edge-based methods. Large filters may result in dislocation of edges and even false edges, so a smaller filter is used as part of the preprocessing step before the application of these algorithms to the dermoscopic images.

3.4.2 Automatic segmentation

The second and decisive step of the system is the automatic segmentation phase. Several algorithms based on different state-of-the-art segmentation techniques studied along the Section 2.3 have been implemented in order to achieve an efficient segmentation of dermoscopic images. The global and specific limitations for the implementation, which have been previously mentioned in Section 3.2, have been taken into account for the selection and implementation of the different methods.

In this section the implemented methods that are able to generate an analysable automatic generated mask are described. Every algorithm includes the introduced modifications for its better adaptation to the specific segmentation problem of dermoscopic images. The last section inside each segmentation technique includes the alternatives of design that were considered at the implementation phase.

3.4.2.1 Threshold-based methods

3.4.2.1.1 Traditional Otsu method

The Otsu method, as studied in Section 2.3.1.1, is a point-dependent thresholding technique, since a threshold is determined from the gray level of each pixel. The traditional version of the method consists in global thresholding because the entire image is thresholded by one threshold value. The process is divided in three steps:

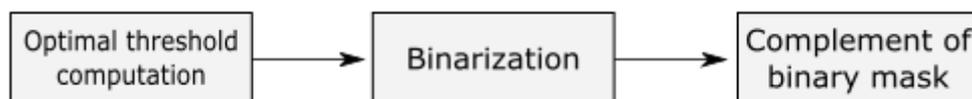


Diagram 3: Traditional Otsu method.

The preprocessed luminance image is the input of the system. Firstly, the optimal threshold value is computed, i.e., the threshold that maximizes the separability of the resultant classes in

gray levels. Then the image is segmented into two classes regarding the threshold value. A binary mask which contains one class for lesion regions and another class for skin regions is obtained, but it is necessary to compute its complement. Since the highest gray levels above the threshold have been assigned to skin areas due to their brighter luminance, the white zones in the resultant binary mask correspond to skin and the black areas to lesion. After computing the mask complement, black and white areas are reversed and the mask can be postprocessed.

The next figures show the original image of two different lesions, followed by the preprocessed luminance image used for Otsu segmentation. In Fig. 32 the luminance histogram shows a sharp valley and the optimal threshold computed by the method can be easily adjusted between lesion and skin areas. However, in Fig. 33 the luminance histogram shows several peaks with several valleys in between, which correspond to colour-variegated lesion areas. This distribution complicates the segmentation of the lesion region by computing a unique threshold and a more complex method might be required.

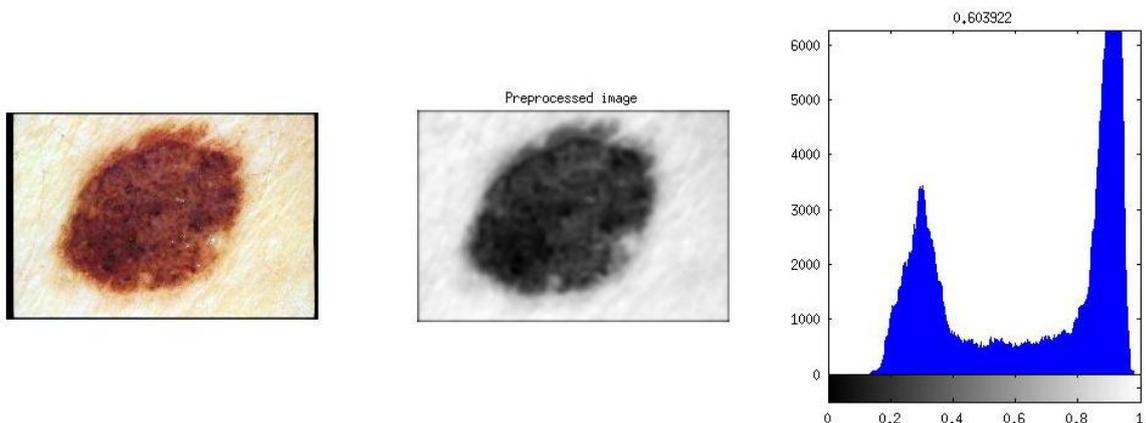


Figure 28: Traditional Otsu method. From left to right: original image, preprocessed luminance image, luminance histogram with threshold value situated on top (0.604).

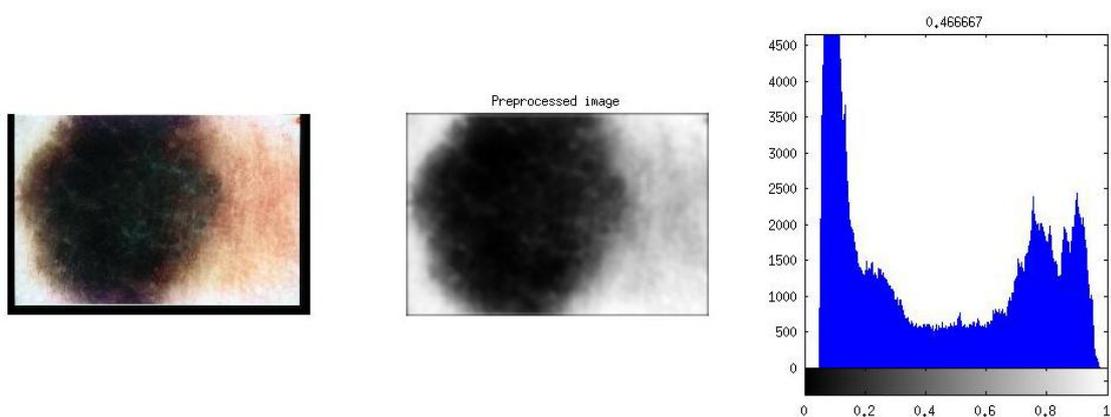


Figure 29: Traditional Otsu method. From left to right: original image, preprocessed luminance image, luminance histogram with threshold value situated on top (0.467).

3.4.2.1.2 Local Otsu method

Local thresholding allows the adaptation of point-dependent thresholding techniques, such as the traditional Otsu method. In local thresholding methods, as studied in Section 2.3.1.2, the image is divided into smaller sections and a threshold is computed for each partition individually. This results in gray level discontinuities at the boundaries of each two partitions. Thus, a smoothing algorithm is required in order to reduce these discontinuities. The preprocessed luminance image is again used as input of the system. The process may be divided in five steps:

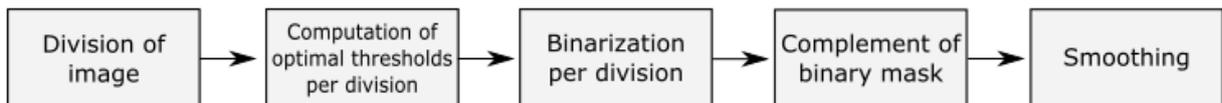


Diagram 4: Local Otsu method.

Regarding the number and shape of the partitions in which the image is divided and thresholded, two different choices versions of the local-thresholding were implemented.

➤ Local Otsu on grid

In this implementation of the algorithm each partition is equivalent to a tile of a grid that covers the entire image. The number of tiles was established to be 6x6. Therefore, each partition has rectangular shape and contains 1/36 area of the image. Regarding the location of each tile, each one can exclusively contain part of skin area, lesion area or pixels from both classes. The optimal threshold is computed on each partition. Therefore, a 6x6 matrix of threshold values is obtained and it constitutes a binary mask, whose complement is computed in order to have lesion areas as positive values and skin areas as negative ones.

In order to reduce the discontinuities between tiles within the grid, two smoothing algorithms are analysed: Gaussian filter and circular averaging filter, i.e., pillbox filter. The Gaussian filter consists in a Gaussian low-pass filter with an assigned size and standard deviation, whereas the pillbox filter has a circular top and straight sides.

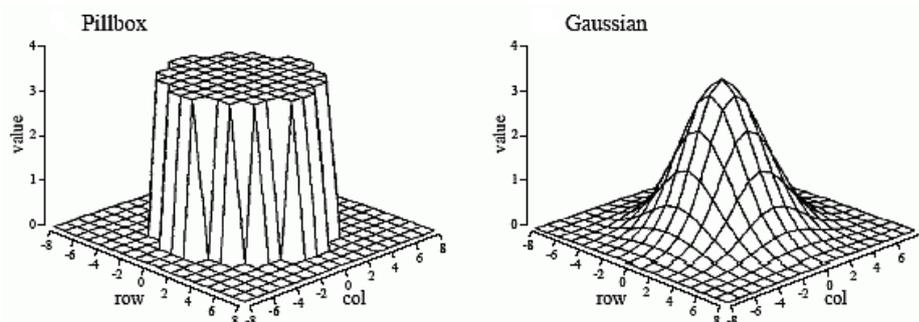


Figure 30: Pillbox and Gaussian filters [XVI]

Large filters show better smoothing of discontinuities than small filters. If a large Gaussian filter is selected, there is good concatenation between the partitions that constitute the lesion region but spurious lesion areas are amplified instead. If a large

pillbox filter is applied, the smoothing of discontinuities improves and less spurious lesion areas remain. This can be observed in the next figure.

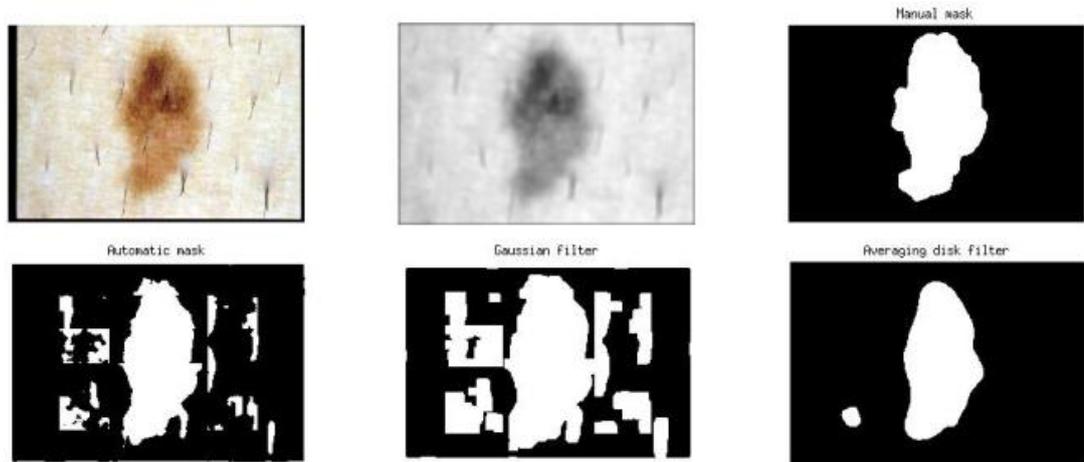


Figure 31: Local Otsu on grid. On top: original image (left), preprocessed luminance image (center), manual mask (right). On bottom: binary mask generated by local Otsu on grid (left), mask after Gaussian filtering (center), mask after pillbox filtering (right).

➤ **Local Otsu on concentric rectangular partitions**

In this version of the algorithm the image is divided into three different areas. Each area is concentric and rectangular and aims to represent one of the three common parts of dermoscopic images: the inner partition is associated with the center of the lesion, the middle partition is associated with the union between lesion and skin, and the external partition aims to represent the skin areas. The next figure shows how the partitions of the image are generated from three different binary masks and how the resulting thresholding looks like for each one.

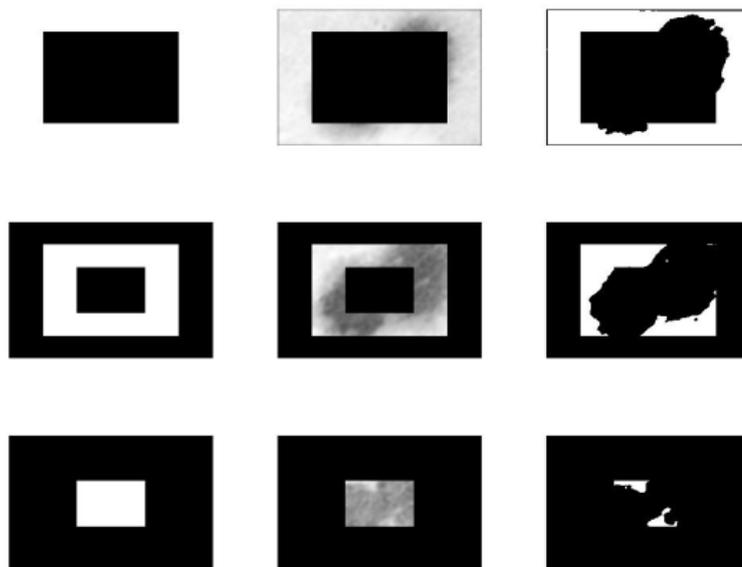


Figure 32: Local Otsu on concentric partitions. From left to right: binary masks of external (top), middle (center) and inner (bottom) partitions, areas of image under each partition, thresholded partitions.

The three thresholded partitions are then concatenated and the complement from the resulting binary mask is obtained. A pillbox filter is used as smoothing algorithm to reduce the discontinuities. This time the pillbox filter is smaller than the pillbox filter applied to discontinuities on grid. Since the number of regions to concatenate is smaller, the discontinuities are less numerous and there are also fewer spurious lesion areas within the binary mask.

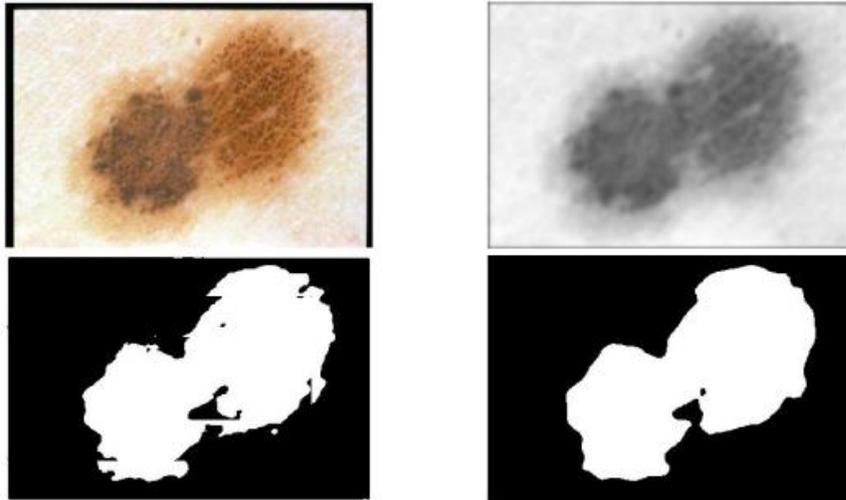


Figure 33: Local Otsu on concentric partitions. On top: original image (left) and preprocessed luminance image (right). On bottom: binary mask before (left) and after smoothing (right).

Regarding the concatenation of partitions thresholded by the local application of the Otsu method, the implementation based on concentric rectangular partitions was selected over the grid implementation in order to be compared with other techniques in the evaluation phase afterwards. The generated binary masks were visually more compact and analyzable.

3.4.2.1.3 Multithresholding Otsu method

Many global thresholding methods, including the traditional Otsu method, can be adapted to multithresholding. The main advantage of the multithresholding Otsu method over the previously implemented traditional algorithm, which establishes a single threshold value for the entire image, is that allows a more accurate segmentation of those lesions that contain several colour/texture-variegated areas. Those areas can be individually detected by extracting several threshold values from the luminance histogram. This way the segmentation of complex images that cannot be easily divided into two classes is more complete and exact.

The procedure is may be divided in five steps:

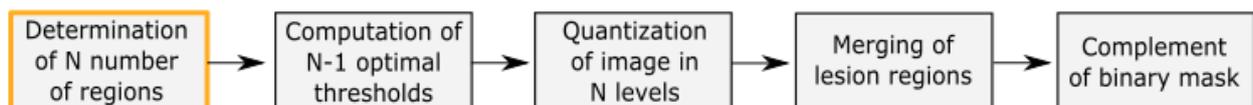


Diagram 5: Multithresholding Otsu method. Determination of number of regions in Diagram 6.

This algorithm achieves more than a direct binarization for complex lesions, but the number of N different regions within the image must be previously defined in order to compute N-1 threshold values and quantize the image regarding those values. The method proposed in this project to determinate the number of regions into which the image must be segmented is based on the analysis of local maxima in the histogram of luminance image and is described below.

➤ **Method to determine the number of regions within a dermoscopic image based on the analysis of local maxima in the luminance histogram**

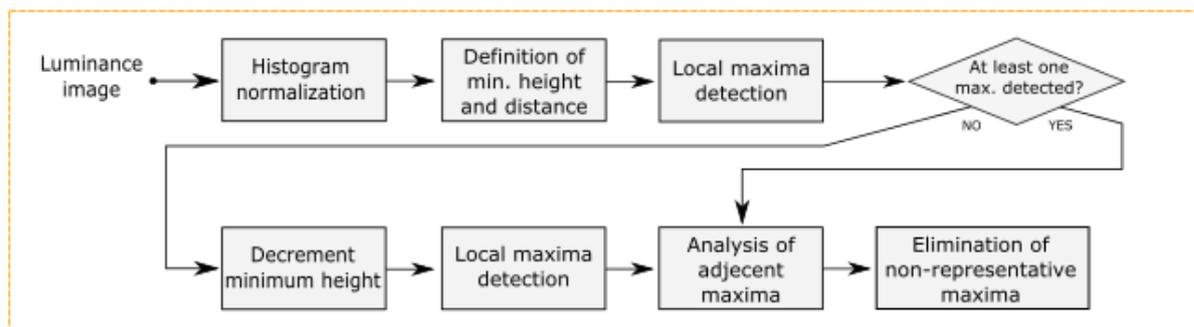


Diagram 6: Method to determine number of regions within a dermoscopic image.

The input is the preprocessed luminance image, whose histogram is first normalized. By empirical tests throughout the dermoscopic images within the database, the optimal values for the minimum distance between peaks and the minimum peak height were obtained.

In a normalized luminance histogram the height of a peak is equivalent to the percentage of occurrence of a certain gray level. Regarding the wide range of levels in a dermoscopic image due to the variety of areas, textures and features, it was observed that a level that supposes the 3% of the total image is enough to generate a representative maximum. Thus, this value is considered as the minimum percentage of occurrence.

Additionally, a minimum difference in luminance between local peaks must be also determined, in order to discard those maxima that are too close and may represent too similar gray levels. The range of luminance goes from 0 to 255 and this value was set to 40. Thus, regarding the distance between local maxima, a total number of 6 maxima could be detected. This supposes a first control measure to avoid oversegmentation problems.

Those peaks that surpass the established threshold values in height and distance between levels are to be considered local maxima. If these minimum values are not surpassed due to dispersed levels of luminance, the minimum height is decremented until the 0.1% of the image, whereas the minimum distance is maintained.

After relevant local maxima in the normalized luminance histogram are extracted regarding the previous parameters, an extra limitation between close adjacent maxima is imposed in order to reduce the influence of non-representative maxima and a posterior oversegmentation of the image. When two adjacent local maxima have a gray level that is close enough, i.e., 60/255, they are further analysed: if one of the two maxima has a height near to the minimum height value, it is then considered as a non-relevant maximum and is consequently discarded, since the adjacent maximum has more weight within the image and already represents the luminance level of the discarded one.

The number of required threshold values must be equivalent to the number of determined regions separated by representative local maxima minus one. Once the approximated number (N) of regions within the image is determined, N-1 thresholds are computed by the multithresholding adaptation of the Otsu method, studied in Section 2.3.1.3.

As a result, the pixels within each segmented region have one value from 1 to N. The algorithm assigns the lowest value to the region that contains lower levels of luminance, as the traditional algorithm, but in this case there may be more than one region inside the lesion area. Therefore, it is necessary to merge all regions considered as lesion into one. After the merging process, there are finally two classes within the image, i.e., a binary mask is obtained. Since the lesion areas must be represented by white pixels and skin areas by black pixels, the complement of the binary mask is computed.

The next figure shows a colour-variegated lesion, whose luminance histogram shows several peaks and valleys, so more than one threshold is required in order to achieve an accurate segmentation. The number of extracted local maxima is three, so two threshold values are computed by the multithresholding Otsu method. The image is then segmented into three different regions. Two of them are part of the lesion and are merged into one to obtain one single lesion region. The complement of the generated binary mask is obtained and ready to be postprocessed.

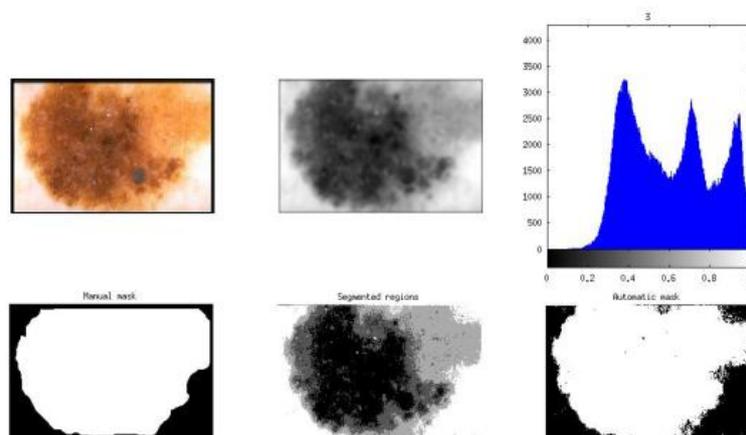


Figure 34: Multithresholding Otsu method. On top: original image (left), preprocessed luminance image (center), luminance histogram with number of required thresholds above: 3 (right). On bottom: manual mask (left), segmented regions (center), binary mask after merging of regions and complement computing (right).

3.4.2.1.4 Alternatives of design

The **local Otsu method** shows several alternatives of design, such as the size and shape of the selected partitions in which the algorithm is separately applied. As for the grid division, square tiles with a certain side dimension in pixels were proposed, but the total number of partitions may change if the image size varies. Therefore, a fixed number of partitions was implemented, so that matrix of known dimensions that contains the threshold values of each partition is obtained from each image, which makes easier to study the performance of the algorithm among lesions. In the case of the concentric rectangular partitions, more concentric divisions could have been implemented, but this would lead to higher computational time and a potential oversegmentation of certain areas.

The local application of the Otsu method using a sliding window along the image was also proposed. But this technique was discarded because of the computational load and time that this algorithm requires and the increasing number of discontinuities to smooth due to the generation of numerous image partitions.

Eventually, the main alternative of design that the **multithresholding Otsu method** offers is the criterion to determine the number of segmentation regions. The one proposed in this project employs the luminance histogram and minimum threshold parameters, related to gray level occurrence and distance between gray levels, adapted to the properties of dermoscopic images.

3.4.2.2 Region-based methods

3.4.2.2.1 Two-region k-means clustering

K-means algorithm, as studied in Section 2.3.2.1, is a partitional clustering technique that requires a priori specification of the number of clusters, that is, it is necessary to introduce the number of final regions into which the image is divided. If the accordant number of regions for each image is not known, the first and most general approach to dermoscopic image segmentation is the determination of two initial classes: one for lesion regions and one for skin regions.

The entire procedure may be divided in five steps:

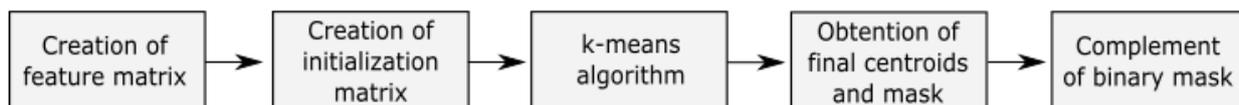


Diagram 7: Two-region k-means clustering.

The first step consists in the generation of a matrix that contains the features of each pixel in the image, which will be analysed by the algorithm in order to assign each pixel its nearest centroid. Therefore, regarding the values of its features, a pixel will join the cluster associated with lesion regions or the one associated with non-lesional skin. The features that define each

pixel and k-means algorithm will use in order to create the clusters can be selected from a wide range of components and variables. This design contemplates two main pixel features: colour components and location within the image.

Different colour spaces, which are described in Appendix A.1, have been considered and three different versions of the algorithm were implemented, each one with a specific combination of colour components:

- The *R*, *G* and *B* components from the *RGB* colour space.
- The *S* and *V* components from the *HSV* colour space.
- The *L*, *a* and *b* components from the Lab colour space.

In the feature matrix each pixel is assigned its value for each one of the selected components, which have been previously disposed of the black frames using the bounding-box dimensions stored during the preprocessing phase of the system. Each colour space provides different information and, regarding the relevance of this information, the algorithm achieves a more or less accurate partition of the data into two clusters.

In dermoscopic images the particular location of pixels within the image may contain significant information about the distribution of lesion and skin regions. Regarding the importance of spatial information, the location of each pixel is an additional feature included in the matrix that is common to all versions of the algorithm. Since the prevailing distribution of the lesion area in dermoscopic images goes from the center towards the borders of the image, the location assigned to every pixel is based on the modulus of radial distance from the pixel to the center of the image. The center of the image is considered to have (0,0) coordinates, so that the vertical and horizontal dimensions are considered as *y* and *x* axes, respectively, whose range goes from -0.5 to 0.5. Therefore, the radial modulus of each *i*-th pixel is computed as follows:

$$\text{Modulus}_i = \sqrt{x_i^2 + y_i^2}$$

Once the components inside the feature matrix are concatenated, the next step is the creation of the initialization matrix. K-means algorithm is highly sensitive to initialization, so for each one of the two initial cluster centroids a particular value from each of the components inside the feature matrix must be assigned and introduced in the initialization matrix. The values must be representative of both clusters (lesion and skin), since the algorithm will interpret these values as a reference to assign the pixels to one of the two clusters. For both centroids the value introduced of each colour component, whose range goes from 0 to 1, is 0.5, as no precise value for each component is primarily associated with each region. Regarding the value of location introduced for each initial cluster centroid, the lesion centroid is assigned the center of the image, i.e., modulus equal to 0, whereas the skin centroid is assigned one of the corners, i.e., the maximum modulus, which is 0.707 and corresponds to the left top corner of the image.

The third step is the application of the k-means algorithm, introducing the number of final clusters (one for lesion and one for skin), the feature matrix and the initialization matrix. After the iterative procedure of the algorithm is finished because the convergence condition has

been reached, the final centroids and a binary mask are obtained. Due to the introduced fixed initialization of the cluster centroids, the algorithm always assigns each one of the clusters to the same tag: pixels which belong to the skin cluster are always tagged as white. Thus, it is necessary to compute the complement of the obtained mask as a final step.

3.4.2.2.2 Multi-region k-means clustering

The division of all images within the database into two regions may be simplistic for those lesions that contain several colour-variegated areas and complex features. Therefore, the determination of the number of clusters in which the k-means algorithm divides the data set can be adapted to a dynamic selection of the adequate number of regions for each image. A specific method to determine the most suitable number of regions has been implemented for this system.

The entire procedure may be divided in the following steps:

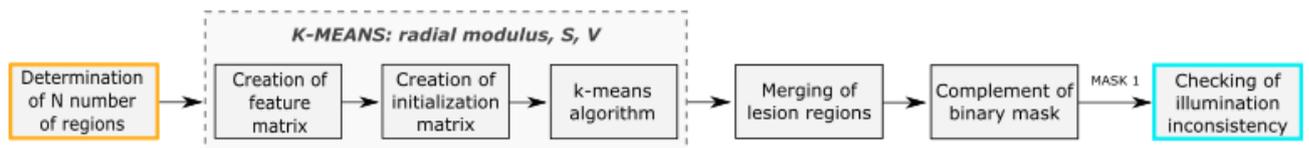


Diagram 8: Multi-region k-means clustering. Determination of number of regions in Diagram 6. Checking of illumination inconsistency in Diagram 9.

The first step is the determination of the number of regions in which the image must be divided, i.e., the number of clusters in which the k-means algorithm will group the pixels. The method proposed in this project for the determination of a suitable number of regions within the image is based on the analysis of local maxima in the histogram of luminance image and is described in Section 3.4.2.1.3, since it is also used before the application of multithresholding algorithms.

Once the method is applied, the total number of representative local maxima is considered as the number of clusters to be introduced in the k-means algorithm. If only one representative maximum has been detected, the number of clusters is automatically two. Among the different versions of the algorithm that have been implemented, one of them was selected as a step belonging to the multi-region adaptation of k-means because of the obtained results (showed in Section 4.3.2.1). The implementation whose feature matrix contains the radial distance of each pixel to the center of the image along with the components S and V from the HSV colour space is the one applied.

All the steps inside the k-means phase are the same as previously described, but the initialization matrix must be now adapted to the determined number of clusters. In case there are more than two clusters, the values of S and V for each initial cluster centroid are again 0.5, but the value of their radial distance is now dynamically applied regarding the number of clusters. On a straight line which goes from the center to the left top corner of the image, each j -th centroid is assigned a value at a proportional distance from the center, following the next numerical relation:

$$Modulus_j = \frac{\max modulus}{n - 1} \times (j - 1)$$

where *max modulus* denotes the maximum possible value of modulus, which is 0.707 and corresponds to the left top corner of the image, *n* is the number of local maxima detected and *j* is the number of centroid for which it is computed the modulus, taking values from 1 to the number of local maxima, so that the first centroid is always initialised as the center of the lesion and the last centroid as the furthest one. That last centroid is the one which will always correspond to skin.

After the k-means algorithm converges, a mask with *n* classes is obtained. The applied initialization of the clusters guarantees a stable distribution of classes and makes possible to merge all regions that belong to the lesion into one, separating them from the most external region, which corresponds to non-lesional skin.

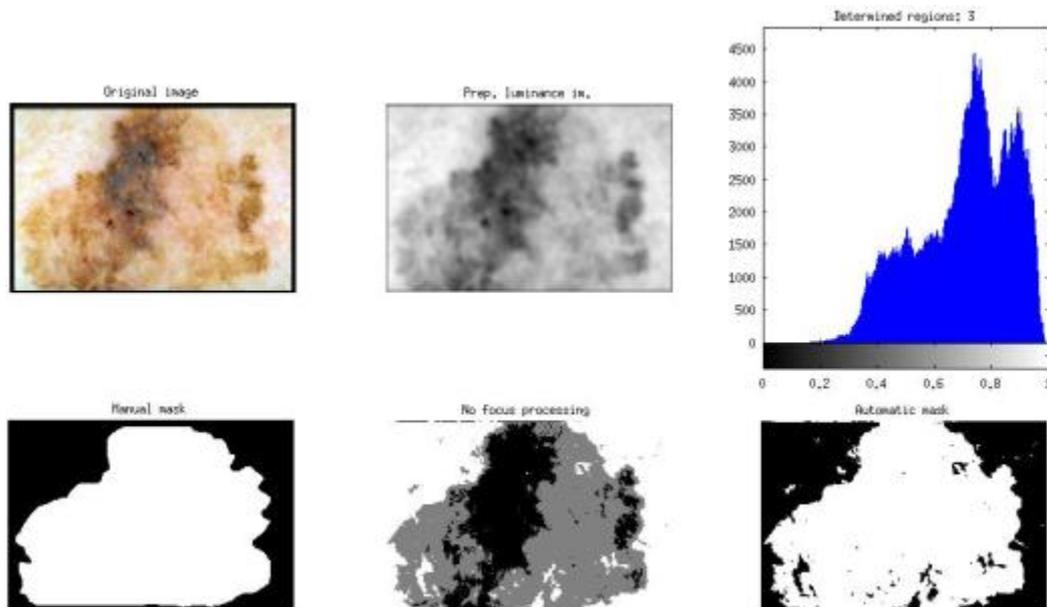


Figure 35: Multi-region k-means. On top: original image (left), preprocessed luminance image (center), luminance histogram with number of determined regions above, 3 (right). On bottom: manual mask (left), mask before region merging (center), mask after region merging and complement (right).

The last step of the procedure is oriented to solve the segmentation problems caused by the illumination inconsistency in some images within the database. Since dermoscopic images depend on the device that was used to take the image, it was observed that changes in illumination may be a common issue that needs specific processing.

➤ **Method to address illumination inconsistency in dermoscopic images**

Illumination changes within a dermoscopic image suppose an external feature that can affect the output of the segmentation phase. Both lesion and skin areas may adopt colours or textures that are no longer representative of those regions and therefore

some segmentation algorithms, including k-means clustering, cannot achieve an accurate separation of lesion areas.

The most common consequence of illumination inconsistency is the apparition in the mask of an ellipsoidal region around the real lesion due to the focus of light that is present over the lesion region in the original image. It was observed that around 1/10 of the database used for experiments was affected by this problem at the beginning, so it was concluded that image inconsistency might be a persistent handicap related to dermoscopic images and needs to be solved in order to obtain an accurate segmentation, especially for those simple and compacted lesions that present no other difficulty at segmentation apart from this one.



Figure 36: Multi-region k-means. From left to right: original image, manual mask, automatic mask before image inconsistency processing

The procedure applied to find a solution to illumination inconsistency follows the next order:

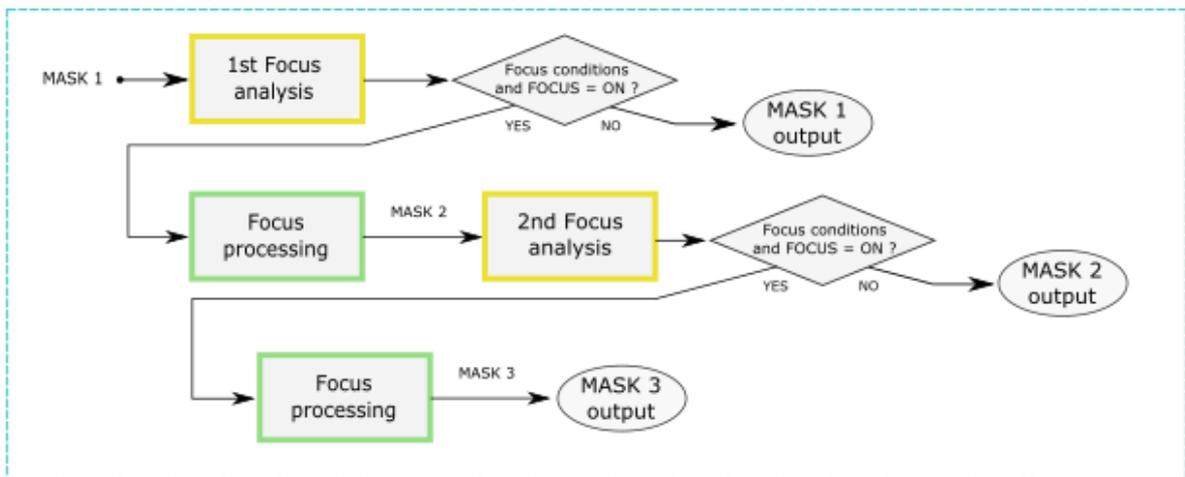


Diagram 9: Method to address illumination inconsistency in dermoscopic images. Focus analysis in Diagram 10. Focus processing in Diagram 11.

The first step of the method is called focus analysis and has two branches: shape analysis of the lesion region and maxima analysis in luminance histogram. If the two branches satisfy the imposed focus conditions, the focus processing is applied. If not, the image is not processed because it is considered not to have illumination inconsistency.

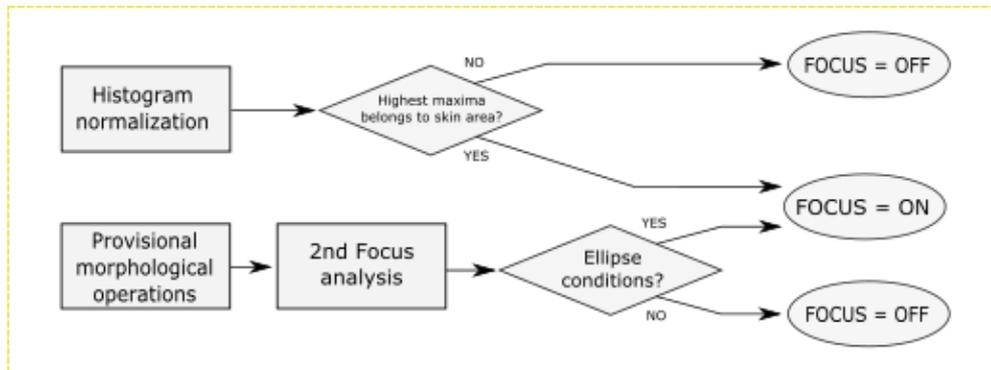


Diagram 10: Focus analysis.

Since the aim is to localize ellipsoidal lesion regions that can be a result of image inconsistency problems, one branch is based on the analysis of the shape of the lesion region generated by the previous segmentation algorithm. As a first step, it is necessary to carry out provisional morphological operations in order to generate a compact mask and extract its shape properties. The parameters used for this morphological opening and closing operations are the optimal ones obtained after the cross-validating postprocessing of the k-means based on components S and V and radial distance, procedure explained in Section 3.4.3.

Once the compact mask is obtained, several shape properties are extracted from the lesion region. If there are multiple lesions, only the region with biggest area is selected. The next shape properties were selected for evaluation: eccentricity (from 0, circle, to 1, line segment), orientation of horizontal axis (from -90° to 90°) and solidity (percentage of compactness). They have a big influence in the definition of an ellipse. If certain values, according to a standard ellipse shape, for the three parameters are satisfied, then the focus condition is positive for this branch.

An additional condition is imposed on the other branch: the highest local maximum belongs to the skin region. The focus area covers a compact lesion in the center and a big area of skin in the surroundings.

If the conditions are met, then the image is potentially affected by illumination inconsistency and the focus processing is applied.

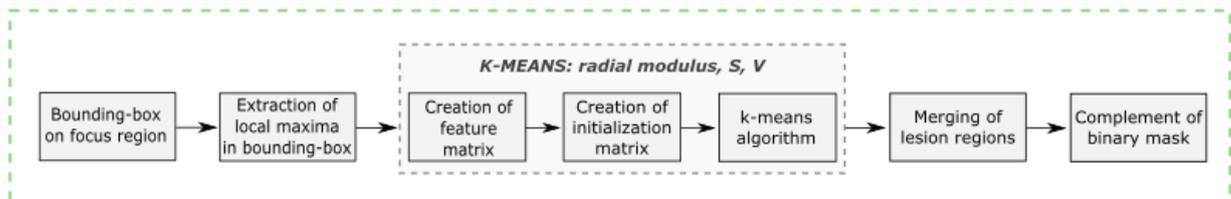


Diagram 11: Focus processing.

The coordinates of the smallest convex polygon that contains the focus region are obtained. The edges of the polygon are located, so that a bounding-box which only contains the focus area is extracted from the original image. Then, within the

bounding-box, the local maxima are extracted following the previous procedure and the multi-region k-means algorithm is applied. The lesion regions are then merged into one and the complement of the mask is computed.

After the focus processing, the ellipsoidal focus within the mask is expected to have disappeared and the new mask is considered as the definitive one.

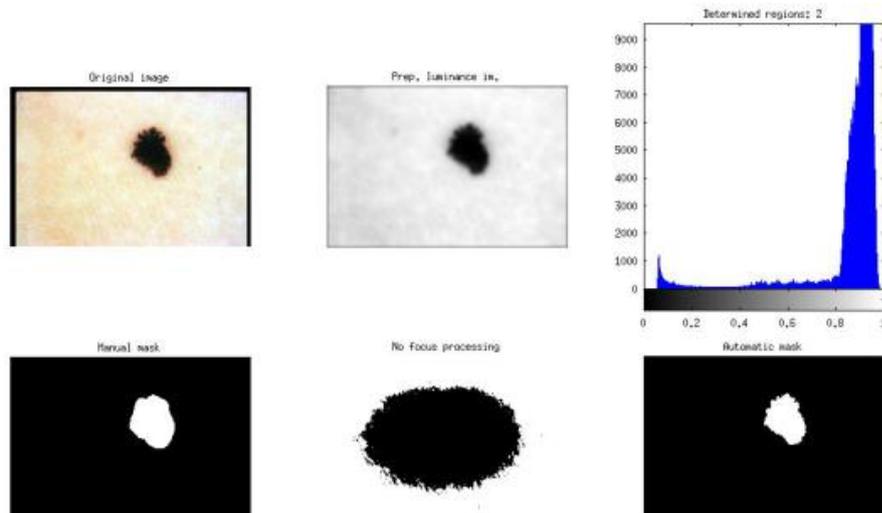


Figure 37: Multi-region k-means. On top: original image (left), preprocessed luminance image (center), luminance histogram (number of determined regions: 2) (right). On bottom: manual mask (left), mask previous to focus processing (center), automatic mask after one focus processing (right).

If not, the ellipsoidal area is still present in the mask but its ellipsoidal shape is likely to be smoother. Thus, a second focus analysis is carried out, but this time the imposed conditions on eccentricity, orientation and solidity are less restrictive. The same conditions on local maxima are imposed. Then a second and last focus processing is carried out and the generated mask is now the definitive one.

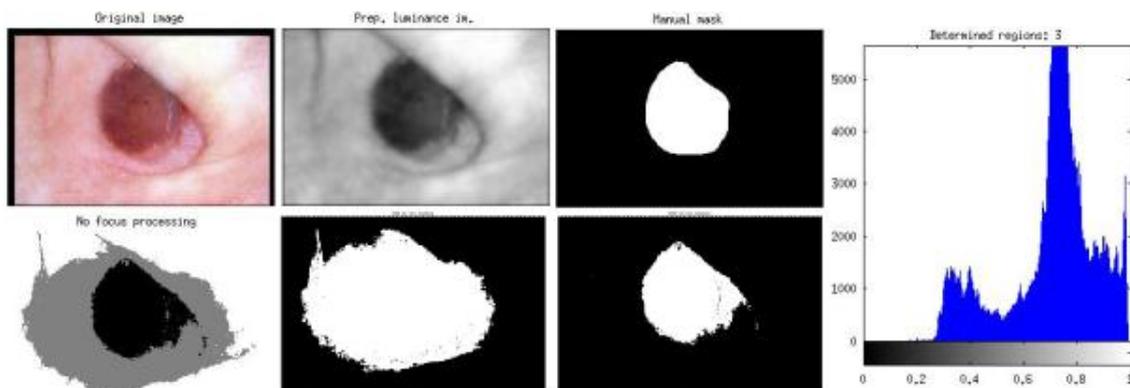


Figure 38: Multi-region k-means. On top: original image (left), preprocessed luminance image and manual mask (center), luminance histogram (number of determined regions: 3) (right). On bottom: mask before region merging and focus processing (left), mask after first (center) and second (right) focus processing.

3.4.2.2.3 Alternatives of design

The main alternative of designs that the k-means algorithm offers are related to two different aspects: feature matrix and distance minimization criterion.

The first aspect consists in the features that are selected to define the pixels and associate them with the most similar cluster. Several options were contemplated. Combinations between different colour spaces are not recommended in the literature, since their components are measured around different axes and properties. The addition of luminance to colour components was also considered, but early discarded for the same previous reason, since luminance properties may differ enough from the other components. Nevertheless, spatial information was also considered a key feature, since previously studied segmentation methods, such as thresholding-based methods, lack spatial information and sometimes the final regions are generated with this handicap. Additionally, the component H was discarded in the version of the algorithm based on the HSV colour space and only S and V components were used. This is due to the distorted information that the H component provides. As mentioned in [37], the hue becomes unstable near the gray axis, which can also visually asserted, and further error propagation analysis is required in order to its addition.

The second aspect is the distance measure used for minimization. Depending on the selected distance measure, the k-means algorithm computes centroid clusters differently. The selected measure over other measures included in [56] was squared Euclidean distance: each centroid is the mean of the points in that cluster. This criterion is the most common one and therefore it makes the results easier to compare with those included in the literature.

3.4.2.3 Segmentation based on deformable models

3.4.2.3.1 Gradient vector flow snakes

Due to the limitations of edge-based methods when applied solely to achieve the segmentation of dermoscopic images, as well as the handicaps of traditional snakes because of their capture range and sensitivity to initialization, gradient vector flow (GVF) snakes were selected over these techniques as part of the system.

GVF snakes, as described in Section 2.3.4.1.2, are a type of parametric deformable models and are based on the energy minimization formulation. In order to find the boundary of the object to be segmented, these curves are initialized within the image domain and are forced to move toward the potential energy minima under the influence of two groups of forces: internal forces, associated with the elasticity and rigidity of the snake, and external forces, which drive the contour toward the edges. GVF snakes employ the gradient vector flow as an additional external force.

The procedure may be divided in the following steps:



Diagram 12: Gradient vector flow snakes.

The first step is the definition of an edge map $f(x, y)$ derived from the image $I(x, y)$, having the property of being larger near the image edges. Any gray-level component can be used, but it is recommendable that the lesion is as emphasized as possible to facilitate the boundaries detection. The blue channel (B) from the RGB colour space was selected due to its high entropy value and its good discrimination between lesion and skin. Additionally, the selected component is smoothed by a Gaussian filter. The selected Gaussian filter is not as large as in the usual preprocessing phase of the other methods implemented for the system, in order to preserve the borders to a greater extent while keeping the blurring of image artifacts and noise. This time the filter size is 6×6 and the standard deviation is 2.5.

The edge map is then computed as follows and showed in the next figure:

$$f(x, y) = -|\nabla[G_\sigma(x, y) * B(x, y)]|^2$$

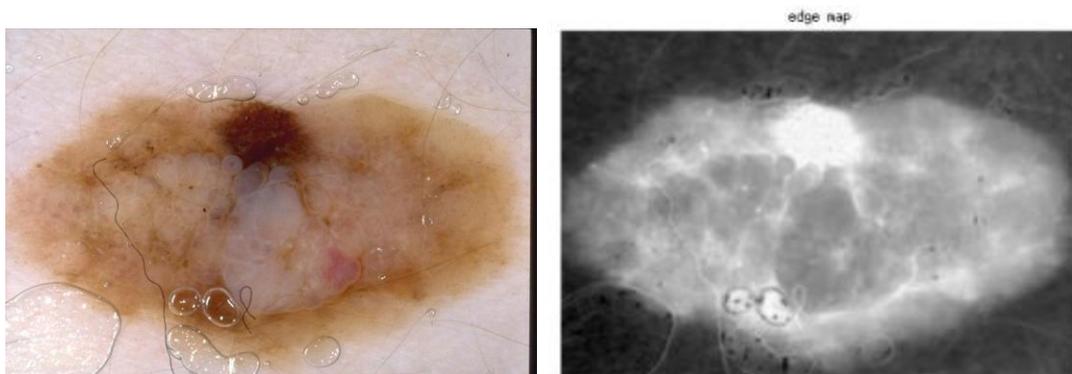


Figure 39: GVF snakes. From left to right: original image and edge map.

Then the numerical gradient of the edge map is computed in the vertical and horizontal directions. These two directional components and the squared magnitude of the gradient field, which is calculated from them, are used at the computation of the GVF field.

The next step consists in the computation of the GVF field. Two parameters must be defined: the number of iterations in which the computation is carried out and μ , which should be set according to the amount of noise in the image (more noise, higher μ). The selected number of iterations was 80 and a small value for the μ parameter was fixed because a better performance was observed, in spite of reducing the strength of the smoothing term. Once the GVF external force is obtained, it is normalized.

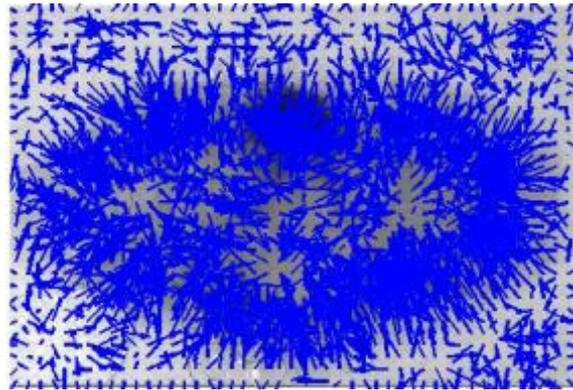


Figure 40: GVF snakes. Normalized gradient vector field.

The previous figure shows a higher concentration of the normalized GVF field around the boundaries of the lesion, while there is also information in the homogeneous areas. As mentioned in Section 2.3.4.1.2, the GVF field aims to keep the desirable property of the gradients near the edges and at the same time extend the gradient map farther away from the edges and into homogeneous regions using a computational diffusion process. The objective is to increase the capture range of the snake and to achieve a better adaptation to concavities within the boundaries. A comparison between the vertical and horizontal components of the gradient and the gradient vector flow can be found in the next figure.

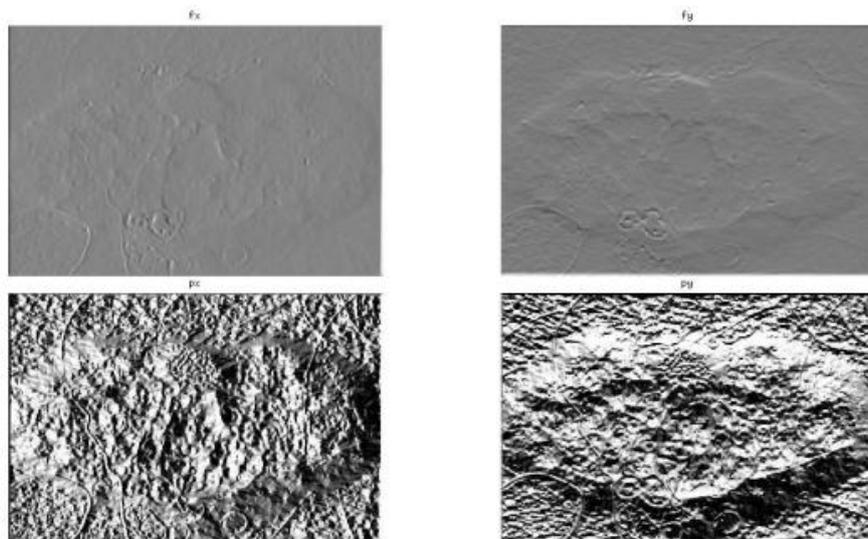


Figure 41: GVF snakes. On top: horizontal and vertical components of gradient. On bottom: horizontal and vertical components of gradient vector flow.

The next step within the procedure is the generation of the initial curve. In order to avoid sensitivity to initialization to the extent possible and better adaptation for each lesion, the initial curve is based on the mask obtained from the multi-region k-means algorithm, which is close to the real lesion boundaries in the vast majority of images within the database. A first curve is extracted from the boundaries of the lesion area within the binary mask and then a radial expansion of the curve is carried out, so as to increment its perimeter and the area inside. The expansion procedure is mathematically described in Section 3.4.3.2. Then 1/10 equidistant points of the total number of points that constitute the curve is selected. This way,

the initial curve is slightly simplified and accelerates the deformation process while maintaining its shape.

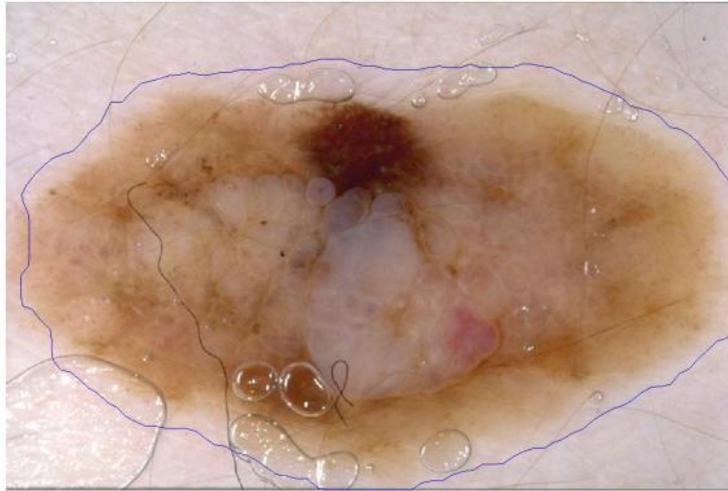


Figure 42: GVF initial curve. Original image without black frames and initial curve on blue.

The following phase is the deformation of the curve. The number of iterations until the curve reaches its final position was defined to be 100. In the deformation process, the number of points in the initial curve is maintained and they only change their coordinates within the dimensions of the external force field, i.e., no interpolation step is carried out.

The curve deformation depends on three main parameters, apart from the number of iterations: α , β and κ . The latter controls the weight of the external forces with respect the internal forces; the bigger its value, the stronger the force toward the image edges. The optimal value for κ is close to the default value of the original algorithm, 0.5, which was the selected value and determines that the external forces have half the weight of the internal ones. On the other hand, α and β regulate the internal forces by controlling the tension and the rigidity of the snake, respectively: α controls the penalty for internal elasticity and bigger values mean less stretching; bigger values of β mean the curve is harder to bend. Since their influence on the deformation of the curve may be decisive, a validation of α and β by a cross-validation procedure was carried out, whose implementation is described in Appendix B.

The accuracy of the final curve with different values of these two parameters was computed for five different groups of randomly selected images within the database. The obtained optimal values were 0.25 for α and 0.05 for β , within a validated range from 0.05 to 0.8. The next figure shows the relation between the values of α , β and the accuracy of the curve in one of the groups inside the cross-validation process.

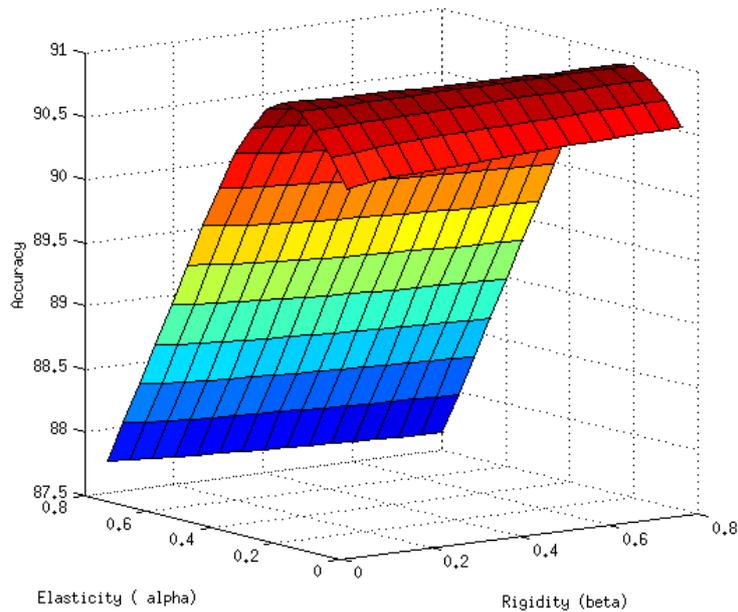


Figure 43: Validation of internal forces parameters. Example of accuracy evolution regarding the modification of the elasticity and rigidity parameters.

Once the deformation process is finished, a binary mask is generated from the final position of the snake. The area inside the perimeter of the snake is considered as lesion region. The pixels outside the snake perimeter are considered to be non-lesional skin.

3.4.2.3.2 Chan-Vese method

The Chan-Vese method, also referred to as active contours model without edges, is a geometric deformable model, since it is based on the theory of curve evolution and the level set method, as described in Section 2.3.4.2.1, the model is supposed to detect boundaries not necessarily defined by gradient.

The procedure may be divided in four steps:



Diagram 13: Chan-Vese method.

In order to facilitate the deformation process and achieve an accurate segmentation, a suitable gray-level component must be selected. The luminance image was selected as a generic input, as well as B channel due to its high entropy value and good discrimination between lesion and skin areas. In parametric deformable models, the smoothing filter applied to the input of the deformation may reduce the noise but it smoothes the edges too. Since the Chan-Vese method stopping term is based on Mumford-Shah segmentation techniques, the model is

supposed to detect better the real boundaries. Consequently, no smoothing is applied this time to the gray-level component before the deformation, as considered in [28].

The next step is the initialization of the contour. Although in this model the initial curve can theoretically be placed anywhere in the image, an initial contour position that is close to the desired object boundaries may allow faster and more accurate segmentation results, as described in [29]. Therefore, the mask obtained from the multi-region k-means segmentation is also used in this method for the generation of the initial curve, due to the accuracy of lesion regions in most images within the database. Nevertheless, the curve obtained from the boundaries of the lesion region in the automatic mask is slightly radially expanded. Then it goes under a simplification procedure before entering the deformation process, so as to improve computation cost and time while maintaining its original shape, by selecting 1/10 equidistant points within the curve. The next figure shows a comparison between the original segmentation mask and the input mask of the deformation: expanded and simplified.



Figure 44: Chan-Vese method. From left to right: original image, multi-region k-means segmentation, Chan-Vese input mask

The gray-level selected component and the initial mask, which specifies the initial state of the active contour, are the inputs of the deformation process. The number of iterations until the snake reaches its final state is defined to be 100, which is also the default number of iterations. The algorithm stops the evolution of the curve if the contour position in the current iteration is the same as one of the contour positions from the most recent five iterations, or the maximum number of iterations is reached, as described in [29].

When the deformation process is finished, the final mask is directly obtained, with the lesion region inside the final position of the contour and the non-lesional skin outside of it.

3.4.2.3.3 Alternatives of design

The implementation of **GVF snakes** as a segmentation technique offers a wide variety of alternatives of design. First of them is the colour component that is selected to define the edge map of the image and, consequently, to compute the GVF field. The luminance image and different colour components were visually analysed and several properties were studied in order to select the most appropriate one.

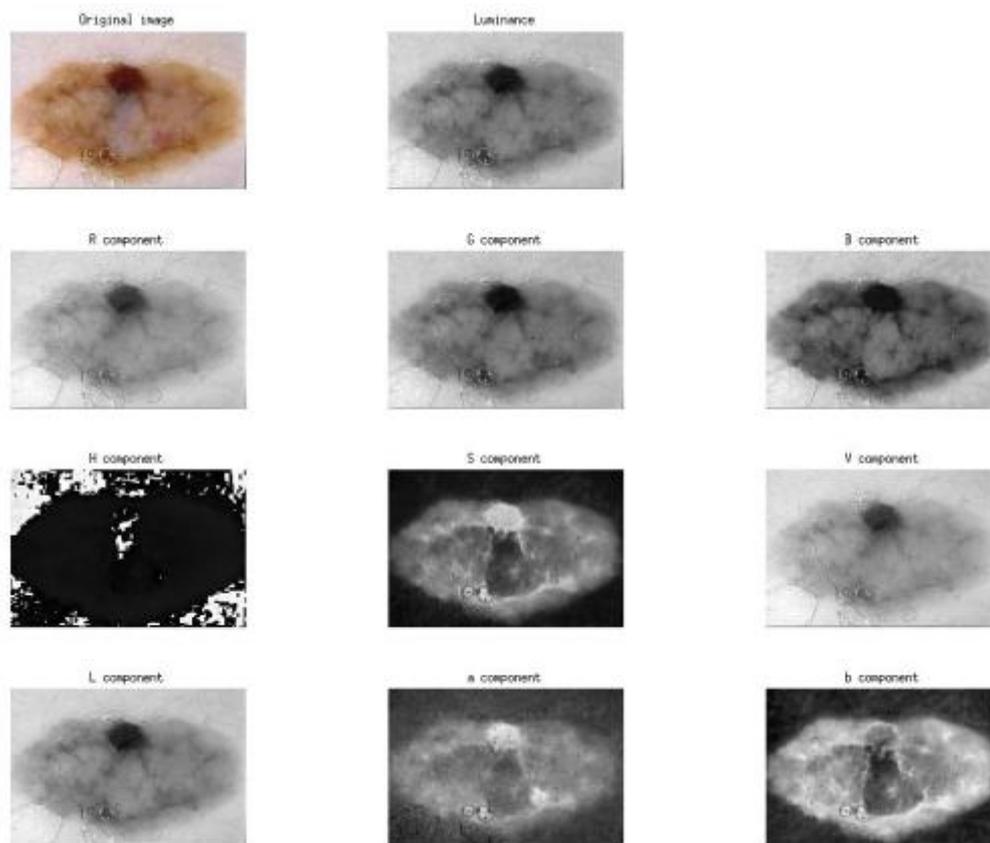


Figure 45: GVF snakes. Colour components analysis to define edge map.

As it may be observed in the previous figure, the component S from HSV colour space and the component B from RGB colour space show a more prominent lesion and not distorted boundaries information. In addition, several studies of the entropy of colour components included in the literature, as mentioned in [57], select the B channel because it provides the highest entropy value, as well as a good discrimination between lesion and skin. Thus, the B component was finally selected.

Another alternative of design is related to the deformation process of the curve. Interpolation between points in the curve with a large value as minimum distance may facilitate the computation, since dermoscopic images are large and have high resolution. But there is a limit for the minimum distance value and if it is increased, the curve may become distorted due to lack of information. In order to avoid interpolation problems and save computation time and cost, the interpolation between points within the curve was discarded and the number of points curve was first reduced and then maintained along the deformation; they only update their coordinates in each iteration.

A last alternative of design of GVF snakes is the initialization curve. Several options for initialization were proposed. The simplest one, which was included in the original algorithm, is a circumference located in the center of the image. It may be the center of mass of certain lesions and help converge the curve on close boundaries, but there are many irregular-shaped lesions and lesions that are not centered in the image along the database. Therefore, it was discarded.

Other options for initialization are based on previous segmentation techniques in order to generate the initial curve from the obtained binary mask. One option is the ellipse containing the lesion region, which may be obtained by combining inside the parametric equation of an ellipse the major and minor axis of the region and other parameters such as the orientation and its centroid. The convex polygon that contains the lesion and whose boundaries are close to the real ones was also considered. However, since a more accurate segmentation was achieved with the multi-region k-means adaptation of the clustering algorithm, it was decided to generate the initial curves directly from the lesion region established on this previous segmentation plus a small radial expansion of the perimeter. Thus, the aim of the GVF algorithm becomes the refinement of that segmentation by adjusting the contour to boundaries and also concavities.

Regarding the **Chan-Vese method**, initialization is an alternative of design too, since several options were contemplated until it was also decided to use as initial the curve the boundaries of the mask obtained by the multi-region k-means segmentation plus a small radial expansion. In order to save computation time and cost while maintaining the shape of the initial curve, a selection of equidistant points within the curve is carried out, so that the curve is simplified before entering the deformation process.

Lastly, there is one parameter related to the deformation process and the creation of the final contour in Chan-Vese method, which is included in [29]. It is the degree of smoothness or regularity of the boundaries of the segmented regions. Higher values produce smoother region boundaries, but can also smooth finer details. Lower values produce more irregularities in boundaries but allow details to be captured. The default smooth factor, which is a low value, has been selected in this system, since the level of detail at boundaries may vary from one lesion to another. If the contour is irregular and the lesion does not need that detail, then it can be smoothed in the post-processing phase.

3.4.2.4 Segmentation based on the bag-of-words model

As previously described in Section 2.3.5, the extraction of local features combined with the bag-of-words model supposes powerful means for pattern recognition and image classification. Due to their proper functioning in these two fields, it was decided to introduce as part of the system an experimental segmentation method based on these two techniques. It is considered to be experimental because, to our knowledge, the specialised literature does not include applications of the extraction of local descriptors and bag-words model to image segmentation. Therefore, the conventional process was followed and adapted to the segmentation of dermoscopic images.

The detection of relevant local features is performed on a set of reference images for the posterior generation of a visual vocabulary that allows the recognition of patterns and their classification. When adapted to the segmentation of dermoscopic images, the classification is reduced to two classes: lesion and skin. Thus, the visual codebook contains words for these two classes in order to achieve the segmentation of an image that enters the system by

matching every keypoints extracted from the image and the closest word inside the vocabulary, which corresponds to lesion region or non-lesional skin.

In the design of the system, the set of reference images used for the generation of the visual vocabulary contains half the images and their manual masks within the database, randomly selected each time a new vocabulary is created. The randomness of the process allows the images used as a reference to create the visual vocabulary to change dynamically in order not to bias the results towards a constant reference set. Simultaneously, the keypoints extracted from every reference set are also randomly selected and introduced in a subset of keypoints (the one used to generate the visual codebook), so that over-adjustments of the images that belong to the reference set are avoided at the matching procedure. It is considered that the amount of images and extracted keypoints is sufficient in order to obtain reference sets that are not significant of a large part of the database when the model is applied.

On the other hand, the final masks are generated for every image of the database, matching its descriptors and the words contained in the previously created visual codebook.

The entire procedure may be divided in the following steps:

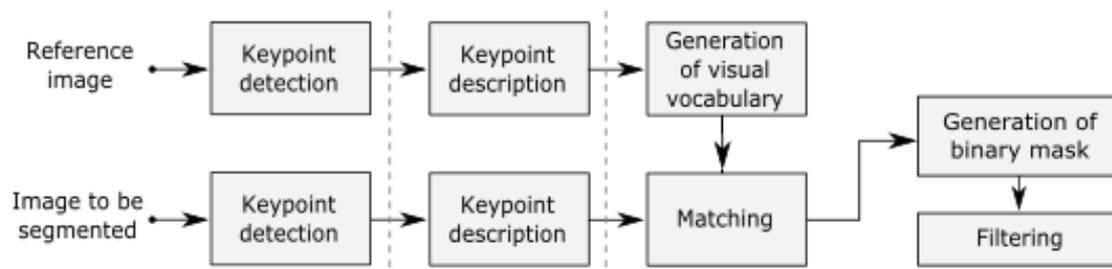


Diagram 14: Segmentation method based on the bag-of-words model.

There is an additional approach for the matching phase and, therefore, the generation of the final mask, that has been implemented for this system: a superpixel-level approach. All the previous phases inside the procedure are common to this approach.

3.4.2.4.1 Detection of keypoints

Two different techniques are proposed for the extraction of keypoints from the set of reference images. The input for both detectors is the preprocessed luminance image.

On one hand, the Harris Laplace detector was selected due to the visual relevance of the extracted keypoints and the density relation of keypoints between lesion and skin. Since the lesion areas are supposed to contain more variations associated with different colours, patterns and textures, it is desirable to capture them as accurately as possible, whereas the skin areas are considered to be more homogeneous. Although the total number of detected keypoints may change from one image to another when the Harris Laplace detector is applied, a higher amount of detected keypoints inside lesion areas means that the mentioned inner variations are taken into account at the posterior generation of the vocabulary.

On the other hand, it was also decided to carry out a dense sampling for keypoints detection. A grid is placed over the entire image and the keypoints are extracted along the grid, equally

distant. The vertical and horizontal step size for detection was set to 6 pixels. This way, the extraction of relevant features does not depend on the performance of the detector, which may be usually affected by artefacts and noise within the image even a previous smoothing filtering has been applied.

Once the keypoints have been extracted from a reference image, they are divided into two groups to be described separately afterwards: lesion keypoints and skin keypoints, regarding the regions inside each manual mask. Since the dense sampling of keypoints results in an enormous number of keypoints per reference image, a reduction is carried out, by keeping a randomly selected set of 2000 lesion keypoints and 2000 skin keypoints.

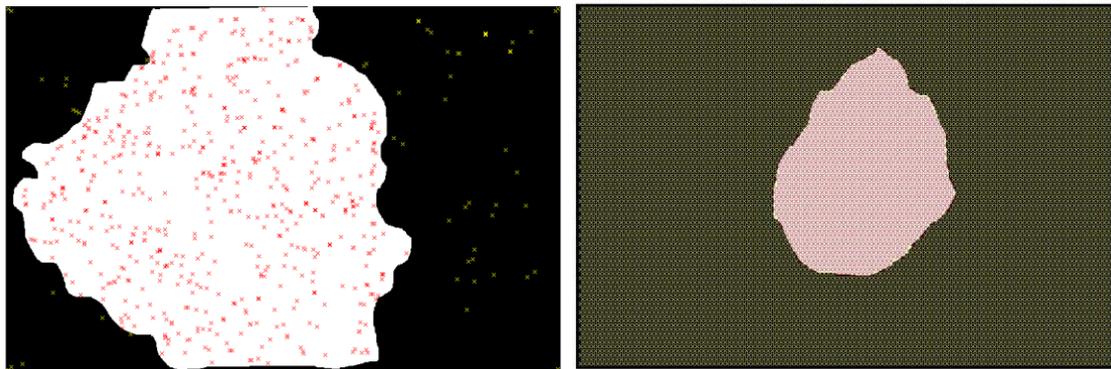


Figure 46: Extraction of keypoints. From left to right: manual mask with Harris Laplace detected keypoints, manual mask with dense sampling detected keypoints. Lesion keypoints are shown on red and detected skin keypoints on yellow.

Regarding the extraction of keypoints from the images that enter the system to be segmented, there is only one option for detection: dense sampling, since it is necessary to create grid of keypoints that provide uniformly distributed information and allow preserving the relation and properties of the original image in order to generate a binary mask afterwards. The vertical and horizontal step size for detection was also set to 6 pixels.

3.4.2.4.2 Description of keypoints

The next step consists in the description of the extracted keypoints. Two different descriptors were selected for implementation: SIFT, which works with the luminance image as input, and OpponentSIFT, a Colour-SIFT descriptor based on the channels of the opponent colour space that works with the *RGB* image. Regardless of the selected descriptor, the same one must be applied for both the detected keypoints from the set of reference images to create the visual vocabulary and the detected keypoints in each image to be segmented.

The next figure shows a random selection of fifty keypoints extracted by the Harris Laplace detector and described by the SIFT descriptor. As mentioned in Section 2.3.5.2.1, SIFT descriptor allows for significant shift in gradient positions around the keypoint. The generated orientation histograms can be observed over the image. Each orientation histogram shows eight directions, with the length of each arrow corresponding to the magnitude of that histogram entry.

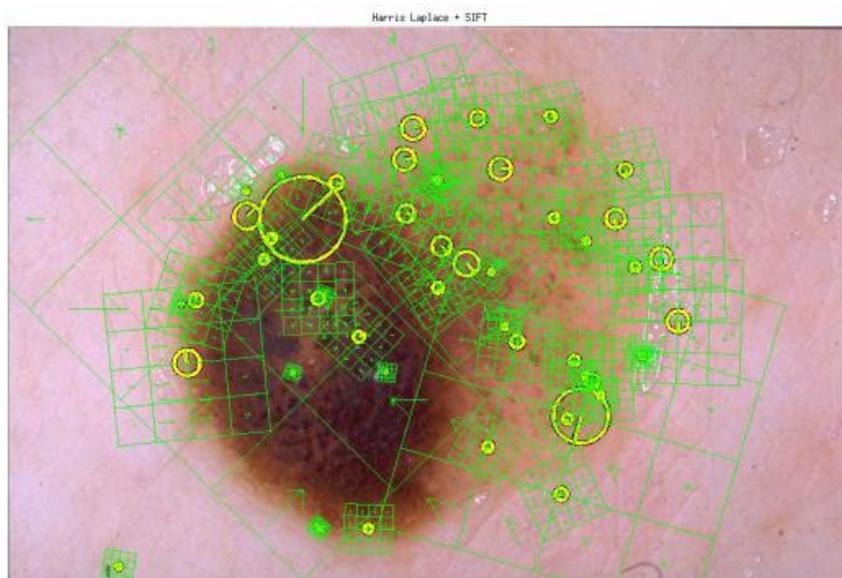


Figure 47: Description of keypoints. Random selection of 50 keypoints detected by Harris Laplace detector and described by SIFT descriptor.

3.4.2.4.3 Bag-of-words model: generation of visual vocabulary

Once the lesion keypoints and skin keypoints extracted from the set of reference images have been described separately, the bag-of-words model is applied. Two different visual vocabularies are generated: one to define lesion regions and one to define skin regions.

The lesion and skin codebooks generated by the system follow a data-driven approach by using k-means clustering. The desired number of words in the vocabulary (n) is introduced, along with the number of times the clustering is applied, which is set to 5 in order to obtain greater stability and consistency of the resultant vocabulary. As previously studied, k-means algorithm also requires the initialization of the cluster centroids. Each centroid is equivalent to a visual word and the keypoints are assigned the closest one in each iteration of the algorithm. Among the described keypoints, n keypoints are randomly selected to be the initial centroids of the k-means algorithm.

A vector quantization on the space of the extracted features is then carried out. K-means algorithm aims to find those cluster centroids that minimize the distance between data points and the nearest centroid. When the algorithm converges, a bag-of-words or visual vocabulary is obtained. It contains n words which will be used to match the keypoints of every image to be segmented.

Once the visual vocabulary of n lesion words and the visual vocabulary of n skin words are generated, they are concatenated into a single vocabulary of size $2n$.

Visual vocabularies containing a different number of words have been implemented in order to compare the influence of the vocabulary size on the classification of lesion and skin areas.

3.4.2.4.4 Matching of keypoints and visual words

The standard matching approach is based on the grid of detected and described keypoints from every image that enters the system to be segmented. The generated visual codebook, which contains words that represent both lesion and skin areas, is used to analyse each entering image, compare its patch features with the visual words inside the vocabulary and assign them to the closest one, i.e., the centroid that minimizes the Euclidean distance.

The applied function computes the Euclidean distance from each keypoint to each codeword and is supposed to reject those matches that are too ambiguous. There exists a threshold parameter for matching uniqueness, which was set to 0.5 and measures the uniqueness of a pair, as described in *VLFeat* library documentation for the function. The descriptor D1 is matched to a descriptor D2 only if the distance between the two descriptors multiplied by the threshold is not greater than the distance between D1 and all other descriptors.

After each keypoint is assigned to a lesion or skin word, it is necessary to generate a binary mask based on the keypoints classification. They are assigned a label: one for lesion or zero for skin. Each label is located within a binary mask regarding the original image size and a nearest-neighbour interpolation. A filtering phase is then required in order to reduce the noise on skin areas and compact the lesion region. The selected smoothing filter is a large pillbox filter.

The next figure shows the output of the matching process using a vocabulary that contains 32 words for lesion regions and 32 words for skin regions. The skin contains numerous spurious lesion areas, but the boundaries of the lesion can be discerned. After the pillbox filter is applied, the spurious areas within the skin are reduced and the lesion is compacted. The postprocessing phase is especially necessary for the masks generated from this method, in order to fill holes inside the lesion area and eliminate possible remaining spurious area on the skin. Additionally, no expansion will be needed, since the resulting masks show already expanded lesion regions due to the nature of the matching and the filtering.

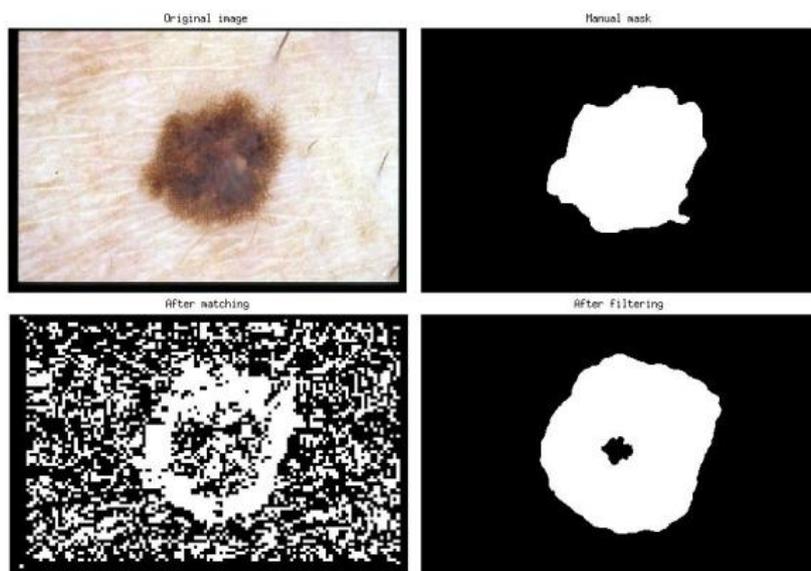


Figure 48: Matching of keypoints and visual words. On top: original image (left) and manual mask (right). On bottom: generated binary mask after matching (left) and mask after the pillbox filter is applied (right).

3.4.2.4.5 Superpixel-level approach

Once the visual vocabulary is obtained, another matching approach is proposed. It is based on superpixel extraction, the computation of one histogram per superpixel that contains the frequency of each word in the visual vocabulary, the generation of a set of histograms of reference subregions (which has been also referred to as “dictionary of histograms”) to represent lesion and skin areas, and the posterior intersection of each computed histogram with the histograms dictionary to carry out the superpixel matching.

The procedure may be divided into several steps:

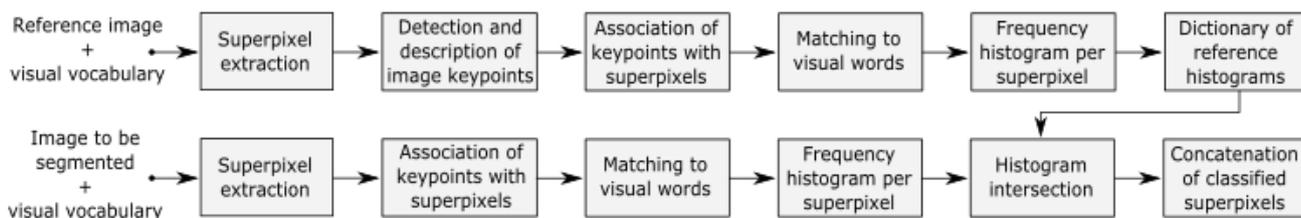


Diagram 15: Superpixel-level approach.

The first step is the generation of superpixels in both set of reference images and images to be segmented. The SLIC algorithm, described in Section 2.3.5.4.1, is the one applied for the extraction of superpixels. It works in *RGB* colour space and two parameters must be defined regarding the desired superpixels size and the regularity of their shape. An average of 50 regular-shaped superpixels is generated per image.

Since the set of reference images used to create the visual vocabulary is not desired to be the same as the set used now for the generation of the dictionary of histograms, another set of reference images is randomly created and the keypoints of these images are then detected and described. This way, the randomness of the process allows again the images used as a reference to create the “dictionary of histograms” to change dynamically in order not to bias the results towards a constant reference set.

The detector and the descriptor must be the same ones that were used to implement the visual vocabulary. Regarding the images within the database that are object of segmentation, the keypoints were already detected on a grid and described. The applied descriptor must be the same one that was used to create the visual vocabulary and the same that is used now to generate the histograms dictionary.

The manual masks from the set of reference images are taken as reference to separate the superpixels within a reference image into two groups: skin superpixels and lesion superpixels.

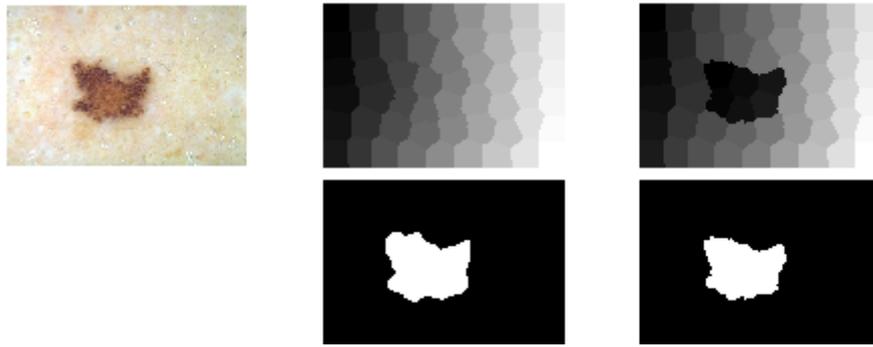


Figure 49: Generation of superpixels. On top: original reference image without black frames (left), generated superpixels (center), superpixel separation between lesion and skin (right). On bottom: manual mask (left), appearance of the mask generated from superpixels division (right).

All the keypoints within an image and the visual vocabulary are then matched, i.e., each keypoint is assigned to the closest visual word: lesion word or skin word.

Then each keypoint is associated with the superpixel in which is located regarding its coordinates. Each keypoint in the superpixel has been matched to a visual word. For each superpixel a histogram is computed. The histogram measures the frequency of each visual word within that superpixel. Therefore, if the employed visual vocabulary contains n words, the superpixel histogram contains n bars. The frequency histogram related to each superpixel can be considered as the new feature vector that characterizes that superpixel.

The next step consists in the creation of the set of histograms of reference subregions, which can be interpreted as a new version of the visual vocabulary adapted to superpixel-level. It is generated from the lesion histograms and skin histograms extracted from the reference images. Since an average of 50 superpixels per image was extracted, there is also an average of 50 frequency histograms per image. Thus, a random selection of a reduced number of histograms is executed, equally to lesion and skin histograms, so as to obtain a medium-sized dictionary of reference histograms and avoid over-adjustments of the images that belong to the reference set are avoided at the matching procedure. It is considered that the amount of extracted subregions and keypoints is sufficient in order to obtain reference sets that are not significant of a large part of the database when the model is applied.

Lesion and skin selected histograms are then concatenated to constitute the set of histograms of reference subregions. “Dictionaries of histograms” of different sizes, i.e., containing different amounts of reference histograms, have been implemented in order to analyse the influence of size on the segmentation.

Once the “dictionary of histograms” is obtained, each superpixel from an image to be segmented can be classified. The histogram intersection method, described in Section 2.3.5.4.2, is applied to measure the similarity between histograms, in order to assign to each superpixel of the image to be segmented one superpixel from the histograms dictionary, which represents lesion or skin region. Additionally, the assignment is modified to be softer, i.e., not to take only into account the most similar histogram within the collection, but the five most similar ones. If the majority of the five closest histograms are lesion histograms, the superpixel is classified as lesion area. The same in the case the majority are skin histograms.

Once all the superpixels of the image to be segmented are classified as lesion or skin, they are concatenated to generate an assessable binary mask.

3.4.2.4.6 Alternatives of design

Several alternatives of design are present along the development of a method based on the extraction of local features and bag-of-words model.

Regarding the **detection of keypoints**, different detectors are based on particular techniques to extract relevant features. The Harris Laplace and Harris multiscale detectors are the optimal ones regarding the density of points extracted on lesion over skin regions. In the literature, the Harris Laplace detector, which simultaneously detects corners and blobs, is popular among classification applications. Therefore, this detector was selected over other detectors.

The number of detected features depends on several parameters, as described in the *VLFeat* Documentation of the detection function. The peak threshold sets the minimum absolute value of the cornerness measure to accept a feature; a larger threshold results in fewer extracted features. Features are also filtered by the curvature of the cornerness measure, which is controlled by the edge threshold. This threshold allows the elimination of those keypoints whose curvature is too small and yield badly localized frames; a larger threshold causes more features. Additionally, the Harris Laplace detector, which uses peaks in the response of the multi-scale Laplace operator to select the scale of the detected frames, allows further adjustment of these peaks with the Laplacian peak threshold. The values of the involved thresholds were conveniently adjusted in order to extract an adequate number of representative features per image, avoiding the detection of keypoints on image external artefacts and non-relevant areas to a great extent.

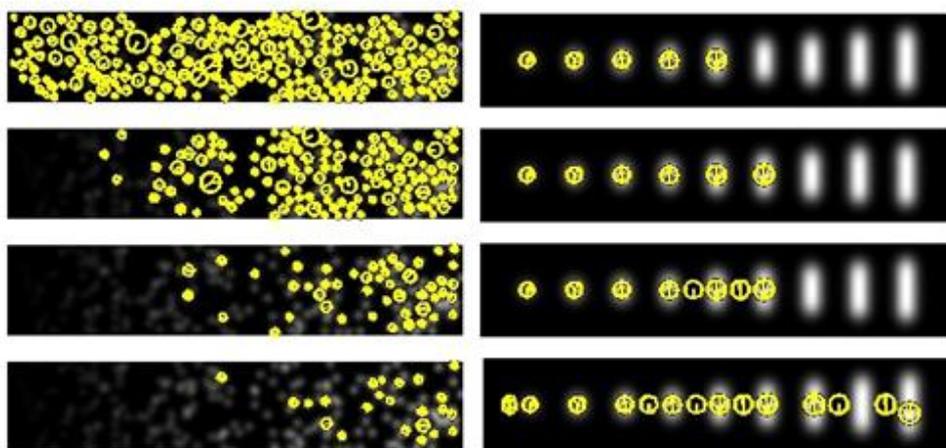


Figure 50: Detection of keypoints. From left to right: detected frames increasing from top to bottom the peak threshold (0, 10, 20, 30), detected frames increasing the edge threshold from top to bottom (3.5, 5, 7.5, 10). [XVII]

When dense sampling is applied for the extraction of equidistant features, the size of the grid can be selected. The larger the distance between adjacent features, the less uniform the obtained information and, therefore, the larger and the less exact the interpolation that is necessary to generate the binary mask afterwards, once the matching is performed.

The **description of keypoints** offers several alternatives of design too, when it comes to select the applied descriptor. The original SIFT descriptor was selected over other proposed descriptors because it is one of the most popular keypoint descriptors due to its invariance to rotation, scale and illumination. Due to its popularity, there are numerous publications in the specialised literature, which allowed an exhaustive analysis of its functioning in order to adapt the procedure to dermoscopic image segmentation. Additionally, colour descriptors, especially those based on SIFT, were also proposed to be included as part of the experimental method. Among the colour-SIFT descriptors studied in Section 2.3.5.2.2, the OpponentSIFT descriptor was selected because when choosing a single descriptor and little prior knowledge of the response of the data set to colour description is available, it is the one recommended, as concluded in [37].

In relation to the application of the **bag-of-words model**, the main alternative of design is related to the size of the generated visual vocabulary. Large vocabularies, i.e., 4000 words of each class, were proposed, but they imply higher computational time and less selective relevant keypoints. Therefore, small (32 lesion words and 32 skin words) and medium-sized (200 lesion words and 200 skin words) vocabularies were finally considered, as also proposed in [33]. Additionally, the relation between the number of lesion and skin words contained within the codebook may be modified. Firstly, a larger number of lesion words was considered, due the numerous variations that these areas present. However, the skin regions suppose larger areas in most dermoscopic images; therefore, they must be also well represented. Thus, the number of lesion and skin words to describe the keypoints of entering images was set to be equal.

Regarding the **matching of keypoints and visual words**, the threshold that controls the uniqueness of a pair of descriptors may be altered in order to increase or decrease matching ambiguity. Due to the enormous variations within features in dermoscopic images, it is hard to assign each keypoint one out of two classes. Therefore, the threshold was set to a smaller value than the default value, in order to obtain a match for every descriptor of the image.

At **superpixel-level**, the assignment technique itself offers several alternatives of design. After matching by histogram intersection, a softer assignment was imposed over a hard one in order to assign each superpixel a lesion or skin classification.

The performance of superpixel-level approach is also influenced by the parameters that govern the SLIC algorithm, which are described in Section 2.3.5.4.1. The size of the generated superpixels is adapted so as to obtain neither an excessive nor a small number of superpixels. The size must be adequate to represent anatomical structures; therefore, too many or too few generated superpixels may not be representative enough. Besides, an excessive number of superpixels may result in high computational load. The value for superpixels size within the system is set in order to generate an average of 50 superpixels per image. Additionally, the shape of superpixels is also of paramount importance to represent structures correctly. The strength of spatial regularization is controlled by the regularizer parameter, which is set to 1. This value allows certain irregularities in shape for a better adaptation, as well as not excessive regularity, so as not to create a grid on the image.



Figure 51: Variations in superpixel shape regarding the value of the regularizer parameter [XVIII].

3.4.3 Postprocessing

The output of the automatic segmentation phase is a binary mask that separates the lesion area from the skin regions. Due to spurious lesion or skin areas within the mask, as well as the presence of irregularities in the lesion boundaries, the mask needs to be postprocessed in order to obtain an analysable mask, which allows evaluation of the system and the posterior comparison between applied segmentation methods. As previously studied in Section 2.5, there are several techniques to address the refinement of the final binary mask.

The main targets of the postprocessing step within this system are the removal of islands (small isolated regions tagged as lesion in the label image) and holes (small regions inside the main lesion region that have been tagged as skin), the smoothing of lesion borders and the expansion of the lesion area. The applied procedures for these three targets are common to all implemented segmentation methods from the previous section.

Additionally, some implemented methods based on the obtaining of more than two regions within the image, such as multithresholding or clustering-based techniques, have included previously a region merging process in order to merge all detected lesion regions into one while excluding the skin region, whose description is included in the corresponding section to each method.

Thus, the postprocessing phase may be divided in the following steps:

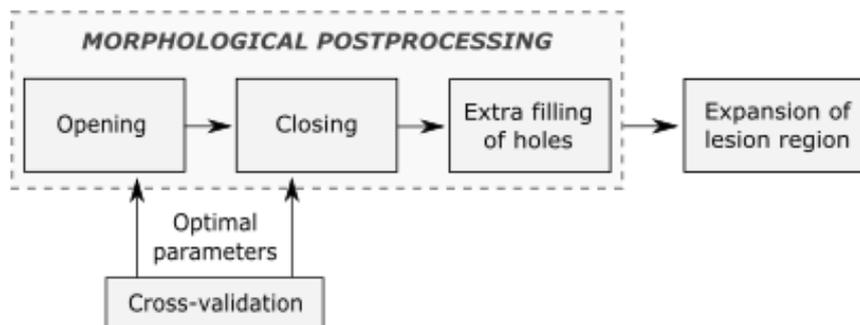


Diagram 16: Postprocessing.

3.4.3.1 Morphological postprocessing

The first step of the postprocessing procedure is based on three concatenated morphological operations, whose aim is to eliminate spurious lesion and skin areas and to smooth lesion border irregularities.

The first process, opening, is the combination of applying an erosion operation followed by a dilation operation, both with the same structural element. Its objective is to smooth the contour of the lesion, break narrow connectors and eliminate thin protuberances. On the other hand, the second process, closing, is the combination of applying first a dilation operator and then an erosion operator, both with the same structural element. This process also tends to smooth the lesion contour but, as opposed to opening, it generally fuses narrow breaks and long thin gulfs, eliminates small holes and fills gaps in the contour.

As it may be observed, the shape and size of the two structural elements used for opening and closing operations, respectively, must be determined. It is desirable to find the optimal values that allow the most accurate removal of islands and holes within the mask, as well as the most suitable borders smoothing, in order to obtain an automatic mask as similar as possible to the manual mask, considered as ground truth. A cross-validation procedure, whose implementation is described in Appendix B, has been applied for the validation of morphological parameters in order to determine the optimal combination.

The selected shape of the structural elements employed for opening and closing, respectively, is a disk. The radius of the disk is the parameter to be validated.

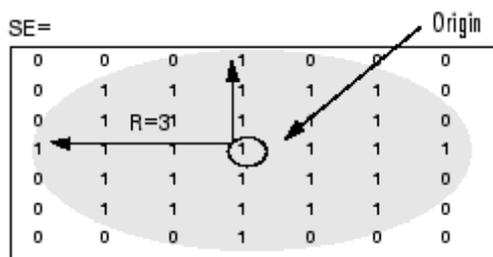


Figure 52: Disk-shaped structural element [XIX].

The radius values are validated within a range of 9 pixels to 42 pixels, increasing 3 pixels each time the value changes. In each iteration, the morphological opening and closing operations are applied to each automatic binary mask within a train set.

The accuracy of the automatic mask with different values of the radiuses of the opening and closing structural elements is computed for five different groups of randomly selected images within the database.

The next figure shows the accuracy evolution of the masks within a train set regarding the modification of the morphological structuring elements. The employed automatic masks are the output of the two-region k-means segmentation based on pixel radial modulus and H and S components. As it may be observed, the peak of the 3-D graphic corresponds to an opening radius of 9 and closing radius of 36. These are the most repeated obtained values, due to the

good achievement at the removal of spurious areas and the generation of lesion borders that do not contain excessive details, as the lesion regions defined in the manual masks.

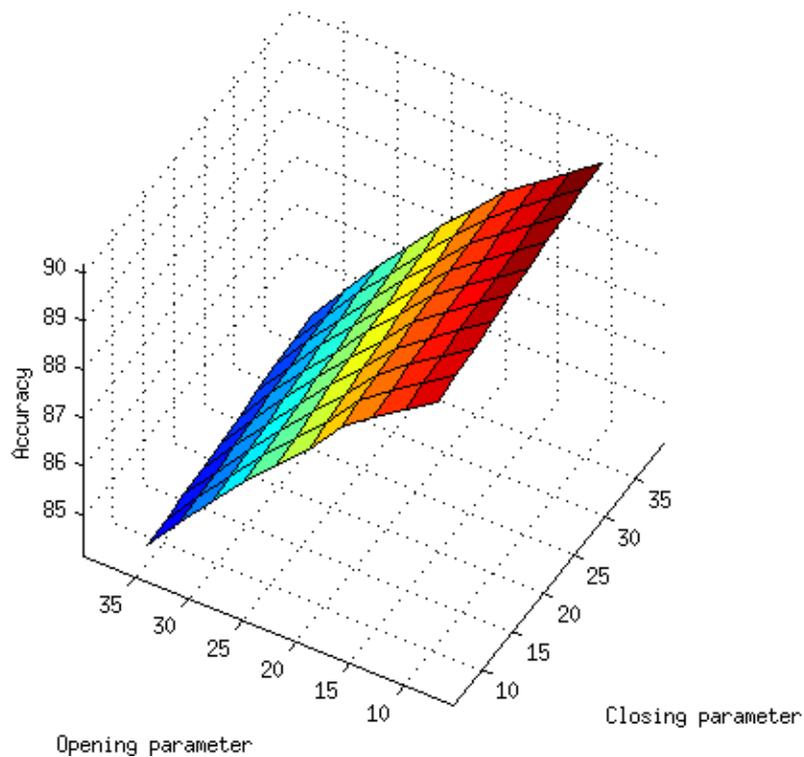


Figure 53: Morphological parameters validation. Example of accuracy evolution regarding the modification of the radius of opening and closing structural elements.

Once the opening and closing operations are applied with the validated morphological parameters, an additional processing in order to guarantee the automatic filling of holes within the lesion region is carried out. As described in [58], a hole is considered a set of background pixels that cannot be reached by filling in the background from the edge of the image. The process is based on a flood-fill operation with no defined seeds and automatic detected holes, since the process is desired to be completely automatic.

3.4.3.2 Expansion of lesion region

As previously mentioned in Section 2.5, most automatic segmentation methods generate lesion regions whose borders are contained within the lesion areas determined by the manual masks. In order to reduce the discrepancy between the generated automatic masks and the ground truth, an expansion of the computer-detected lesion boundary is implemented as a last postprocessing step within the system.

Firstly, the coordinates of the points that constitute the lesion boundaries are obtained, as well as the center of the mass or centroid of the lesion region. The lesion border points are then expanded along the straight lines that join the centroid and each point. Their position moves in the normal direction, from the lesion border toward the image edge. The expansion distance is

previously determined. Its value was set to 20 pixels, which is a constant value empirically optimized that allows a sufficient and accurate lesion expansion.

Every point is expanded separately in an iterative process. In each iteration, the point is expanded by a small constant, d , which was set to 1 pixel, until the expansion distance is completed. Before the coordinates of the point are updated after each iteration, it is checked if the edges of the image have been reached. If so, the expansion is stopped and a new point enters the process. The whole expansion process can be mathematically defined as follows.

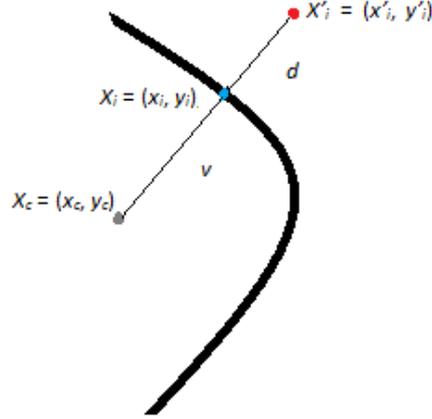


Figure 54: Expansion of lesion region.

Given the centroid of the lesion region $X_c = (x_c, y_c)$ and the point to be expanded $X_i = (x_i, y_i)$, the vector v from the centroid of the region and the point can be denoted as:

$$v = X_i - X_c = (x_i - x_c, y_i - y_c) = (v_x, v_y)$$

The normalization of v is given by:

$$n = \left(\frac{v_x}{|v|}, \frac{v_y}{|v|} \right) = (n_x, n_y)$$

The vector v' that connects the centroid of the region with the position of the point after an expansion of d pixels is denoted by:

$$v' = n \times (|v| + d)$$

Therefore, the position of the point after an expansion of d pixels, $X'_i = (x'_i, y'_i)$, can be computed as follows:

$$X'_i = (x'_i, y'_i) = X_c + v' = X_c + [n \times (|v| + d)] = f(X_c, X_i, d)$$

As it has been indicated in Section 3.4.2.4.4, those masks generated from the extraction of local features and matching with a visual vocabulary are already expanded due to the nature of the matching and the required filtering. These masks are the only ones which do not undergo the process of lesion region expansion.

4. System evaluation

4.1 Database

The database used for the analysis and comparison of all the algorithms included in our study consists in 724 dermoscopic images extracted from the frequently used EDRA Interactive Atlas of Dermoscopy [1]. These images contain several types of melanocytic lesions, benign as well as malignant, with a great variety in shape, size, colour and other external and intrinsic cutaneous features. Therefore, the level of segmentation difficulty changes from one image to another.

Every image within the database is paired with a manual mask that establishes the expected lesion segmentation. It is considered as ground truth and used in the evaluation of performance of every automatic mask generated by the system. Thus, the comparison between the implemented algorithms bases also its conclusions on the stored manual segmentations.

4.2 Evaluation measures

As in the case of general image segmentation procedures, there are two major evaluation methods for the segmentation of dermoscopic images: subjective and objective. The former involves the visual examination of the border detection results by one or more dermatologists, but this technique does not allow comparison among different automatic segmentation methods, since there is no objective measure of quality involved. On the other hand, the latter involves the quantification of automatic detected borders using dermatologist-determined borders as ground truth [2].

The evaluation of performance of the implemented automatic segmentation methods is objective and the selected quantitative evaluation measures are based on the concepts of true/false positive/negative, as shown in the following figure.

Definitions of True/False Positive/Negative

		Detected pixel	
		Lesion	Background
Actual	Lesion	True positive (TP)	False negative (FN)
Pixel	Background	False positive (FP)	True negative (TN)

Table 1: Evaluation concepts [XV].

If a lesion pixel is detected as part of the background skin, this pixel is considered to be a false negative (FN), whereas if it is correctly detected as lesion, the pixel is considered to be a true positive (TP). On the other hand, if a background pixel is detected as part of the lesion, it is considered as a false positive (FP), whereas it is correctly included inside the non-lesional skin, it is considered to be a true negative (TN) [50].

In this project several quantitative evaluation measures have been selected to study and compare the performance of the implemented algorithms.

❖ **Sensitivity and specificity**

Sensitivity, also referred as to TP rate or recall, and specificity, also named after TN rate, are commonly used in medical applications. In dermoscopic image segmentation, the former corresponds to the percentage of correctly detected lesion pixels, whereas the latter corresponds to the percentage of correctly detected background pixels.

$$Sensitivity = \frac{TP}{TP + FN} \times 100\%$$

$$Specificity = \frac{TN}{FP + TN} \times 100\%$$

Thus, sensitivity measures the avoiding of false negatives (all lesion pixels are identified as lesion) and specificity quantifies the avoiding of false positives (no skin pixels are identified as lesion). It may be observed that an automatic border that encloses the manual border will have a perfect sensitivity (100%). On the other hand, an automatic border that is completely enclosed by the corresponding manual border will have a perfect specificity. Therefore, it is crucial not to interpret these measures separately because there must be a trade-off between them [50].

Perhaps, taking into consideration the selected quantitative measures, the concept associated to sensitivity becomes especially important from a clinical point of view, since the main objective is the posterior study of all the points in the area of interest, i.e., the lesion region. The highest value of sensitivity for an automatic segmentation means that all the lesion pixels are supposed to be included inside the mask and no lesion areas are being discarded for analysis. Nevertheless, as it has been previously mentioned, the specificity value has also to be taken into account for a trade-off, because an excessive quantity of false positives inside the final mask may lead to altered results.

❖ **Accuracy**

The accuracy measure is equivalent to the proportion of true detections, both positive and negative, among the total number of detected pixels.

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \times 100\%$$

Since it mixes negative and positive true detections, it does not allow marginal interpretations. Therefore, when interpreted isolated, accuracy may lead to misleading

conclusions and must be combined with other quantitative measures, such as sensitivity and specificity, which provide marginal ratios.

❖ **Balanced accuracy**

Another useful performance measure is the balanced accuracy, which avoids inflated performance which may be a result of imbalanced datasets. It is defined as the arithmetic mean of sensitivity and specificity, i.e., the average accuracy obtained on either class.

$$\text{Balanced accuracy} = \frac{\text{Sensitivity} + \text{Specificity}}{2}$$

If the detection performs equally well on lesion areas and non-lesional areas, balanced accuracy is equivalent to the conventional accuracy value (correct detections divided by the total number of detections). On the other hand, if the conventional accuracy is above chance because the classifier takes advantage of an imbalanced dataset, then the balanced accuracy drops to chance [55].

4.3 Experiments and results

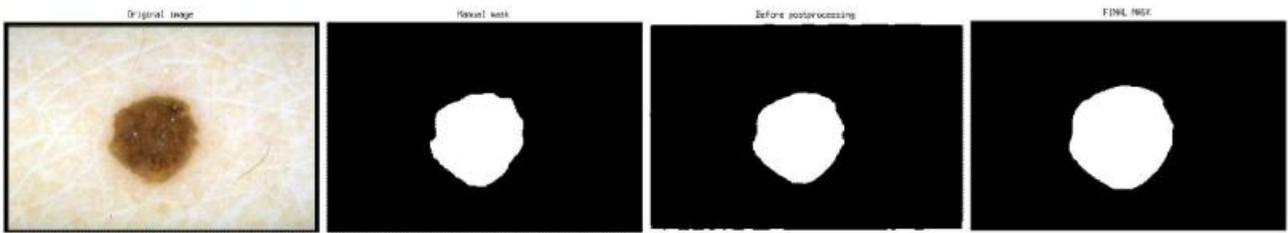
The database of 724 images was introduced in the system and processed using all the implemented segmentation methods. The segmentation results were analysed regarding the manual masks considered as ground truth. The performance of each method was evaluated after the postprocessing phase by the metrics described in the previous section: sensitivity (SE), specificity (SP), accuracy (AC) and balanced accuracy (BAC). The performance was measured after the postprocessing phase, separately for five cross-validated test sets.

In this section several examples of the generated automatic masks are shown for each implemented method, along with the correspondent original image, manual mask and several steps before the postprocessed result if required. For each one of the five cross-validated test sets, which have been described in Section 3.4.3, tables with the performance statistics are also included, which allow the comparison between methods that belong to the same category regarding the segmentation technique, as well as between methods based on different techniques.

4.3.1 Threshold-based methods

4.3.1.1 Traditional Otsu method

The traditional implementation of Otsu method divides the luminance histogram in two different classes. Therefore, compact lesions with high contrast between lesion and skin achieve good segmentation results (Ex. 1), even at the presence of odd skin textures (Ex. 2). Expanded lesions are also well isolated when the contrast is high (Ex. 3).



Example 1: Traditional Otsu method. Compact lesion. From left to right: original image, manual mask, mask before postprocessing, final mask.

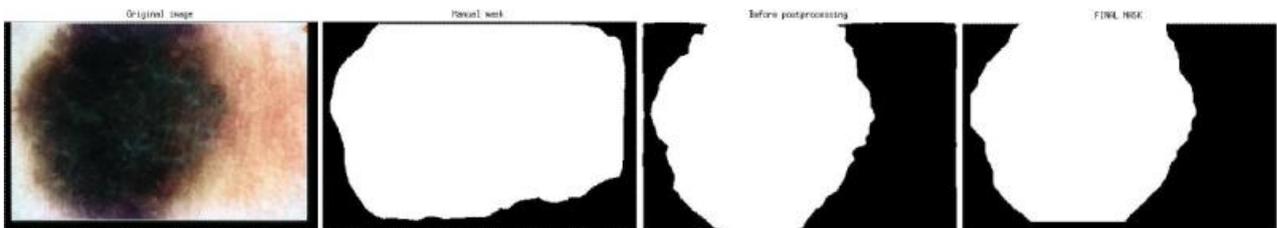


Example 2: Traditional Otsu method. Compact lesion with skin texture. From left to right: original image, manual mask, mask before postprocessing, final mask.

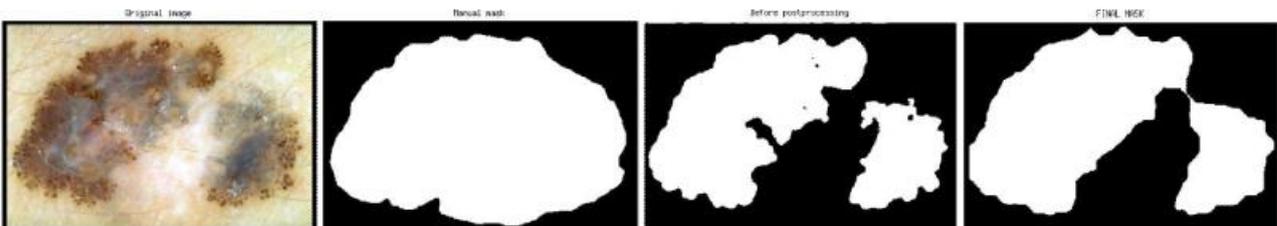


Example 3: Traditional Otsu method. Expanded lesion. From left to right: original image, manual mask, mask before postprocessing, final mask.

Due to its simple segmentation approach, in those colour-variegated lesions (Ex. 4) and those which contain either inner regression (Ex. 5) or blue-white veil areas, several zones are excluded from the lesion region.

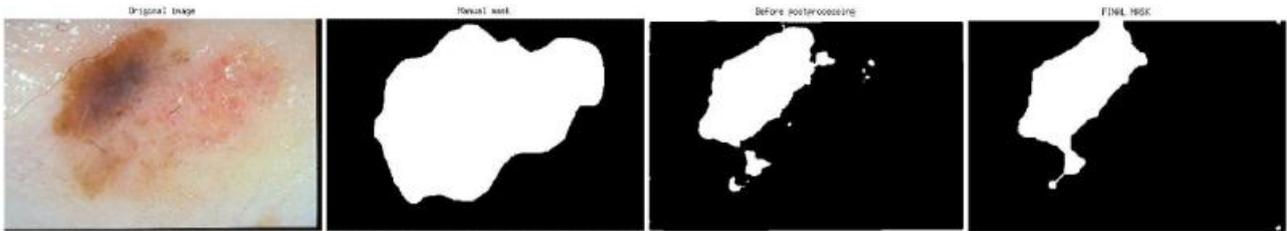


Example 4: Traditional Otsu method. Colour-variegated lesion. From left to right: original image, manual mask, mask before postprocessing, final mask.

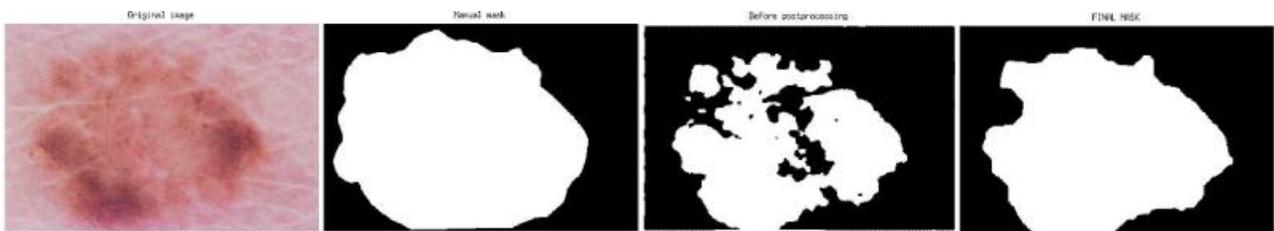


Example 5: Traditional Otsu method. Lesion with inner regression areas. From left to right: original image, manual mask, mask before postprocessing, final mask.

Additionally, low-pigmentation lesion types are rarely recognised as lesional skin, due to the low contrast between lesion and the background skin (Ex. 6). The skin must be even lighter in order to segment these lesions correctly (Ex. 7).

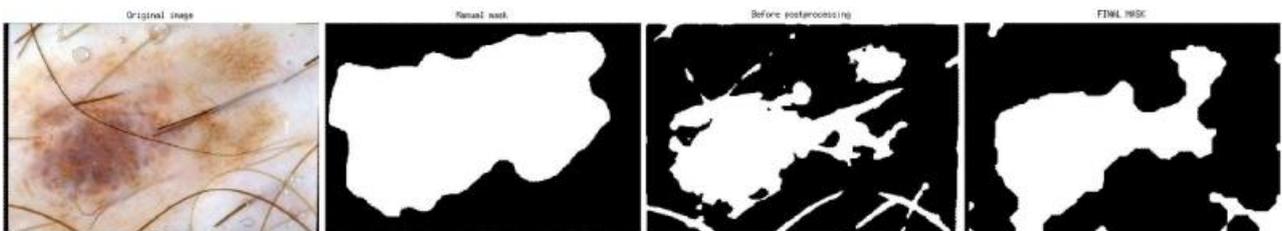


Example 6: Traditional Otsu method. Low-contrast between lesion and skin. From left to right: original image, manual mask, mask before postprocessing, final mask.

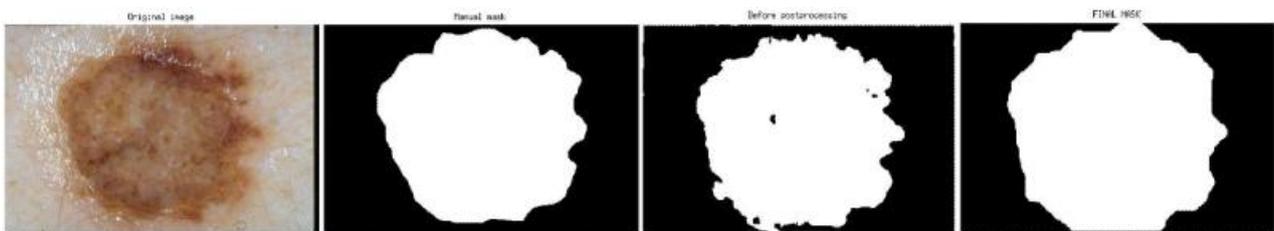


Example 7: Traditional Otsu method. Low contrast between lesion and skin. From left to right: original image, manual mask, mask before postprocessing, final mask.

The algorithm performs irregularly at the presence of external artefacts. External artefacts such as hairs cause inconsistency at its performance (Ex. 8). However, the algorithm is usually unaffected by other external artefacts such as traces of liquid (Ex. 9).

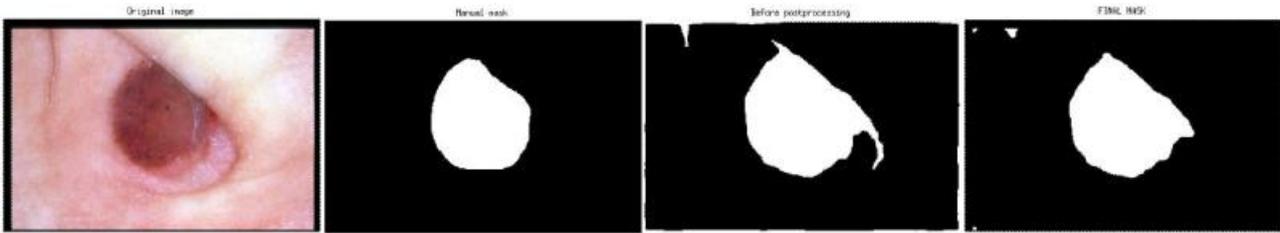


Example 8: Traditional Otsu method. Presence of hairs. From left to right: original image, manual mask, mask before postprocessing, final mask.



Example 9: Traditional Otsu method. Presence of traces of liquid. From left to right: original image, manual mask, mask before postprocessing, final mask.

Nevertheless, the segmentation of images with illumination inconsistency is better addressed (Ex. 10).



Example 10: Traditional Otsu method. Illumination inconsistency. From left to right: original image, manual mask, mask before postprocessing, final mask.

The next table shows the individual evaluation measures computed from the segmentation results of every test set and the global average for each evaluation metric. Noticeable differences between sensitivity and specificity values can be observed. A high specificity means that a low percentage of healthy areas have been identified as lesional, but many lesional areas are excluded from the lesion region too due to the mentioned aspects, which is reflected by the obtained low values of sensitivity.

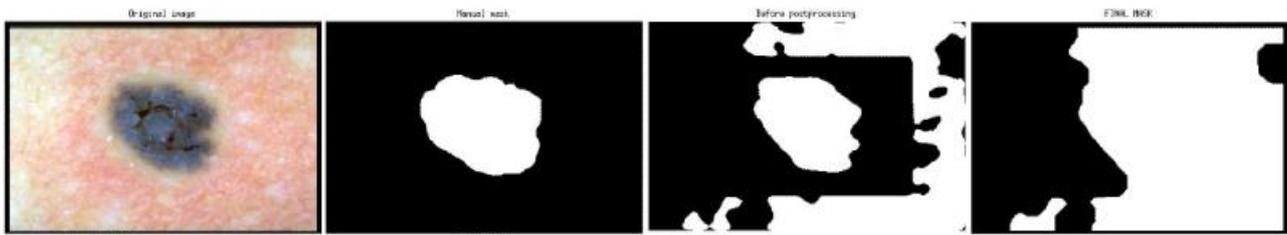
	TRADITIONAL OTSU			
	SE (%)	SP (%)	AC (%)	BAC (%)
Test 1	82,20	97,45	91,77	89,82
Test 2	77,16	97,30	89,06	87,23
Test 3	80,08	97,35	91,43	88,71
Test 4	80,26	97,29	91,06	88,78
Test 5	78,52	97,09	89,79	87,81
AVERAGE	79,64	97,30	90,62	88,47

Table 2: Evaluation of traditional Otsu method.

4.3.1.2 Local Otsu method

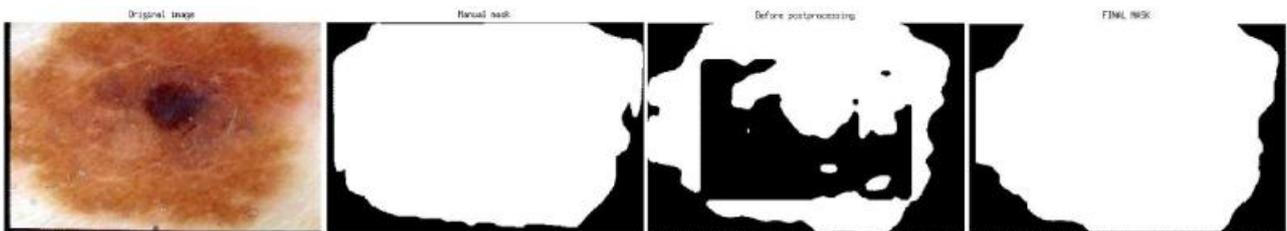
As mentioned in Section 3.4.2.1.2, the local application of the Otsu method on rectangular concentric partitions was selected over the grid approach, since the smoothing of discontinuities between partitions was facilitated by the number of image divisions and their distribution. Each image is divided in three concentric rectangular areas and the Otsu method is applied on each one separately. As a result, there is a binarization of the external, middle and inner areas, which are then concatenated.

The local implementation of the Otsu method does not show appreciable improvements regarding the performance of the traditional approach. Segmentation problems related to low contrast between lesion and skin areas, as well as the presence of cutaneous features such as regression or external features such as hairs, have not been solved. Besides, additional segmentation irregularities that were not present before have emerged for those images with strong skin patterns and illumination inconsistency. Changes of illumination and skin texture within an image cause unstable thresholds, mainly in the external partition (Ex. 11).



Example 11: Local Otsu method. Unstable thresholds causes by skin patterns. From left to right: original image, manual mask, mask before postprocessing, final mask.

Only for those expanded, colour-variegated lesions the results may have improved. Three different thresholds may identify different areas within a lesion and the postprocessing phase helps the closing of the lesion region (Ex. 12). Nevertheless, since the partitions are concentric, no improvements are achieved when the lesion is not centered in the image.



Example 12: Local Otsu method. Colour-variegated, expanded lesion. From left to right: original image, manual mask, mask before postprocessing, final mask.

The following table includes the performance statistics from five test sets of images. The trade-off between sensitivity and specificity changes with respect to the traditional Otsu approach. The increment in sensitivity may be due to the inclusion of colour-variegated areas inside the lesion region that were not included before. However, the decrement of specificity is related to the unstable thresholds, which separate between lesion and skin even those areas which belong totally to non-lesional skin, for example, in images of compact lesions surrounded by textured skin.

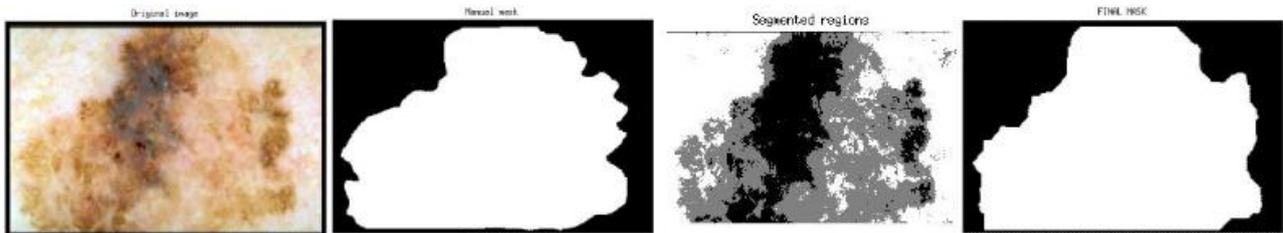
	LOCAL OTSU (CONCENTRIC PARTITIONS)			
	SE (%)	SP (%)	AC (%)	BAC (%)
Test 1	82,74	94,47	89,22	88,61
Test 2	81,72	92,73	87,58	87,23
Test 3	80,27	95,64	88,83	87,96
Test 4	83,01	93,09	88,30	88,05
Test 5	82,09	95,19	89,20	88,64
AVERAGE	81,97	94,23	88,63	88,10

Table 3: Evaluation of local Otsu method

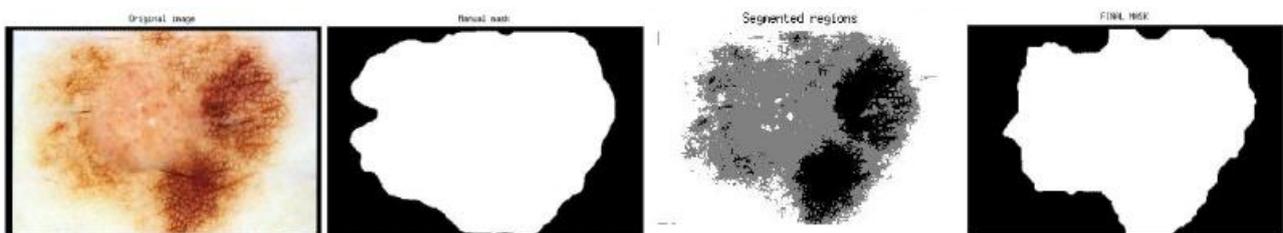
4.3.1.3 Multithresholding Otsu method

The adaptation of the traditional Otsu method to the computation of several global, optimal thresholds shows diverse improvements with respect to the other Otsu approaches. The previous determination of a suitable number of regions within the image helps the identification of different colour-variegated or textured lesion areas, which were not detected

correctly before by a direct binarization of the image. Therefore, the performance of the multithresholding approach surpasses the previous results for both expanded (Ex. 13) and colour-variegated lesions (Ex. 14).

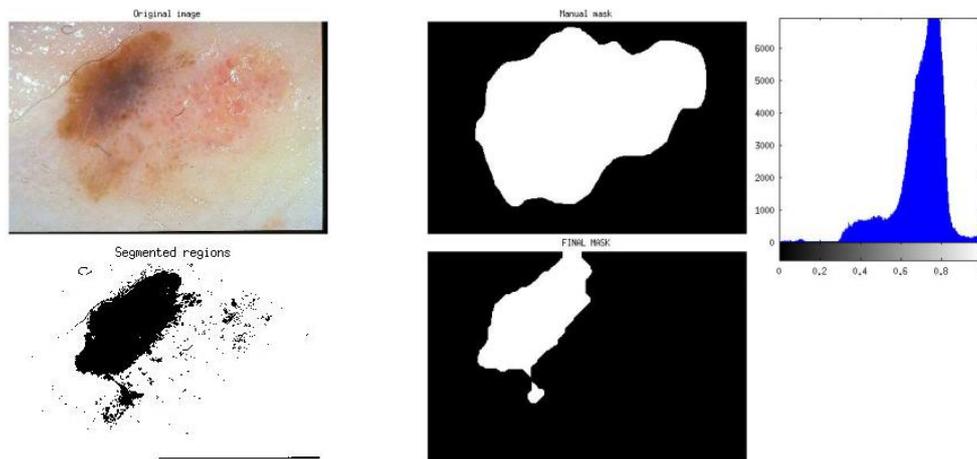


Example 13: Multithresholding Otsu method. Expanded lesion. From left to right: original image, manual mask, segmented regions, final mask.



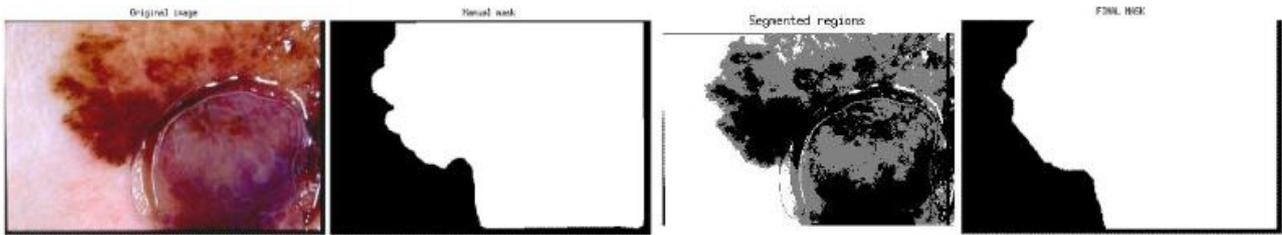
Example 14: Multithresholding Otsu method. Colour-variegated lesion. From left to right: original image, manual mask, segmented regions, final mask.

Nevertheless, the problem for images with low contrast between lesion and skin is still present. If only two regions are detected due to the low contrast and lack of local maxima in the luminance histogram, then the algorithm applies a direct binarization and those lesional areas with low pigmentation are not included inside the lesion region (Ex. 15).



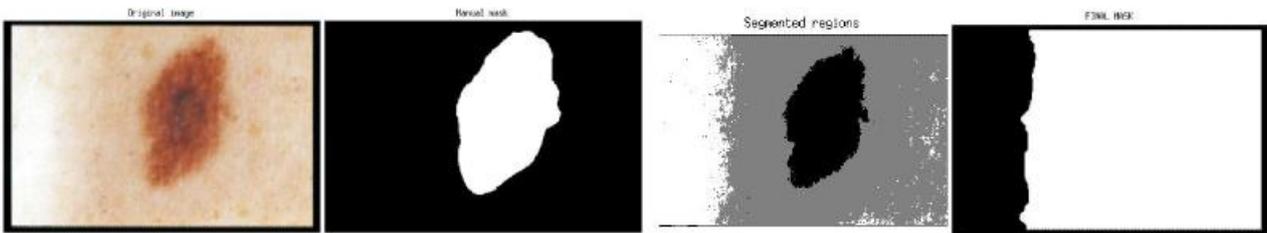
Example 15: Multithresholding Otsu method. Low contrast between lesion and skin. On top: original image (left), manual mask (center) luminance histogram (right). On bottom: segmented regions (left), final mask (right).

Regarding the performance at the presence of artefacts, this adaptation of the algorithm is largely resistant to cutaneous features, such as blue-white veil, and external features, such as traces of liquid (Ex. 16).



Example 16: Multithresholding Otsu method. Presence of liquid traces. From left to right: original image, manual mask, segmented regions, final mask.

The main handicap of this algorithm, as a threshold-based technique, is the lack of spatial information when the optimal thresholds are computed. Therefore, the segmentation of those images which present illumination inconsistency may be still unstable (Ex. 17).



Example 17: Multithresholding Otsu method. Illumination inconsistency. From left to right: original image, manual mask, segmented regions, final mask.

The next table shows the evaluation metrics obtained from the application of the multithresholding Otsu method in the segmentation phase to five test sets of images. As it may be observed, the global percentages of sensitivity and accuracy are the highest inside the groups of threshold-based methods.

	MULTITHRESHOLDING OTSU			
	SE (%)	SP (%)	AC (%)	BAC (%)
Test 1	83,67	95,37	90,70	89,52
Test 2	86,02	95,22	91,98	90,62
Test 3	85,11	93,13	90,67	89,12
Test 4	86,88	94,54	92,17	90,71
Test 5	83,71	93,95	89,63	88,83
AVERAGE	85,08	94,44	91,03	89,76

Table 4: Evaluation of multithresholding Otsu method.

The following table summarizes the global average results obtained from the application of each implemented thresholding-based segmentation method to the 724 images within the database and facilitates the comparison between techniques inside the same category.

METHOD	AVERAGE EVALUATION MEASURES: THRESHOLDING			
	SE (%)	SP (%)	AC (%)	BAC (%)
Traditional Otsu	79,64	97,30	90,62	88,47
Local Otsu	81,97	94,23	88,63	88,10
Multithresholding Otsu	85,08	94,44	91,03	89,76

Table 5: Average evaluation measures for each implemented thresholding-based method.

4.3.2 Region-based methods

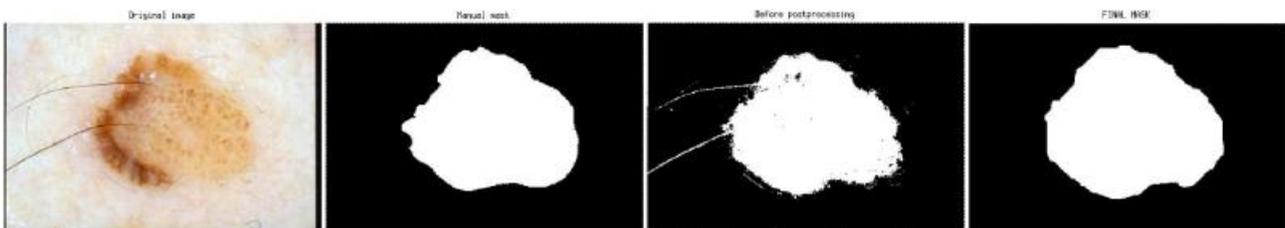
4.3.2.1 Two-region k-means clustering

The most general approach to dermoscopic image segmentation based on clustering techniques is the determination of two initial classes: one for lesion regions and one for skin regions. As described in Section 3.4.2.2, the feature matrix of the implemented k-means approaches always include the radial distance of each pixel to the center of the image, so that spatial information about the distribution of pixels is present when the image is divided into regions. Additionally, the values of each pixel for several colour components are taken into account. Regarding the selected colour space, three different approaches are distinguished.

The initialization matrix contains the data of the two initial centroids: centered in the image and middle values for the colour components in the case of the lesion cluster centroid; located on the image corner and also middle values for the colour components in the case of the skin cluster centroid.

4.3.2.1.1 Radial distance and *R, G, B* components

The most noticeable improvement of k-means clustering with respect to threshold-based segmentation is the inclusion of lesion areas with low pigmentation inside the final lesion region. Many centered light areas are now identified as lesion even if the contrast between lesion and skin is low, due to the spatial information used by the algorithm (Ex. 18).



Example 18: Two-region k-means clustering (radial distance, *R, G, B*). Low contrast between lesion and skin. From left to right: original image, manual mask, mask before postprocessing, final mask.

Additionally, the *RGB* k-means approach provides enhancement to the segmentation of colour-variegated lesions with respect to basic thresholding techniques. Nevertheless, the multithresholding approach still shows better performance with these lesions, due to the extraction of several inner lesion regions (Ex. 19).



Example 19: Two-region k-means clustering (radial distance, *R, G, B*). Colour-variegated lesion. From left to right: original image, manual mask, mask before postprocessing, final mask.

However, there is an appreciable introduced improvement with respect to multithresholding Otsu approach, regarding illumination inconsistency. The spatial information also allows a better segmentation of those images which contain areas with big variations in external illumination (Ex. 20).



Example 20: Two-region k-means clustering (radial distance, R, G, B). Illumination inconsistency. From left to right: original image, manual mask, mask before postprocessing, final mask.

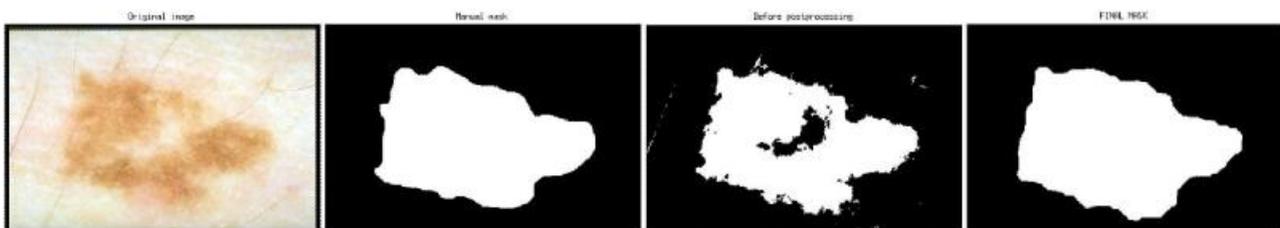
The next table contains the performance statistics obtained from applying this approach of k-means clustering to five different test sets of images. The global values of the evaluation metrics show a general improvement with respect to the thresholding-based methods, including the multithresholding approach.

	TWO-REGION K-MEANS CLUSTERING (RGB)			
	SE (%)	SP (%)	AC (%)	BAC (%)
Test 1	87,29	96,01	91,12	91,65
Test 2	87,70	96,75	91,88	92,22
Test 3	89,25	96,38	93,00	92,81
Test 4	89,37	95,66	92,45	92,51
Test 5	85,05	94,59	89,10	89,82
AVERAGE	87,73	95,88	91,51	91,80

Table 6: Evaluation of two-region k-means clustering (radial distance and R, G, B components).

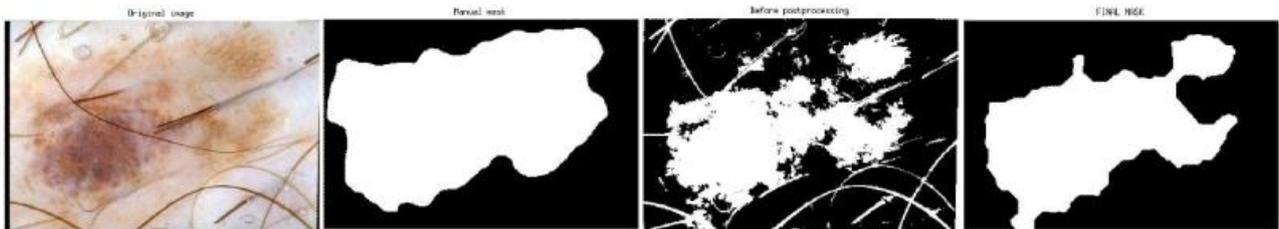
4.3.2.1.2 Radial distance and L, a, b components

This different approach of two-region k-means clustering also allows a better segmentation of images with skin textures and low contrast between lesion and skin (Ex. 21). The use of spatial information seems to be indispensable at the improvement of segmentation of areas with low pigmentation that the clustering-based techniques provide.



Example 21: Two-region k-means clustering (radial distance, L, a, b). Low contrast between lesion and skin. From left to right: original image, manual mask, mask before postprocessing, final mask.

Nevertheless, further development is needed to clustering techniques regarding the presence of image artefacts. Those lesions covered by hair contain areas that are still not correctly detected (Ex. 22).



Example 22: Two-region k-means clustering (radial distance, L , a , b). Presence of hair. From left to right: original image, manual mask, mask before postprocessing, final mask.

This approach of k-means clustering, which employs the components inside the Lab colour space, offers certain enhancement to the segmentation of colour-variegated lesions with respect to the approach based on the RGB colour components (Ex. 23).



Example 23: Two-region k-means clustering (radial distance, L , a , b). Colour-variegated lesion. From left to right: original image, manual mask, mask before postprocessing, final mask.

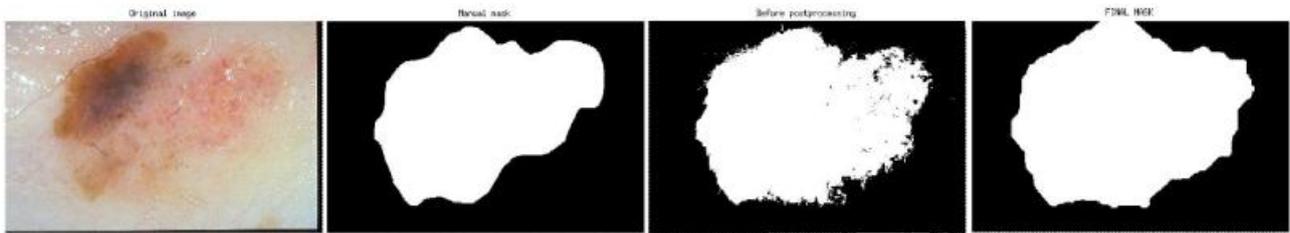
The following table shows the performance of the k-means clustering approach based on the radial distance and the L , a , b components. The evaluation measures extracted from five different test sets are highly similar to those obtained from the application the approach based on the RGB colour space.

	TWO-REGION K-MEANS CLUSTERING (Lab)			
	SE (%)	SP (%)	AC (%)	BAC (%)
Test 1	89,23	93,39	90,52	91,31
Test 2	87,81	94,42	90,92	91,11
Test 3	87,56	94,00	90,04	90,78
Test 4	88,30	92,89	89,67	90,60
Test 5	90,07	94,62	91,72	92,34
AVERAGE	88,59	93,86	90,57	91,23

Table 7: Evaluation of two-region k-means clustering (radial distance and L , a , b components).

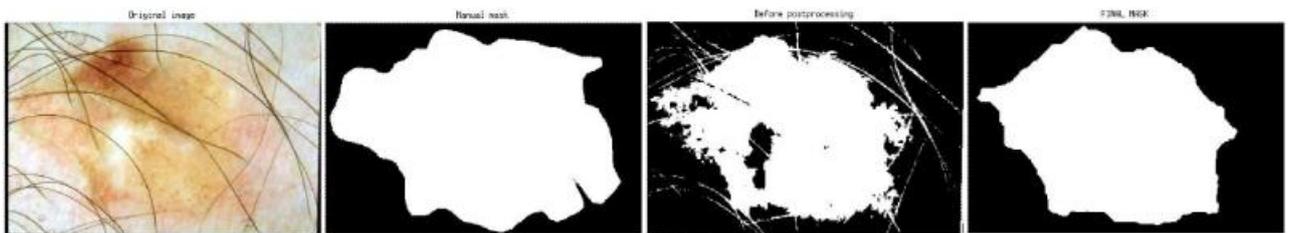
4.3.2.1.3 Radial distance and S , V components

The S and V components of the HSV colour space were selected as a third implementation of the k-means algorithm. The exclusion of the H component due to distorted information is explained in Section 3.4.2.2.3. As both previous implementations of two-region k-means clustering, the segmentation of images with low contrast between lesion and skin shows significant improvement (Ex. 24).



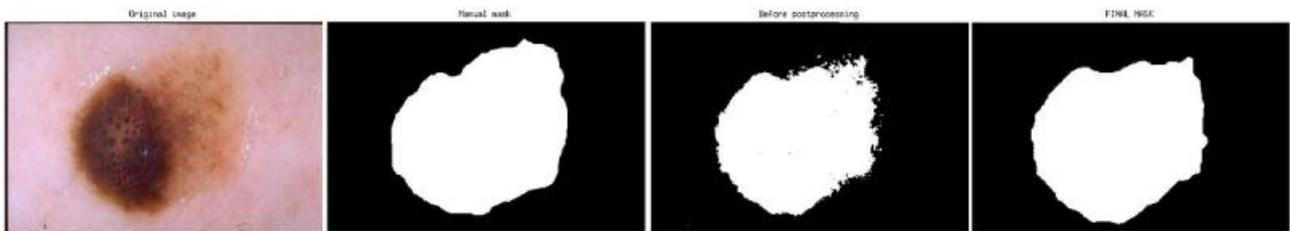
Example 24: Two-region k-means clustering (radial distance, S , V). Low contrast between lesion and skin. From left to right: original image, manual mask, mask before postprocessing, final mask.

Additionally, this implementation allows an increment of detected lesion areas which are covered by image artefacts, such as hairs (Ex. 25), with respect to the previous versions of the algorithm.

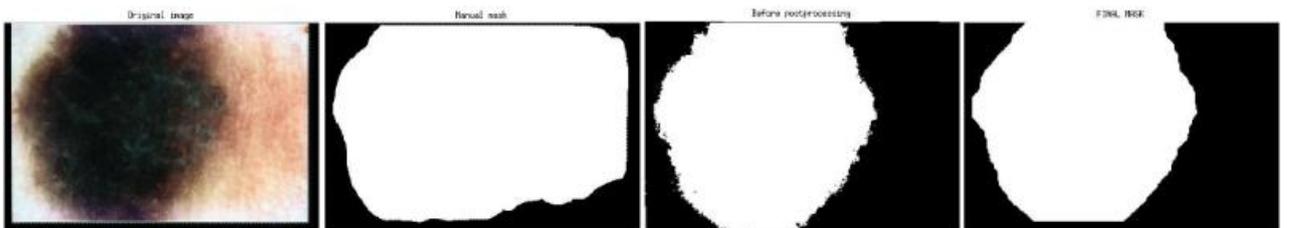


Example 25: Two-region k-means clustering (radial distance, S , V). Presence of hair. From left to right: original image, manual mask, mask before postprocessing, final mask.

The introduction of the S and V components also allows a better segmentation of colour-variegated areas that were not entirely included inside the lesion region with the previous implementations (Ex. 26). However, it can be observed that it is still necessary the division of the image into more than two regions, since the resultant masks of many lesions do not change with respect to those obtained from direct thresholding binarization (Ex. 27).

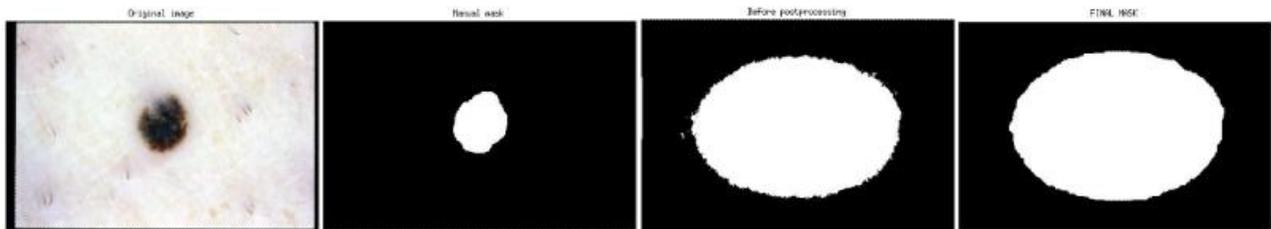


Example 26: Two--region k-means clustering (radial distance, S , V). Well-segmented, colour-variegated lesion. From left to right: original image, manual mask, mask before postprocessing, final mask.



Example 27: Two--region k-means clustering (radial distance, S , V). Poorly-segmented, colour-variegated lesion. From left to right: original image, manual mask, mask before postprocessing, final mask.

Finally, it has been observed that this implementation is affected by certain illumination problems, which cause the creation of a focus area around the lesion within the generated binary masks from those images with illumination inconsistency (Ex. 28).



Example 28: Two--region k-means clustering (radial distance, S, V). Illumination inconsistency. From left to right: original image, manual mask, mask before postprocessing, final mask.

Despite these segmentation handicaps due to illumination inconsistency, the next table shows that the evaluation metrics are globally higher than the metrics obtained from the other two versions of two-region k-means clustering. Therefore, the implementation based on radial distance and the S and V components can be considered to have the best performance among the two-region k-means segmentation methods.

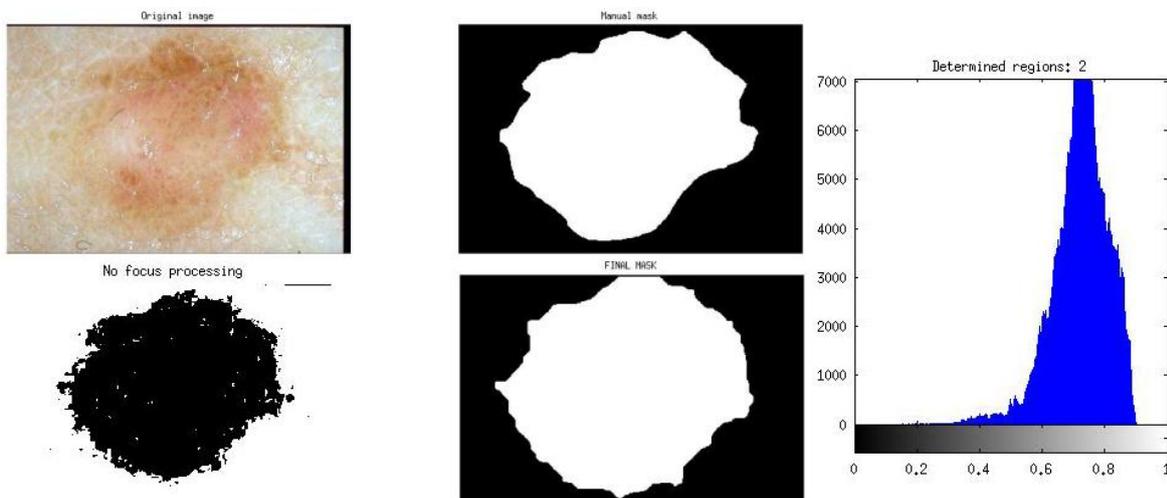
	TWO-REGION K-MEANS CLUSTERING (SV)			
	SE (%)	SP (%)	AC (%)	BAC (%)
Test 1	91,13	96,14	92,78	93,64
Test 2	89,10	95,30	90,79	92,20
Test 3	91,37	95,56	92,68	93,46
Test 4	90,11	96,07	92,73	93,09
Test 5	87,94	95,87	91,18	91,90
AVERAGE	89,93	95,79	92,03	92,86

Table 8: Evaluation of two-region k-means clustering (radial distance and S, V components).

4.3.2.2 Multi-region k-means clustering

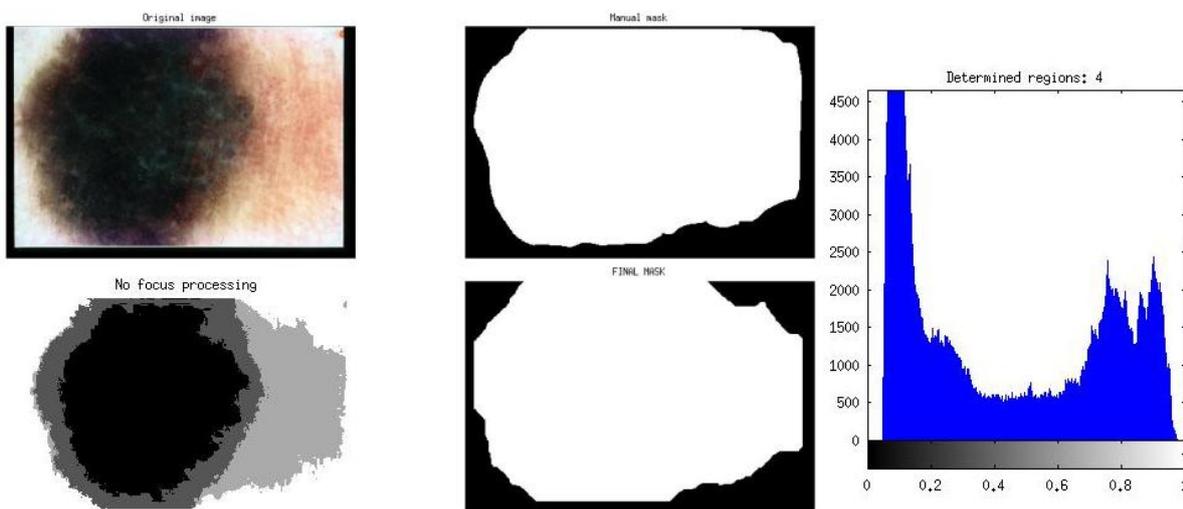
Since numerous dermoscopic images require the detection of more than two regions (one associated with lesion areas and one associated with skin) because of the variability of colour and texture within the lesion region, the multi-region k-means approach provides an advanced solution in order to improve the segmentation performance. The suitable number of regions within the image is first determined regarding the local maxima in the luminance histogram. Thus, the number of initial cluster centroids is equivalent to the determined number of regions. Each initial cluster is provided with middle values of colour components and an adjustable radial distance, so that the cluster corresponding to skin is initialized as the farthest one. Regarding the selected colour components, the k-means algorithm based on S and V components was selected over the other implementations due to its better performance, as showed above by the obtained global evaluation measures.

The multi-region k-means approach maintains the good segmentation performance with those images which contain lesion areas with low pigmentation (Ex. 29). As the example shows, although only two regions are determined, the spatial information employed by the algorithm facilitates the inclusion inside the final lesion region of lighter areas in the center of the image.



Example 29: Multi-region k-means clustering. Low contrast between lesion and skin. On top: original image (left), manual mask (center) luminance histogram (right). On bottom: segmented regions (left), final mask (right).

The multi-region implementation introduces important advances regarding the segmentation of colour-variegated lesions (Ex. 30). The accurate separation of inner lesion regions allows the inclusion of lesional areas that were not detected as lesion in any of the previous methods, neither threshold-based nor two-region k-means approaches.

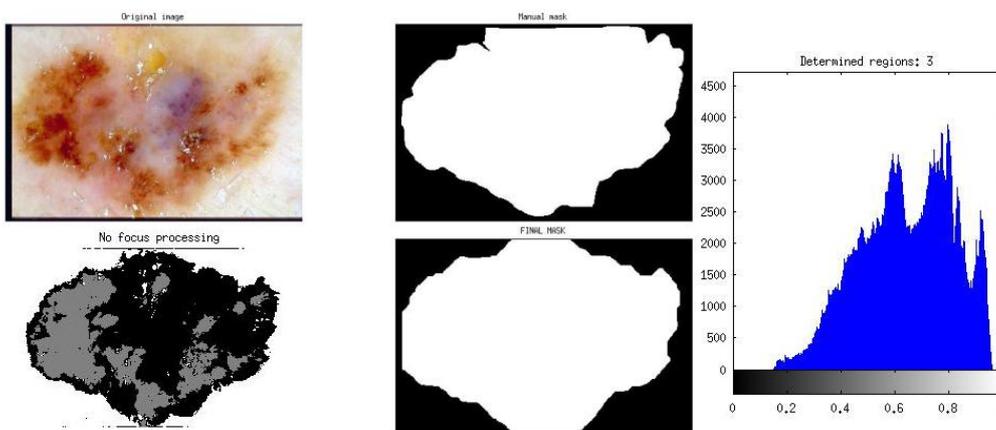


Example 30: Multi-region k-means clustering. Colour-variegated lesion. On top: original image (left), manual mask (center) luminance histogram (right). On bottom: segmented regions (left), final mask (right).

Additionally, by taking into account several regions inside the clustering algorithm, considerable progress is achieved respecting the presence of both external and cutaneous features within the image. Regression areas are included inside the lesion region (Ex. 31). Besides, those lesions affected simultaneously by intrinsic and external features are also well segmented; for example, despite the presence of blue-white veil and traces of liquid (Ex. 32).

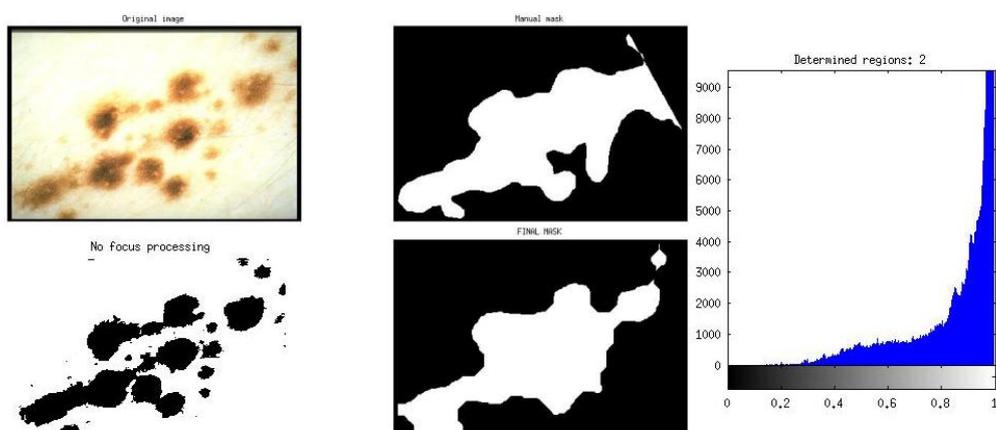


Example 31: Multi-region k-means clustering. Regression areas inside the lesion. On top: original image (left), manual mask (center) luminance histogram (right). On bottom: segmented regions (left), final mask (right).



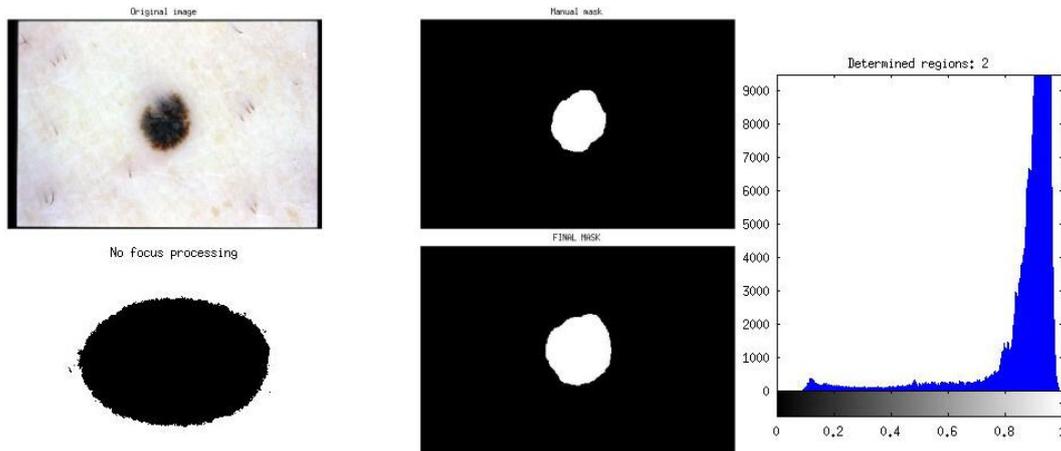
Example 32: Multi-region k-means clustering. Presence of blue-white veil and traces of liquid. On top: original image (left), manual mask (center) luminance histogram (right). On bottom: segmented regions (left), final mask (right).

It has been observed that even those fragmented lesions achieve a sufficiently compact segmentation after the postprocessing phase, which maintains the similarities with the segmentation fixed by their corresponding manual mask (Ex. 33).



Example 33: Multi-region k-means clustering. Fragmented lesion. On top: original image (left), manual mask (center) luminance histogram (right). On bottom: segmented regions (left), final mask (right).

The segmentation problems regarding illumination inconsistency, which result from the k-means clustering based on S and V components, were addressed by analysing the shape properties of the lesion region along with the location of the highest maxima in the luminance histogram. This method was entirely described in Section 3.4.2.2.2. Therefore, the focus created around the lesion region within the generated mask is eliminated in most cases and the real lesion contour is extracted (Ex. 34).



Example 34: Multi-region k-means clustering. Addressed illumination inconsistency problem. . On top: original image (left), manual mask (center) luminance histogram (right). On bottom: segmented regions (left), final mask (right).

The next table demonstrates the high evaluation values obtained from the application of the multi-region k-means approach to five different test sets, especially an improvement of the sensitivity value, the most significant one from a clinical point of view. Despite certain persistent illumination handicaps and some difficulties with the complicated segmentation of images with low contrast between lesion and skin, the performance of this method is the best accomplished among all the implemented thresholding-based methods and simpler clustering versions.

MULTI-REGION K-MEANS CLUSTERING				
	SE (%)	SP (%)	AC (%)	BAC (%)
Test 1	91,83	91,16	91,10	91,49
Test 2	90,75	93,78	92,27	92,27
Test 3	90,72	94,45	92,56	92,58
Test 4	92,69	93,43	92,87	93,06
Test 5	92,59	93,95	93,15	93,27
AVERAGE	91,71	93,35	92,39	92,53

Table 9: Evaluation of multi-region k-means clustering.

The following table summarizes the global average results obtained from the application of each implemented clustering-based segmentation method to the 724 images within the database and facilitates the comparison between techniques inside the same category.

AVERAGE EVALUATION MEASURES: CLUSTERING				
METHOD	SE (%)	SP (%)	AC (%)	BAC (%)
Two-region k-means (RGB)	87,73	95,88	91,51	91,80
Two-region k-means (Lab)	88,59	93,86	90,57	91,23
Two-region k-means (SV)	89,93	95,79	92,03	92,86
Multi-region k-means	91,71	93,35	92,39	92,53

Table 10: Average evaluation measures for each implemented clustering-based method.

4.3.3 Segmentation based on deformable models

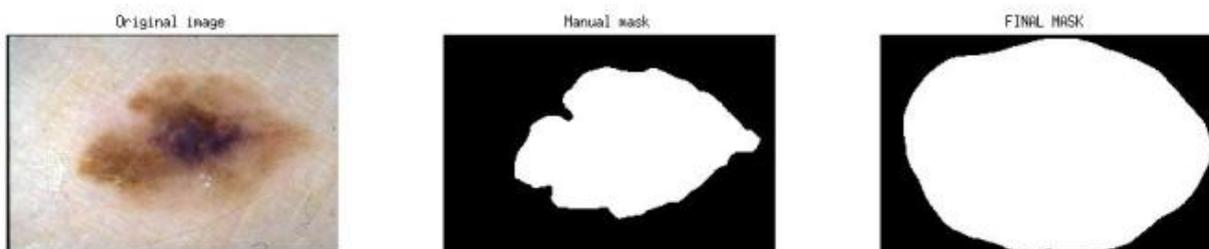
4.3.3.1 Gradient vector flow snakes

As described in Section 3.4.2.3.3, among different alternatives for the initialization of the algorithm, the initial curve is extracted from the automatic mask generated from the multi-region k-means method, which provides the highest segmentation performance. The objective is to reduce the sensitivity of the algorithm to an initial contour located far from the real boundaries. Additionally, the lesion contour is then expanded and the number of points that constitute it is decremented, so that the snake is simplified and the computational load is diminished.

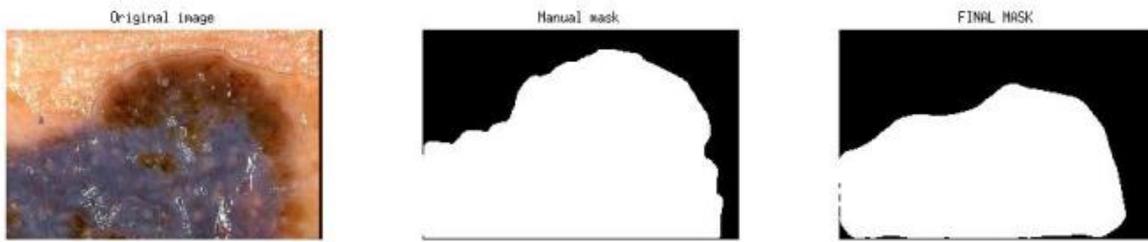
This initial snake is then iteratively deformed regarding the energy minimization formulation, studied in Section 2.3.4.1.2, taking into account the gradient vector flow (GVF) as an additional external force, computed from the gradient of the edge map of the B component.

The main handicap observed in the resulting masks from the application of this deformable model is that the snakes do not adjust to the real boundaries precisely. Two particular cases can be differentiated: when the initial curve is not near the real boundaries and when the initial curve is already sufficiently adjusted.

In the former case, if the original snake was not close to the real lesion boundaries due to an imprecise segmentation, the snake does not deform enough in order to reach the contour, either outside (Ex. 35) or inside (Ex. 36) the lesion areas.

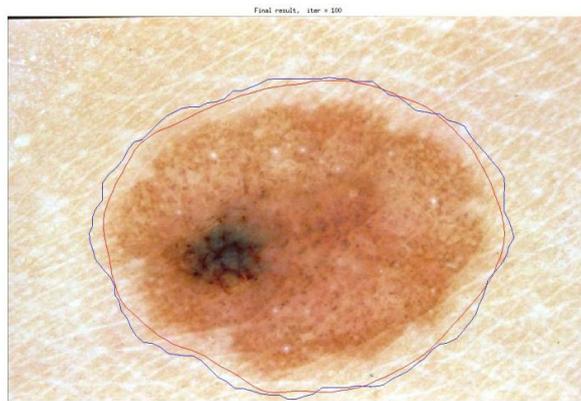


Example 35: GVF snakes. Final curve outside the lesion region. From left to right: original image, manual mask, final mask.



Example 36: GVF snakes. Final curve inside the lesion region. From left to right: original image, manual mask, final mask.

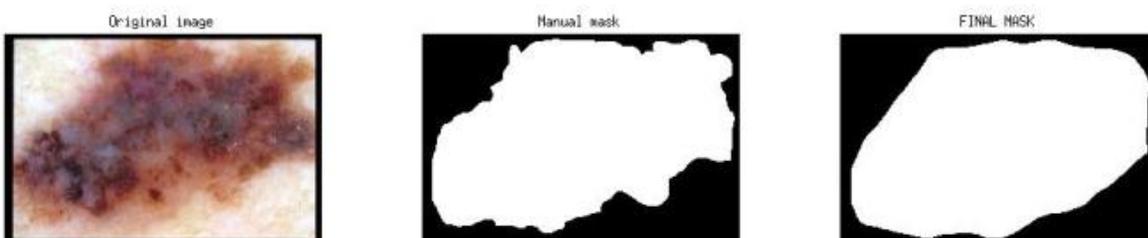
In the latter case, when the initialization is close enough to the real boundaries, the GVF deformable model tends to reduce the details of the former contour, i.e., causes under-segmentation (Ex. 37). As a consequence, the sensitivity of the final masks may be increased because a bigger area is being included inside the lesion region, but the specificity decreases simultaneously, since healthy skin is also being considered as lesion. This happens in centered (Ex. 38) and expanded (Ex. 39) lesions without distinction.



Example 37: GVF snakes. Comparative of initial (blue) and final (red) snake, after 100 iterations.



Example 38: GVF snakes. Centered lesion that has been under-segmented. From left to right: original image, manual mask, final mask.



Example 39: GVF snakes. Expanded lesion has been under-segmented. From left to right: original image, manual mask, final mask.

Provided that the initial curve is over-adjusted and some lesion areas are being excluded, the reduction of details within the contour may help the final segmentation. Nevertheless, if the contour is already refined enough, this process becomes unnecessary. It can be observed, that even after the analysis of the numerous parameters that intervene in the deformation procedure, along with the validation of those which govern the internal forces, the application of the model results in under-segmentation of the lesions.

Variability within the image due to skin and lesion textures, as well as the presence of features, may avoid the algorithm from converging at the real boundaries, because the increment in number of iterations was proved to not improve the results. It can be observed in the next table how the obtained average sensitivity from the multi-region k-means method is maintained, whereas the other metrics deteriorate, especially the specificity value. Thus, the GVF deformable model may not be an appropriate tool for segmentation refinement.

	GVF SNAKES			
	SE (%)	SP (%)	AC (%)	BAC (%)
Test 1	92,21	90,37	90,39	91,29
Test 2	91,43	91,24	91,41	91,33
Test 3	92,89	91,23	91,30	92,06
Test 4	92,57	90,23	91,15	91,40
Test 5	89,39	91,15	90,56	90,27
AVERAGE	91,70	90,84	90,96	91,27

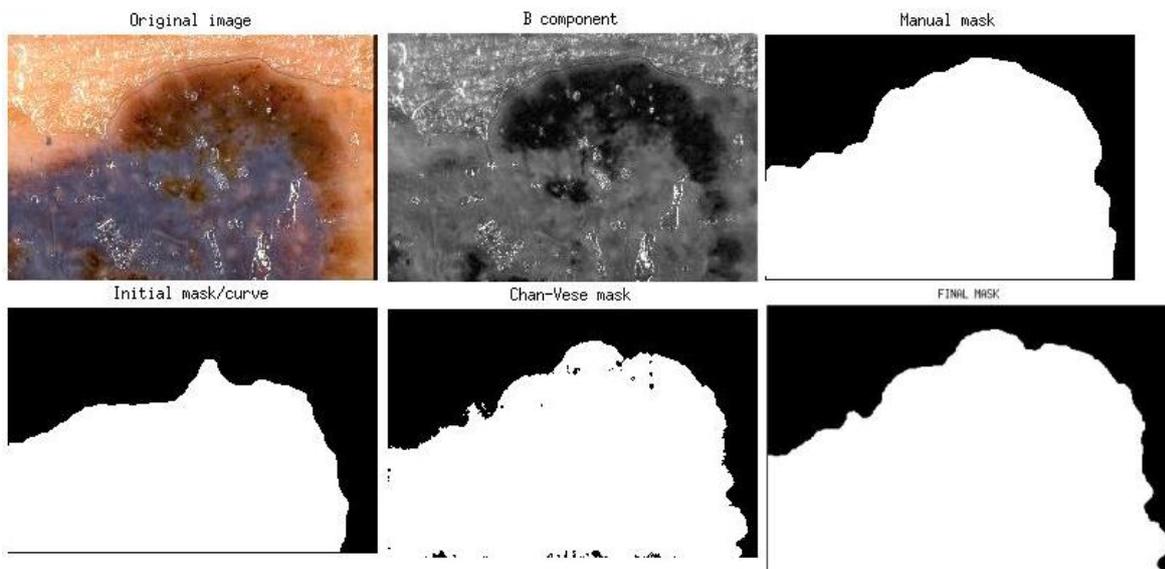
Table 11: Evaluation of GVF snakes

4.3.3.2 Chan-Vese method

The selected initialization for the Chan-Vese method was also the obtained segmentation from the multi-region k-means clustering due to its high performance. Two different approaches of the method were implemented regarding the initial curve: for one of them the original contour was maintained, whereas the number of points within the curve was reduced for the second approach. Since the speed of computation of the algorithm was faster than the GVF model, both approaches were finally considered.

The deformation procedure, which is based on level sets formulation (studied in Section 2.3.4.2.1), was applied to the luminance image and the *B* component, but due to a better performance, the latter was finally selected as the gray level image to be introduced in the algorithm.

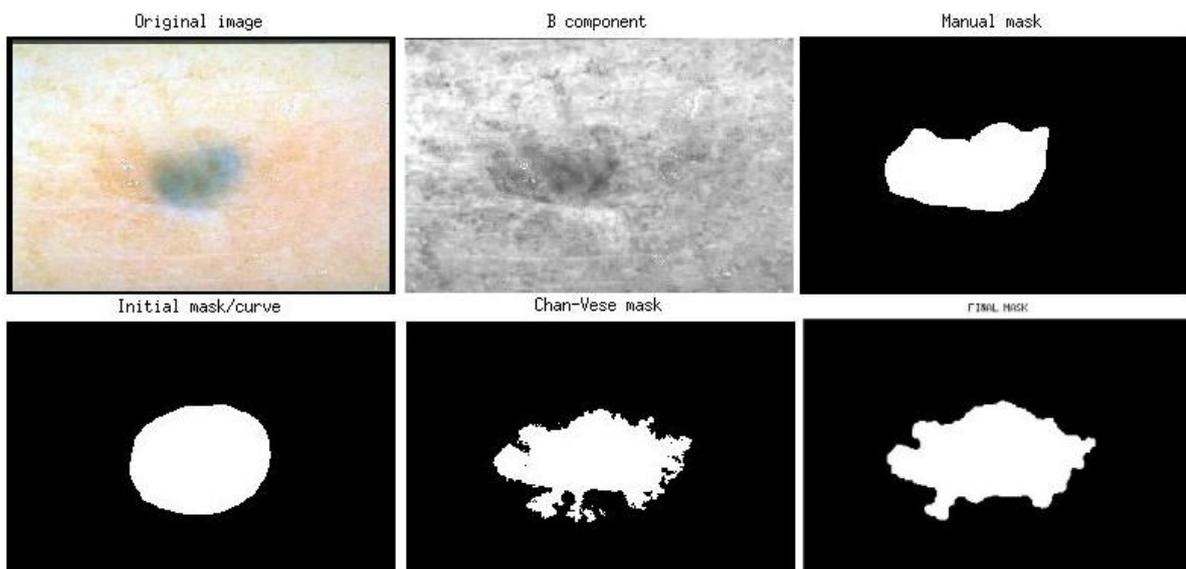
The contours of this model perform without requiring edges information; therefore, a better adjustment of the final curve to the real boundaries can be globally observed, even at the presence of intrinsic features, such as blue-white veil, and external features, such as traces of liquid (Ex. 40).



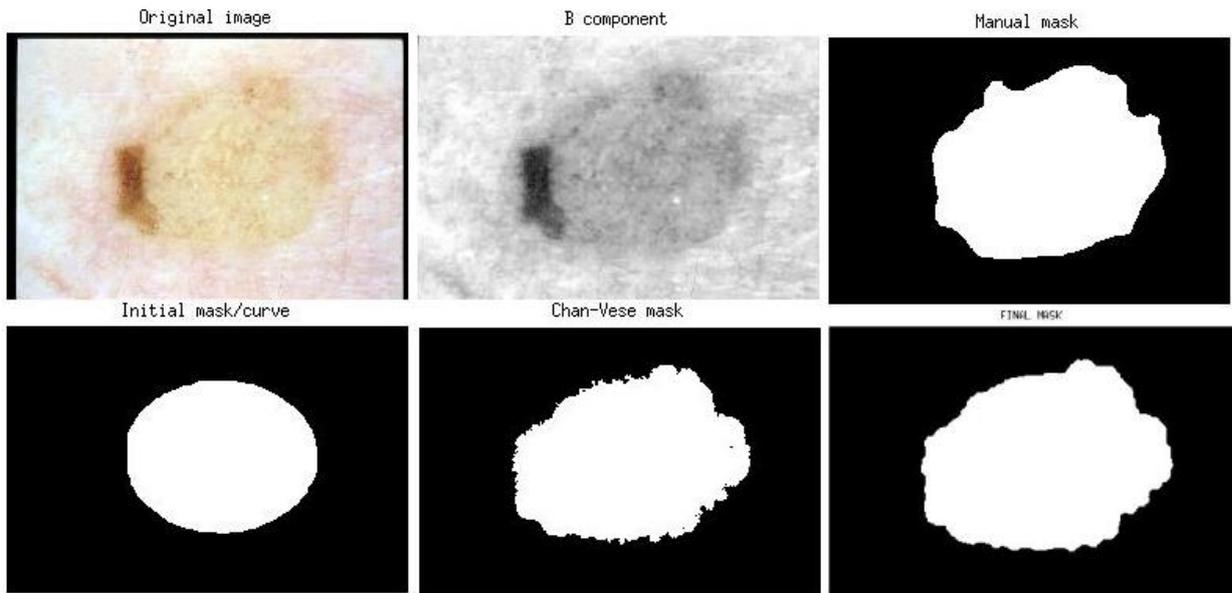
Example 40: Chan-Vese method. Adjustment of final curve to real boundaries. On top: original image (left), B component (center), manual mask (right). On bottom: initial mask (left), mask before postprocessing (center), final mask (left).

In some cases, where the initial contour is not close enough to the lesion boundaries and is required to deform, it is not accurately adjusted after the deformation process (Ex. 41). The number of iterations was incremented from 100 to 300 in order to allow the snake to continue its deformation in case the convergence was not achieved, but the real contour was not reached and there was also cases where the curve became too detailed and some areas of the lesion boundaries were excluded from the final lesion region.

Nevertheless, when the initial curve is sufficiently close to the real boundaries and its borders are not necessarily well defined, the adjustment is well performed and a more accurate segmentation is achieved, even if there is low contrast between lesion areas and skin (Ex. 42).



Example 41: Chan-Vese method. Final contour not completely adjusted. On top: original image (left), B component (center), manual mask (right). On bottom: initial mask (left), mask before postprocessing (center), final mask (left).



Example 42: Chan-Vese method. Final contour well adjusted. On top: original image (left), B component (center), manual mask (right). On bottom: initial mask (left), mask before postprocessing (center), final mask (left).

Regarding the improved segmentation, mainly in those cases where the initial snake is close enough to real boundaries but the borders are not well defined, the Chan-Vese deformable model become a useful tool for addressing the illumination inconsistency problems that were persistent after the multi-region k-means segmentation, where a focus area surrounded the lesion boundaries within the final mask.

This model also allows the adjustment of many snakes that were previously not well determined due to low-contrast between lesion and skin, variability in colour or texture and cutaneous or external features. Therefore, the Chan-Vese method can be considered as a good option for segmentation refinement, which can be observed in the following tables.

CHAN-VESE METHOD (B COMPONENT)				
	SE (%)	SP (%)	AC (%)	BAC (%)
Test 1	92,25	93,84	93,02	93,05
Test 2	93,68	91,92	92,58	92,80
Test 3	93,88	93,88	93,78	93,88
Test 4	92,34	93,15	92,30	92,75
Test 5	93,09	93,12	92,63	93,11
AVERAGE	93,05	93,18	92,86	93,12

Table 12: Evaluation of Chan-Vese method on B component.

	CHAN-VESE METHOD (B COMPONENT, REDUCED POINTS)			
	SE (%)	SP (%)	AC (%)	BAC (%)
Test 1	94,01	90,49	91,21	92,25
Test 2	95,36	91,03	93,13	93,20
Test 3	95,34	89,93	92,20	92,63
Test 4	95,03	90,81	91,80	92,92
Test 5	96,10	92,05	93,68	94,07
AVERAGE	95,17	90,86	92,41	93,01

Table 13: Chan-Vese method on B component and reduced points on initial curve.

Both implementations of the Chan-Vese method show a global improvement of the segmentation performance with respect to the results obtained from the multi-region k-means clustering. The approach based on an initialization with reduced points achieves the highest sensitivity value among all the implemented methods. The loss of detail at the contour borders assures the inclusion of lesion areas, but simultaneously there is also inclusion of certain healthy skin areas, which causes the decrease in the specificity value. Regarding the approach that does not reduce the points within the initial curve, the evaluation metrics of the multi-region k-means are approximately maintained, but there is noticeable increase in the sensitivity value, which is clinically the most significant one.

The following table summarizes the global average results obtained from the application of each implemented segmentation method based on deformable models to the 724 images within the database and facilitates the comparison between techniques inside the same category.

METHOD	AVERAGE EVALUATION MEASURES: DEFORMABLE MODELS			
	SE (%)	SP (%)	AC (%)	BAC (%)
GVF snakes	91,70	90,84	90,96	91,27
Chan-Vese method	93,05	93,18	92,86	93,12
Chan-Vese method (red. points)	95,17	90,86	92,41	93,01

Table 14: Average evaluation measures for each implemented segmentation method based on deformable models.

4.3.4 Segmentation based on the bag-of-words model

The applied implementations based on this experimental segmentation method are first discussed before definitive examples and statistics are shown. Different results were obtained from different combinations of keypoints detectors, keypoints descriptors and visual vocabulary properties.

Regarding the detection of keypoints from the set of reference images, the Harris Laplace detector and a dense sampling were both implemented. It was observed that the Harris

detector placed numerous points on noisy features within the dermoscopic images, which may not be as representative as desired. Artefacts such as hairs or traces of liquid can be considered as relevant areas in the image by the detector, excluding other potential keypoints from being selected (Ex. 43).



Example 43: Harris Laplace detection at the presence of traces of liquid.

This may have consequences regarding the generation of the visual vocabulary. The masks obtained from the application of Harris Laplace as detector contained distorted information, so that the lesion contour could not be well identified. Therefore, the dense sampling technique for the extraction of local features, which is necessarily applied to the images to be segmented, was finally considered for the set of reference images too. This way, the visual codebook is generated from uniformly distributed and separated lesion and skin keypoints.

On the other hand, the implementation of colourSIFT descriptors provided noisy matching and the generated masks could not be analysable. The high variability in lesion and skin textures among the colour components within the opponent colour space, as well as the matching technique, may be the cause of the low performance of the OppSIFT descriptor with the dermoscopic images within the database. Several vocabulary sizes were implemented, as well as its combination with the mentioned Harris Laplace detector and dense sampling, but no improvement was achieved.

The SIFT descriptor, which describes keypoints within the luminance domain, achieved better results and was the one selected. However, the size of the visual vocabulary and the proportion of lesion and skin words are crucial. A small vocabulary may contain a smaller number of representative keypoints, but the matching phase is simplified due to the reduction in the number of possible matches. As the total number of visual words is reduced and the proportion of lesion and skin words is maintained to be 50/50, the detection of the contour of the lesion region within the generated binary mask is facilitated. The best performance was accomplished with a vocabulary constituted by 32 lesion words and 32 skin words. All the examples and the final statistics included in this section have been achieved with this vocabulary.

Regarding the superpixel-level approach, the employed vocabulary for the creation of the frequency histograms per superpixel and the dictionary of reference histograms was the visual codebook previously mentioned. The resulting masks showed a poor performance of the adaptation of the method to superpixel-level. Along the segmented dermoscopic images, most superpixels found their closest frequency histogram among the skin histograms inside the dictionary after the histogram intersection technique and a softer assignment. Thus, the generated masks contained most regions tagged as skin and were not similar to the corresponding manual masks.

Several sizes for the dictionary of reference histograms were implemented. Although limited improvements were noticed, a smaller dictionary which contained 100 histograms for lesion and 100 histograms for skin showed an increased average of superpixels tagged correctly as lesion (Ex. 44). Consequently, the superpixel approach was discarded from a final implementation due to its low performance.

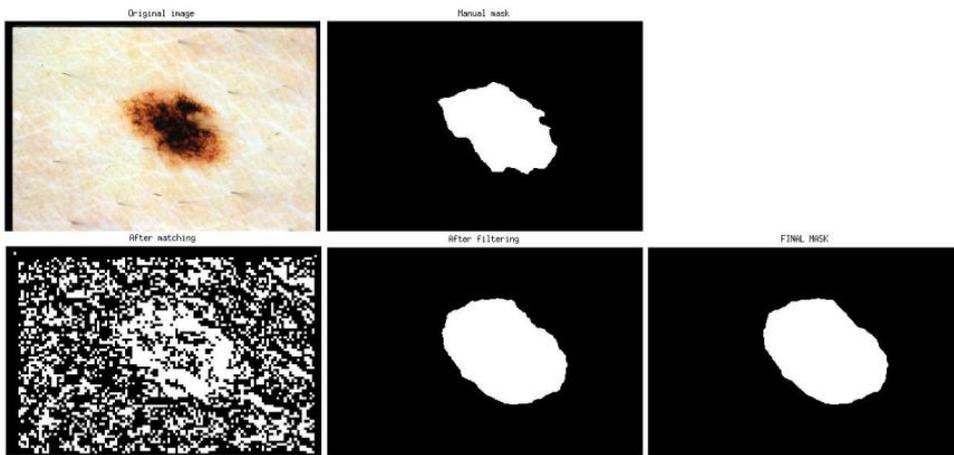


Example 44: Segmentation based on visual vocabulary and histogram intersection at superpixel-level.

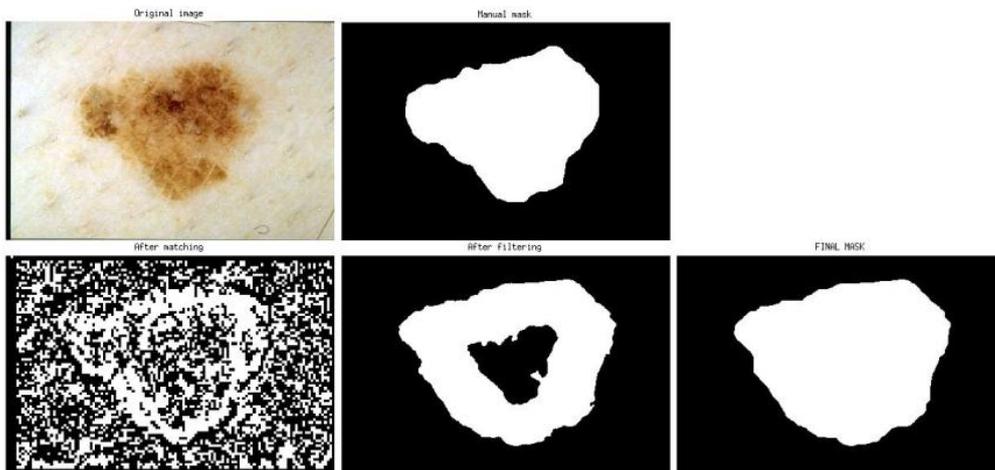
Therefore, the definitive implementation of this experimental segmentation method is based on the following combination of steps:

- Dense sampling for the extraction of local features from both the set of reference images and the images which enter the system to be segmented.
- SIFT descriptor for the description of all the detected keypoints.
- K-means clustering for the generation of a visual vocabulary for lesion regions that contains 32 lesion words and a visual vocabulary for skin regions that contains 32 skin words.
- Individual matching of each keypoint with one of the 64 visual words within the codebook, so that it is classified as lesion or skin.
- Creation of binary mask by nearest-neighbour interpolation and application of pillbox filter in order to eliminate the noise on skin areas and compact the lesion region.

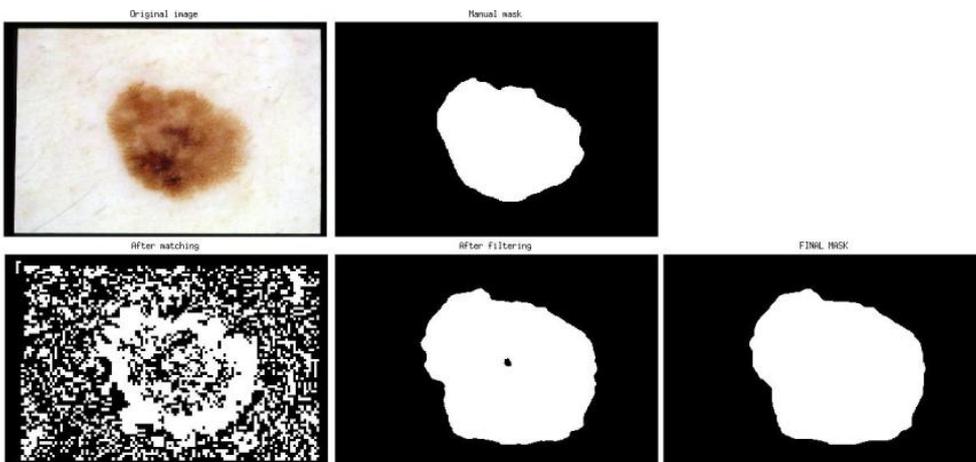
The performance of the method is high for centered and compact lesions, whose lesion borders are well discerned from the rest of the image. Some of them are already compact before the postprocessing phase (Ex. 45), whereas other lesions need the postprocessing to a greater (Ex. 46) or lesser (Ex. 47) extent, mainly for the filling of holes inside the lesion region.



Example 45: Segmentation based on dense sampling, SIFT, 32/32 vocabulary. Compact lesion. On top: original image (left), manual mask (right). On bottom: mask after matching (left), mask after filtering (center), final mask (right).

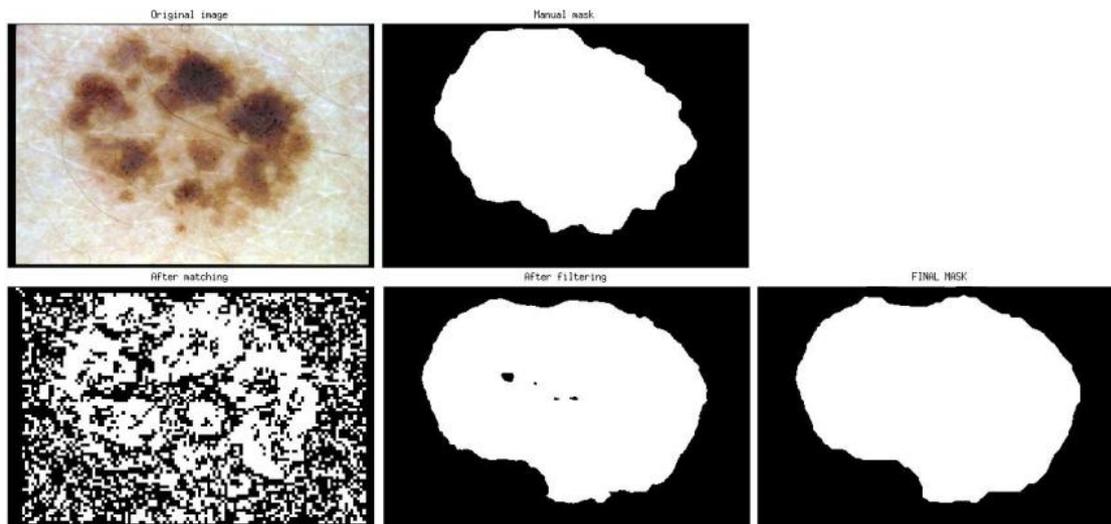


Example 46: Segmentation based on dense sampling, SIFT, 32/32 vocabulary. Compact lesion that requires postprocessing to fill holes. On top: original image (left), manual mask (right). On bottom: mask after matching (left), mask after filtering (center), final mask (right).

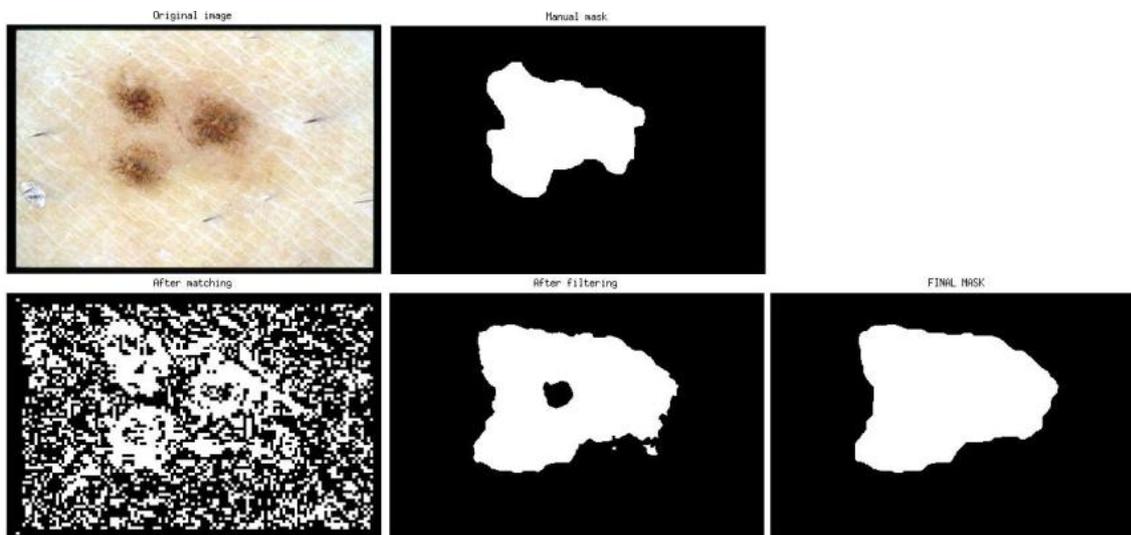


Example 47: Segmentation based on dense sampling, SIFT, 32/32 vocabulary. Compact lesion that requires postprocessing to fill the hole inside the lesion region. On top: original image (left), manual mask (right). On bottom: mask after matching (left), mask after filtering (center), final mask (right).

As it may be observed, the generated masks are rough and certainly expanded. Despite the elimination of the expansion process inside the postprocessing phase, this causes the specificity value to increase due to the inclusion of healthy skin areas inside the final lesion region. Nevertheless, those compact lesions with colour-variegated areas (Ex. 48) or surrounding lesion areas with low pigmentation (Ex. 49) can also be segmented due to the strong filtering process.

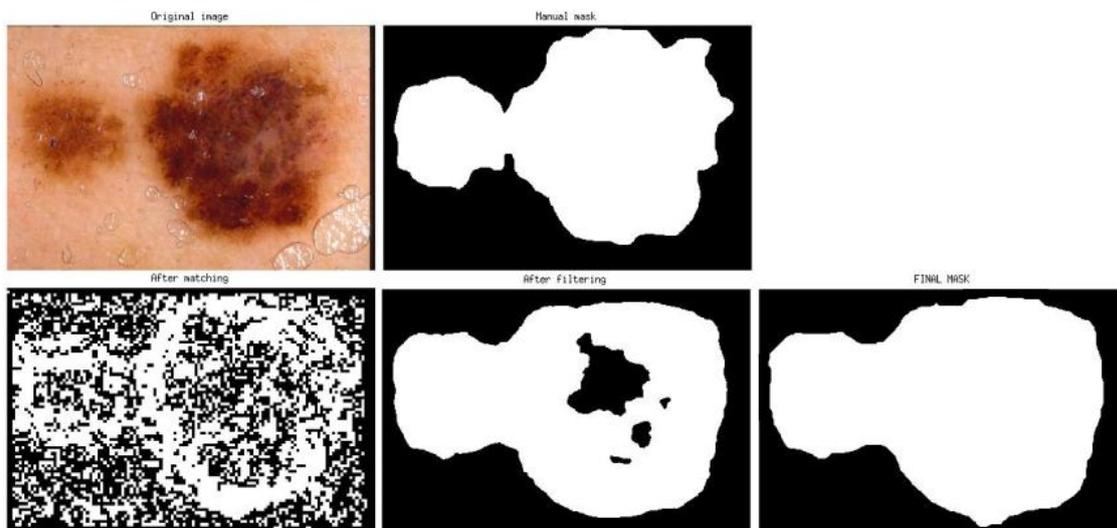


Example 48: Segmentation based on dense sampling, SIFT, 32/32 vocabulary. Compact, colour-variegated lesion. On top: original image (left), manual mask (right). On bottom: mask after matching (left), mask after filtering (center), final mask (right).



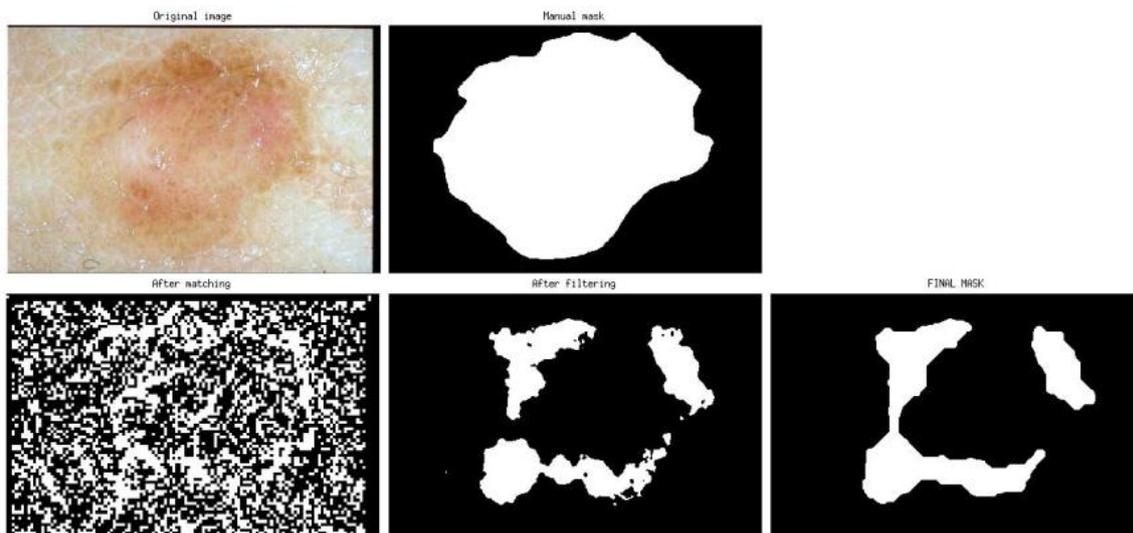
Example 49: Segmentation based on dense sampling, SIFT, 32/32 vocabulary. Compact lesion with areas of low pigmentation. On top: original image (left), manual mask (right). On bottom: mask after matching (left), mask after filtering (center), final mask (right).

Regarding expanded lesions, a good segmentation is also achieved for those lesions that are not defined by extremely fuzzy borders (Ex. 50).



Example 50: Segmentation based on dense sampling, SIFT, 32/32 vocabulary. Expanded lesion. On top: original image (left), manual mask (right). On bottom: mask after matching (left), mask after filtering (center), final mask (right).

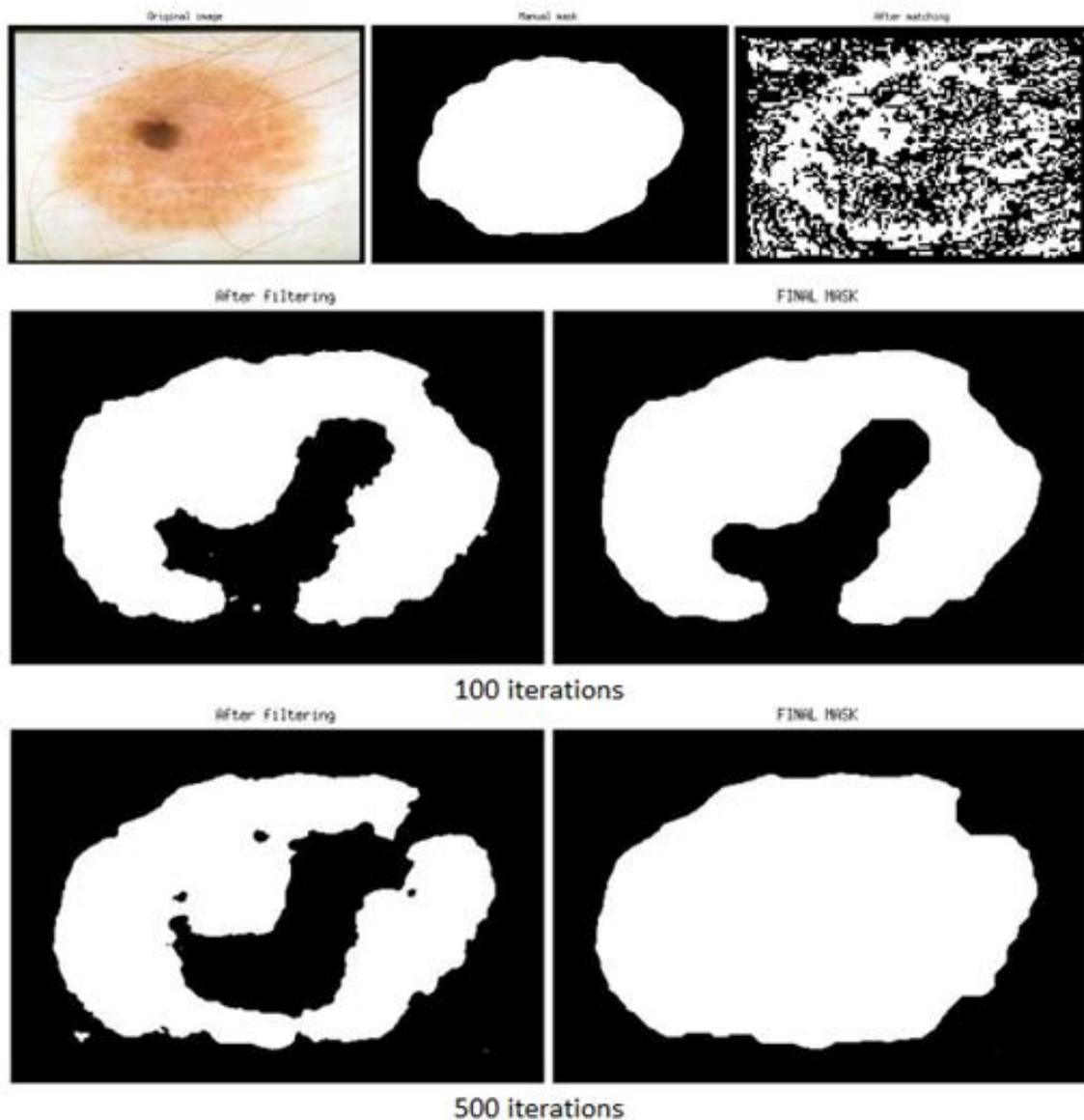
The main problem of this method is associated with the connectivity of lesion borders. The pillbox filter may facilitate the connection of borders in those compact lesions which contain at least some areas with high contrast between lesion and skin. However, for those lesions whose entire lesion region has low pigmentation, the closing of lesion borders is not accomplished by the filtering process neither the postprocessing phase (Ex. 51).



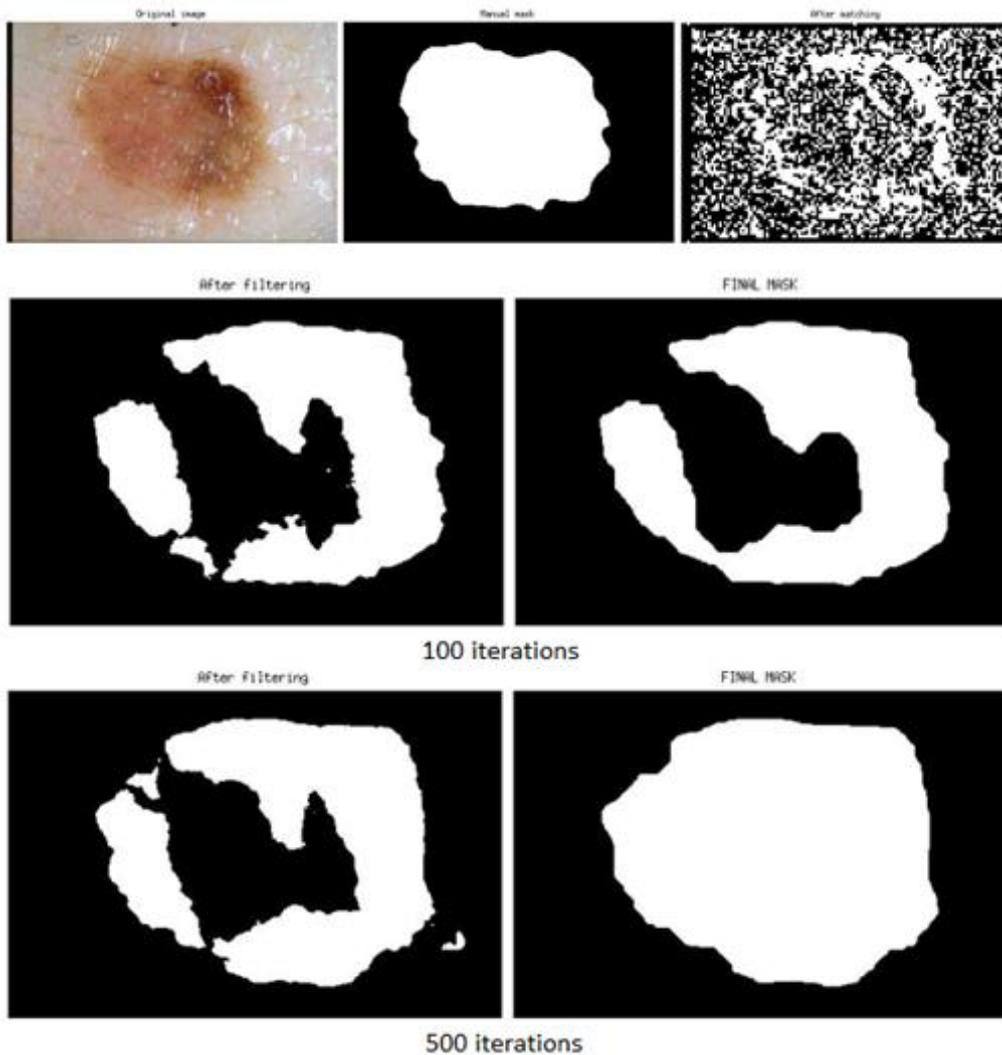
Example 51: Segmentation based on dense sampling, SIFT, 32/32 vocabulary. Low contrast between lesion and skin. On top: original image (left), manual mask (right). On bottom: mask after matching (left), mask after filtering (center), final mask (right).

Nonetheless, it has been observed that the connectivity of lesion borders improves with the increment of iterations of the k-means clustering employed for the generation of both lesion and skin vocabularies. The k-means algorithm is applied five times consecutively and each time there is a number of iterations until it converges. If the number of iterations that the algorithm takes to converge and assign each described keypoint its closest cluster is incremented from

100 iterations to 500 iterations, the matching results do not change visually, but the small variations within the border areas are sufficient in order to improve the connection and close the lesion region, including colour-variegated lesions (Ex. 52) and images affected by the presence of artefacts, such as traces of liquid (Ex. 53).



Example 52: Segmentation based on dense sampling, SIFT, 32/32 vocabulary. Colour-variegated lesion. Improvement in connectivity when the number of iterations of k-means clustering is incremented. On top: original image (left), manual mask (center), mask after matching (right). On center (100 iterations): mask after filtering (left), final mask (right). On bottom (500 iterations): mask after filtering (left), final mask (right).



Example 53: Segmentation based on dense sampling, SIFT, 32/32 vocabulary. Presence of traces of liquid. Improvement in connectivity when the number of iterations of k-means clustering is incremented. On top: original image (left), manual mask (center), mask after matching (right). On center (100 iterations): mask after filtering (left), final mask (right). On bottom (500 iterations): mask after filtering (left), final mask (right).

The following table shows the evaluation metrics obtained from the extraction, description and matching of keypoints within five different test sets.

	DENSE SAMPLING + SIFT, 32/32 WORDS, 500 ITER.			
	SE (%)	SP (%)	AC (%)	BAC (%)
Test 1	87,17	82,85	82,79	85,01
Test 2	86,22	83,42	81,84	84,82
Test 3	87,71	83,30	82,75	85,50
Test 4	85,57	83,39	82,13	84,48
Test 5	85,69	84,35	82,34	85,02
AVERAGE	86,47	83,46	82,37	84,97

Table 15: Evaluation of segmentation method based on the extraction of keypoints by dense sampling, the description of keypoints by SIFT and the matching of keypoints to a vocabulary generated by k-means algorithm with 500 iterations and constituted by 32 lesion words and 32 skin words.

As previously mentioned the specificity values are lower than usual due to the rough and expanded generated masks, which include large areas of healthy skin inside the lesion region. Also low values for the sensitivity and accuracy are obtained, mainly due to the nature of the matching process and the hard assignment, crucial step inside the method.

Nevertheless, although it is considered as an experimental method due to the shortage of specialised literature in order to apply the extraction of local points and the bag-of-words model to automatic segmentation, the global results show the potential of this implementation and the possibilities of improvement for a better performance. Additionally, this method can be easily extended to the classification phase within the diagnosis of dermoscopic images, since the extracted and described keypoints can also be separated regarding the cutaneous features they represent.

5. Conclusions

The implementation and evaluation of several methods were proposed for the automatic segmentation of lesion regions in dermoscopic images. The implemented algorithms include methods based on different state of the art techniques, methods with introduced modifications for a better adaptation to features associated with dermoscopic images, as well as a new proposed approach based on the extraction of local points and the generation of visual vocabularies.

After the corresponding preprocessing phases and the automatic segmentation procedure, the output of each automatic segmentation method was compared with the corresponding manual segmentation in order to evaluate the performance, once the postprocessing phase was completed.

The following table summarizes the global average results obtained from the application of each implemented method to the 724 images within the database and facilitates the comparison between techniques regarding the evaluation metrics.

METHOD	AVERAGE EVALUATION MEASURES			
	SE (%)	SP (%)	AC (%)	BAC (%)
Traditional Otsu	79,64	97,30	90,62	88,47
Local Otsu	81,97	94,23	88,63	88,10
Multithresholding Otsu	85,08	94,44	91,03	89,76
Two-region k-means (RGB)	87,73	95,88	91,51	91,80
Two-region k-means (Lab)	88,59	93,86	90,57	91,23
Two-region k-means (SV)	89,93	95,79	92,03	92,86
Multi-region k-means	91,71	93,35	92,39	92,53
GVF snakes	91,70	90,84	90,96	91,27
Chan-Vese method	93,05	93,18	92,86	93,12
Chan-Vese method (red. points)	95,17	90,86	92,41	93,01
Local features + BoW model	86,47	83,46	82,37	84,97

Table 16: Average evaluation measures for each implemented method.

As it may be observed, the performance of clustering-based method surpasses the results obtained by thresholding-based methods. The two-region k-means approaches outperform the sensitivity and specificity of even the most developed implemented thresholding technique, the multithresholding Otsu method. Only the k-means approach based on the *Lab* colour space achieve accuracy and specificity values lower than the multithresholding technique.

This reflects the paramount importance of spatial information at segmentation of dermoscopic images. All the implemented clustering-based methods employ pixel location information regarding the center of the image, since the distribution of lesions, compact or expanded, is centered and they spread from this point over the image. Based on this premise, the separation of skin and lesion regions is facilitated.

Thus, the multi-region adaptation of the k-means algorithm leads to better performance. Not only spatial information and the highest-achiever combination of colour components (S and V) are being considered for this technique, but also the detection of the most suitable number of regions within the dermoscopic image regarding the local maxima in the luminance histogram. The extraction of the number of regions in the image was also applied to multithresholding, but it is its combination with information about pixel distribution the aspect that leads to considerable segmentation improvement.

With respect to the method based on the extraction and description of local features and the bag-of-words model, by matching each keypoint with the corresponding lesion or skin word contained in a 32/32 visual codebook, the contour of the lesion can be easily discerned within the generated binary mask, but the high amount of noise on the skin regions show the relevance of considering structural image information within the segmentation process. The SIFT descriptor describes each keypoint in the luminance space respecting certain properties of invariability, but location knowledge about the keypoints is lost in the procedure and mismatches occur along the image.

Regarding the evaluation metrics obtained from the application of this method to the database, the average accuracy is the lowest among the implemented methods, but the achieved trade-off between sensitivity and specificity show certain potential response of this approach towards the segmentation problem and may create the possibility of application in this field of such successful classification techniques.

On the other hand, the implemented methods based on deformable models show two different tendencies according to the principles of their functioning. Both gradient vector flow (GVF) snakes and the Chan-Vese method used as initialization the generated segmentation of the highest-performer method, which is the multi-region k-means approach. Therefore, they may be considered as a refinement tool to adjust the obtained lesion contour to the real boundaries and achieve better evaluation metrics.

The Chan-Vese method outperforms the GVF snakes regarding all the metrics. GVF snakes compute the GVF additional external force, which drives the contour toward the edges, from the gradient of a generated edge map. The variability of textures and the presence of external and intrinsic artefacts in dermoscopic images hinder the proper functioning of edge operators and, consequently, of GVF snakes. Besides, the enormous dependency of GVF snakes on the numerous parameters that govern the contour deformation may also prevent the contour from a correct convergence.

Nevertheless, the Chan-Vese method, which bases its functioning on an advanced level sets formulation and does not require edges information, leads the initial contour to a more exact final position. This technique is less sensitive to contour initialization and achieves the adjustment of curves which are initialized relatively far from the real boundaries. The main handicaps of the method are related to excessive-detailed contours that exclude certain lesional areas from the final lesion region, as well as contours that do not reach the real boundaries after the iterations within the deformation process.

Despite of these handicaps, the best results were obtained by the combination of the multi-region k-means clustering and the Chan-Vese method as a refinement tool. As it may be observed in the previous table, if the original lesion contour from the multi-region k-means segmentation is employed as initialization for the Chan-Vese deformation, there is a general increase of the evaluation metrics, achieving the highest global performance among all the implemented methods. When the points within the original curve are reduced and a simpler initial contour is generated, the highest sensitivity value is obtained to the detriment of the specificity and accuracy values. Since the sensitivity, also referred as to true positive rate, is clinically the most significant value and a satisfactory trade-off between it and specificity is maintained, this solution can also be considered to have a high performance.

6. Future work

After the conclusive analysis of the evaluation results, it is important to take into account the main identified handicaps that have hindered the segmentation of dermoscopic images, regardless of the implemented method, and that should be especially addressed at future modifications of the implemented methods in order to improve the performance.

Illumination inconsistency, which highly depends on the device which was used to take the image, supposes a constant difficulty for all methods due to its considerable influence in the division of regions and, therefore, the separation of the lesion from the background skin. A method for addressing illumination inconsistency in dermoscopic images has been proposed in this project, but it is open to additional adaptations in order to solve further cases.

Besides, those images which contain areas or entire lesions with low contrast between lesional and non-lesional skin complicate the generation of a binary mask because the boundaries of the lesion area are hard to determine within the image. These images suppose a common factor of wrong segmentations among the implemented methods and specialised analysis should be considered as future work, since amelanotic melanomas, which are characterized by the lack of melanin, are included in this group of problematic images.

In addition to general modifications focused on the solution of the mentioned main segmentation handicaps, particular modifications of the implemented methods may also be proposed. Regarding clustering-based methods, the addition of other relevant pixel features, along with the already included radial distance and representative colour components, may be considered in order to improve the convergence of pixels inside the cluster of their correspondent image region.

With respect to the implemented first approach of segmentation based on the extraction of local features and the bag-of-words model, several modifications could be introduced. A mathematical adaptation of the Harris Laplace detector to colour space may provide the algorithm with relevant information in different components, so that the colour descriptors performance could minimally improve. Additionally, the keypoints used as seeds to the k-means clustering employed to generate both lesion and skin visual vocabularies could be especially selected instead of randomly selected, in order to avoid noisy and non-representative keypoints from being used as seeds. Besides, the introduction of modifications to the superpixel-level approach, such as alternative soft-assignment techniques, may improve the performance of this interesting adaptation of the method.

Finally, also the addition of procedures within the system and the evaluation of new algorithms are crucial in order to accomplish optimal segmentations. There exist several potential segmentation methods that could be adapted to dermoscopic images once their drawbacks are conveniently analysed in order to assure performance improvement. It is the case of several methods included in Appendix A, such as hierarchical clustering, the introduction of the SIFT descriptor into the deformation procedure of active contours or the adaptation of active shape models.

7. Planning and methodology

The project was divided in several phases of development, which facilitated the organization regarding its complexity and six months of duration. Since the fundamental premise is the comparison of automatic segmentation methods in order to find the optimal one, it was hard to estimate the duration of each of the phases at the beginning. Each implemented method required specific tools and depended on alternatives of design and certain modifications to a greater or lesser extent in order to be adapted to dermoscopic images. Therefore, the chronological planning and development were based on objectives instead of periods of time.

The first phase consisted in a general overview of the automatic segmentation techniques within the state of art in order to make one first informed selection among those methods which could be potentially adapted to the problem, considering technical drawbacks as well as those limitations imposed by the dermoscopic images in the database.

The next phase was the implementation of those segmentation methods which offered a more direct solution, such as thresholding-based techniques, so as to have a first contact with the database and analyse global segmentation difficulties and requisites. According to the acquired first evaluations, new methods belonging to different groups of techniques were incorporated into the system, such as clustering-based segmentation. Modifications were also proposed for a better adaptation to the problem and considerable good performances were obtained. It is the case of the multi-region k-means clustering approach.

In the following phase of the research, methods with specific and more advanced formulations were considered. Deformable models were selected among other candidates. The methods based on this technique required an exhaustive analysis of the parameters and variables that govern the process, as well as the study of the complex deformation procedure itself.

In the next phase the segmentation problem was led to an experimental scope with the adaptation of techniques that are successfully applied in other computer vision areas, such as image classification and patterns recognition. The development of methods based on the extraction of local features and the bag-of-words model was carried out as a last contribution to the comparative of methods, with the goal of including less explored segmentation techniques.

The last phase consisted in a general final evaluation of the implemented methods in order to obtain definitive results and statistics to complete the comparison of automatic segmentation methods applied to dermoscopic images, which included multiple techniques from direct methods and their adapted approaches to advanced models and experimental implementations. This allowed the extraction of final conclusions in order to analyse the optimal methods among the implemented ones, the main detected segmentation handicaps and proposed future lines of work.

On the other hand, the following methodology was common to every implemented method within the comparative:

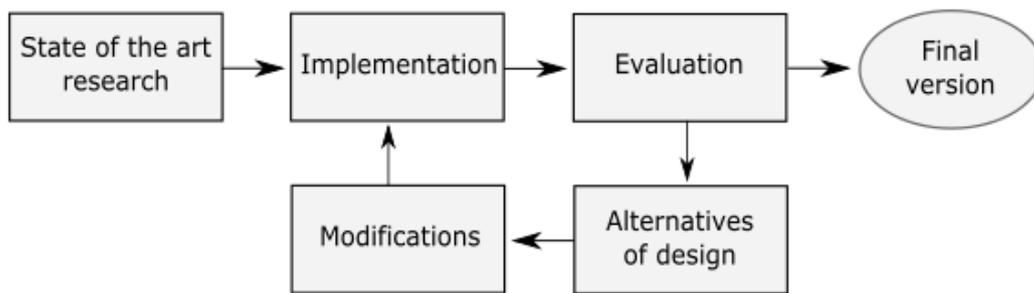


Diagram 17: Methodology per implemented method

Each method required the study of its formulation as well as its pros and cons along specialised literature within the state of the art. Then a first implementation of the method was carried out. After the extraction of first evaluation results, possible alternatives of design are analysed in order to obtain improved segmentation performance. Therefore, particular modifications were implemented, so that the system was better adapted to the properties of dermoscopic images. Then several tests on the system and additional modifications are executed, until the evaluation of performance shows sufficient refinement, i.e., a proper functioning allows the generation of a final version of the method and the extraction of definitive evaluation metrics to be introduced in the comparative. Additionally, this final version was let open to system improvements and new design suggestions.

8. Socio-economic background

8.1 Budget

In this section the budget corresponding to the implementation of the project “A Comparative Study of Algorithms for Automatic Segmentation of Dermoscopic Images” is indicated. The budget includes the costs related to the material which has been necessary for the development of the system, as well as the costs associated with the participants.

8.1.1 Material costs

Material costs are divided into two types of items, hardware and software.

Material	Units	Unit Cost (euros)	Amortization period (years)	Amortized Cost (euros)
Desktop PC	1	500	4	125
Portable PC	1	650	4	162,5
Display	1	100	4	25
Printer	1	110	5	22
External Hard Drive	1	50	5	10
Expendable office supplies	--	75	--	75
TOTAL				419,5

Table 17: Material costs (hardware)

Material	Units	Cost (Euros)
MATLAB Standard License	1	2000
Image Processing Toolbox	1	1000
Computer Vision System Toolbox	1	1250
Microsoft Office	1	70
TOTAL		4320

Table 18: Material costs (software)

8.1.2 Personnel costs

The duration of this project is 6 months. The development of the project is based on 22 working days per month, with one junior engineer working an average of 3 hours per day and one senior engineer as a support during the different phases of the project.

The personnel costs indicated in the following table include indirect costs related to use of infrastructure and other charges.

Category	Hours	Hourly wage (euros)	Cost (euros)
Junior Engineer	420	16,5	6930
Senior Engineer	50	30	1500
TOTAL			8430

Table 19: Personnel costs

8.1.3 Total project cost

The total cost of the project (including material, both hardware and software, and personnel costs) is showed in the following table. The 21% of Value Added Tax (VAT) must be added to the total cost in order to obtain the final cost of the project, which is included at the end of the table.

Item description		Cost (euros)
Material	Hardware	419,5
	Software	4320
Personnel		8430
Total (without VAT)		13169,5
VAT (21%)		2765,6
TOTAL (included VAT)		15935,1

Table 20: Total project cost

The total budget for this project is the amount of FIFTEEN THOUSAND NINE HUNDRED THIRTY-FIVE EUROS AND TEN CENTS.

8.2 Socio-economic environment

At present, cancer is one of the most common causes of natural death. Among the different types of cancer, skin cancer has become one of the most habitual ones. Malignant melanoma causes the majority of deaths among skin cancer types. Therefore, diverse tools and techniques are being developed over the last years in order to help its prevention and cure. These tools assist health professionals in the diagnosis and analysis of skin lesions in order to apply immediate measures which facilitate the clinical procedure. As a consequence, it leads simultaneously to the improvement of the required treatment and economic and time savings. This covers the main current necessities of the “med-tech” market: the reduction of clinical costs and the obtaining of the best solution that satisfies all the involved factors within the procedure, which are health professionals, suppliers and patients.

In medicine it is of paramount importance the early detection of diseases in order to effectively apply the convenient therapies, overtaking complicated or even irreversible situations. Dermoscopic image processing, with lesion segmentation as the first phase within the procedure, contributes to the early detection of melanomas, providing dermatologists with an easy and intuitive tool to use. With a growing appearance of skin cancer due to several factors, this type of diagnosis support becomes essential so as to complement the traditional diagnosis procedure, which consists in the non-error-free visual exploration of the lesion. An automatic diagnosis procedure means an increment of objectivity and reliability, including the finding and analysis of lesion features that are not recognisable to the naked eye. The process, after the segmentation of lesion regions, evaluates certain lesion features, such as colour, texture and shape properties, in order to reach a conclusive diagnosis of the lesion and avoid intrusive medical procedures when possible, such as biopsies.

Automatic dermatological analysis techniques not only aim to facilitate the diagnosis process, but also to achieve a globalised diagnosis, i.e., a uniform protocol to determine if a cutaneous lesion is benign or malign, regardless of the medical center and the health professional. Additionally, since the procedure is exclusively based on the segmentation and classification of dermoscopic images taken from the patient, it facilitates a remote medical treatment, which is especially beneficial in rural environments or places that are difficult to access, by means of teledermatology techniques.

On the other hand, in a digitized society, the processing of dermoscopic images does not only benefit from new devices to capture images, such as high-definition cameras or even smartphones, and fast computation procedures. It also benefits from the growing generation of global image databases, which are easily accessible by broadband networks that provide the medical system with the transmission of high-resolution images and enormous amounts of associated data in real time. These global databases are meant to be stored by cloud computing and constantly available from multiple places. They allow dermatologists to access numerous dermoscopic images, their segmentations and their main parameters already analysed and classified, in order to automate the diagnosis of new entering images. The higher the amount of available data and images, the more precise the results.

References

Text references

1. Argenziano, G., Soyer, H., De Giorgi, V. (2012). *Dermoscopy: A Tutorial*. EDRA Medical Publishing & New Media. Milan. [ONLINE]. Available at: <http://dermoscopy.org/atlas/base.htm>. [Last accessed 26 August 2016].
2. Celebi, M. E., Wen, Q., Iyatomi, H., Shimizu, K., Zhou, H., & Schaefer, G. (2015). A state-of-the-art survey on lesion border detection in dermoscopy images. *Dermoscopy Image Analysis*, 97-129.
3. Mayer, J. (1997). Systematic review of the diagnostic accuracy of dermatoscopy in detecting malignant melanoma. *The Medical Journal of Australia*, 167(4), 206-210.
4. Kumari, R., Sharma, N. A Study on the Different Image Segmentation Technique. *International Journal of Engineering and Innovative Technology (IJEIT) Volume, 4*.
5. Silveira, M., Nascimento, J. C., Marques, J. S., Marçal, A. R., Mendonça, T., Yamauchi, S., ... & Rozeira, J. (2009). Comparison of segmentation methods for melanoma diagnosis in dermoscopy images. *IEEE Journal of Selected Topics in Signal Processing*, 3(1), 35-45.
6. Bogo, F., Peruch, F., Belloni, A., Peserico, E. (2015). Where's the Lesion?: Variability in Human and Automated Segmentation of Dermoscopy Images of Melanocytic Skin Lesions. *Dermoscopy Image Analysis*, 67-95.
7. Soyer, H., Argenziano, G., Menzies, S., Pehamberger, H., Rabinovitz, H. Stolz, W. (1998). *Two-Steps Procedure for the Dermoscopic Classification of Pigmented Skin Lesions: A Unifying Concept*. [ONLINE]. Available at: <http://www.dermoscopy.org/consensus/tutorial.asp>. [Last accessed 31 August 2016].
8. Ó Conaire, C., O'Connor, N. E., Cooke, E., & Smeaton, A. F. (2006). Detection thresholding using mutual information. VISAPP 2006 - International Conference on Computer Vision Theory and Applications.
9. Sahoo, P. K., Soltani, S. A. K. C., & Wong, A. K. (1988). A survey of thresholding techniques. *Computer vision, graphics, and image processing*, 41(2), 233-260.
10. Otsu, N. (1975). A threshold selection method from gray-level histograms. *Automatica*, 11 (285-296), 23-27.
11. Kittler, J., & Illingworth, J. (1986). Minimum error thresholding. *Pattern recognition*, 19(1), 41-47.
12. Jansing, E. D., Albert, T. A., & Chenoweth, D. L. (1999). Two-dimensional entropic segmentation. *Pattern Recognition Letters*, 20(3), 329-336.
13. Varaprasad Raju, G. S., Prasad, G. S., & Srikanth, Y. (2013). Fusion approach for region based image segmentation using k-means clustering. *Computer Science & Telecommunications*, 38(2).

14. Kaur, N., & Jindal, G. (2013). A survey of K means Clustering with modified gradient magnitude region growing technique for lesion segmentation. *International Journal of Innovations in Engineering and Technology (IJJET)*, 2(2).
15. Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R., & Wu, A. Y. (2002). An efficient k-means clustering algorithm: Analysis and implementation. *IEEE transactions on pattern analysis and machine intelligence*, 24(7), 881-892.
16. Han, J., & Kamber, M. (2006). *Data mining: Concepts and techniques* (2nd ed.). San Francisco: Morgan Kaufmann Publishers.
17. Arifin, A. Z., & Asano, A. (2006). Image segmentation by histogram thresholding using hierarchical cluster analysis. *Pattern Recognition Letters*, 27(13), 1515-1521.
18. *Region Growing*, 149-150. [ONLINE] Available at: <http://homepages.inf.ed.ac.uk/rbf/BOOKS/BANDB/LIB/bandb5.pdf>. [Last accessed 28 August 2016].
19. Kandwal, R., Kumar, A., & Bhargava, S. (2014). Review: existing image segmentation techniques. *International Journal of Advanced Research in Computer Science and Software Engineering*, 4(4), 153-156.
20. Dougherty, G. (Ed.). (2011). *Medical image processing: techniques and applications*. Springer Science & Business Media.
21. Jain, R., Kasturi, R., & Schunck, B. G. (1995). *Machine vision* (Vol. 5), Chapter 5. New York: McGraw-Hill.
22. Xu, C., Pham, D. L., & Prince, J. L. (2000). Image segmentation using deformable models. *Handbook of medical imaging*, 2, 129-174.
23. Mendonça, T., Ferreira, P. M., Marques, J. S., Marcal, A. R., & Rozeira, J. (2013, July). PH 2-A dermoscopic image database for research and benchmarking. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 5437-5440). IEEE.
24. Xu, C., & Prince, J. L. (1998). Snakes, shapes, and gradient vector flow. *IEEE Transactions on image processing*, 7(3), 359-369.
25. Zhao, J., Chen, B., Sun, M., Jia, W., & Yuan, Z. (2013). Improved algorithm for gradient vector flow based active contour model using global and local information. *The Scientific World Journal*, 2013.
26. Liu, L., & Bovik, A. C. (2012). Active contours with neighborhood-extending and noise-smoothing gradient vector flow external force. *EURASIP Journal on Image and Video Processing*, 2012(1), 1-6.
27. Erkol, B., Moss, R. H., Joe Stanley, R., Stoecker, W. V., & Hvatum, E. (2005). Automatic lesion boundary detection in dermoscopy images using gradient vector flow snakes. *Skin Research and Technology*, 11(1), 17-26.
28. Chan, T. F., & Vese, L. A. (2001). Active contours without edges. *IEEE Transactions on image processing*, 10(2), 266-277.
29. *Mathworks Documentation: Active Contours*. [ONLINE] Available at: <http://es.mathworks.com/help/images/ref/activecontour.html>. [Last accessed 30 August 2016].

30. Cootes, T. F., Taylor, C. J., Cooper, D. H., & Graham, J. (1995). Active shape models-their training and application. *Computer vision and image understanding*, 61(1), 38-59.
31. *Mathworks Documentation: Local feature Detection and Extraction*. [ONLINE]. Available at: <http://es.mathworks.com/help/vision/ug/local-feature-detection-and-extraction.html>. [Last accessed 30 August 2016].
32. Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2), 91-110.
33. Barata, C., Marques, J. S., & Rozeira, J. (2013, July). Evaluation of color based keypoints and features for the classification of melanomas using the bag-of-features model. In *International Symposium on Visual Computing* (pp. 40-49). Springer Berlin Heidelberg.
34. Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., ... & Van Gool, L. (2005). A comparison of affine region detectors. *International journal of computer vision*, 65(1-2), 43-72.
35. *VLFeat Tutorial: Covariant feature detectors*. [ONLINE]. Available at: <http://www.vlfeat.org/overview/covdet.html>. [Last accessed 31 August 2016].
36. Yang, M., Yuan, Y., Li, X., & Yan, P. (2011). Medical Image Segmentation Using Descriptive Image Features. In *BMVC* (pp. 1-11).
37. Van De Sande, K., Gevers, T., & Snoek, C. (2010). Evaluating color descriptors for object and scene recognition. *IEEE transactions on pattern analysis and machine intelligence*, 32(9), 1582-1596.
38. Van Gemert, J. C., Geusebroek, J. M., Veenman, C. J., & Smeulders, A. W. (2008, October). Kernel codebooks for scene categorization. In *European conference on computer vision* (pp. 696-709). Springer Berlin Heidelberg.
39. Coates, A., & Ng, A. Y. (2012). Learning feature representations with k-means. In *Neural Networks: Tricks of the Trade* (pp. 561-580). Springer Berlin Heidelberg.
40. Beis, J. S., & Lowe, D. G. (1997, June). Shape indexing using approximate nearest-neighbour search in high-dimensional spaces. In *Computer Vision and Pattern Recognition, 1997. Proceedings, 1997 IEEE Computer Society Conference on* (pp. 1000-1006). IEEE.
41. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., & Süsstrunk, S. (2012). SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE transactions on pattern analysis and machine intelligence*, 34(11), 2274-2282.
42. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., & Süsstrunk, S. (2010). *Slic superpixels* (No. EPFL-REPORT-149300).
43. Vedaldi, A. *VLFeat Features Library: Simple Linear Iterative Clustering (SLIC)*. [ONLINE]. Available at: <http://www.vlfeat.org/api/slic.html>. [Last accessed 31 August 2016].
44. Wang, X., Hänsch, R., Ma, L., & Hellwich, O. (2014, January). Comparison of different color spaces for image segmentation using graph-cut. In *Computer Vision Theory and Applications (VISAPP), 2014 International Conference on* (Vol. 1, pp. 301-308). IEEE.
45. *Mathworks: Representing color with the Lab color space*. [ONLINE]. Available at: <http://es.mathworks.com/discovery/lab-color.html>. [Last accessed 1 September 2016].

46. Gonzalez, R. C., & Woods, R. E. (2008). *Digital image processing, 3rd Edition* (Ch. 9, pp.652-662). Nueva Jersey.
47. Wen, Q., Ming, D., Chen, J., & Liu, W. (2013, October). A superpixel based post-processing approach for segmenting dermoscopy images. In *Advanced Computational Intelligence (ICACI), 2013 Sixth International Conference on* (pp. 155-158). IEEE.
48. Gómez, D. D., Butakoff, C., Ersboll, B. K., & Stoecker, W. (2008). Independent histogram pursuit for segmentation of skin lesions. *IEEE Transactions on Biomedical Engineering*, 55(1), 157-161.
49. Joel, G., Philippe, S. S., David, G., Jean Philippe, C., Ralph, B., Joakim, K., ... & Murat, K. (2002). Validation of segmentation techniques for digital dermoscopy. *Skin Research and Technology*, 8(4), 240-249.
50. Celebi, M. E., Schaefer, G., Iyatomi, H., Stoecker, W. V., Malters, J. M., & Grichnik, J. M. (2009). An improved objective evaluation measure for border detection in dermoscopy images. *Skin Research and Technology*, 15(4), 444-450.
51. *MathWorks Products: MATLAB*. [ONLINE]. Available at: <http://es.mathworks.com/products/matlab/>. [Last accessed 3 September 2016].
52. Xu, C., & Prince, J. L. *Active Contours, Deformable Models and Gradient Vector Flow: GVF Software, Matlab (2D Version)*. [ONLINE]. Available at: <http://www.iacl.ece.jhu.edu/static/gvf/>. [Last accessed 3 September 2016].
53. Vedaldi, A., Fulkerson, B. The VLFeat Team. *VLFeat open source. Download: using from MATLAB*. [ONLINE]. Available at: <http://www.vlfeat.org/install-matlab.html>. [Last accessed 3 September 2016].
54. Van De Sande, K. University of Amsterdam. *ColorDescriptor software*. [ONLINE]. Available at: <http://koen.me/research/colordescriptors/>. [Last accessed 3 September 2016].
55. Brodersen, K. H., Ong, C. S., Stephan, K. E., & Buhmann, J. M. (2010, August). The balanced accuracy and its posterior distribution. In *Pattern recognition (ICPR), 2010 20th international conference on* (pp. 3121-3124). IEEE.
56. *Mathworks Documentation: k-means clustering*. [ONLINE]. Available at: <http://es.mathworks.com/help/stats/kmeans.html>. [Last accessed 7 September 2016].
57. Gutkowitz-Krusin, D., Elbaum, M., Szwaykowski, P., & Kopf, A. W. (1997). Can early malignant melanoma be differentiated from atypical melanocytic nevus by in vivo techniques? *Skin Research and Technology*, 3(1), 15-22.
58. *Mathworks Documentation: fill image regions and holes*. [ONLINE]. Available at: <http://es.mathworks.com/help/images/ref/imfill.html>. [Last accessed 8 September 2016].

Figure references

- I. Argenziano, G., Soyer, H., De Giorgi, V. (2012). *Dermoscopy: A Tutorial*. EDRA Medical Publishing & New Media. Milan. [ONLINE]. Available at: <http://dermoscopy.org/atlas/base.htm>. [Last accessed 26 August 2016].
- II. Scikit-image development team. *Image processing in Python*. [ONLINE]. Available at: <http://scikit-image.org>. [Last accessed 28 August 2016].
- III. Arifin, A. Z., & Asano, A. (2006). Image segmentation by histogram thresholding using hierarchical cluster analysis. *Pattern Recognition Letters*, 27(13), 1515-1521.
- IV. Jain, R., Kasturi, R., & Schunck, B. G. (1995). *Machine vision* (Vol. 5), Chapter 5. New York: McGraw-Hill.
- V. Xu, C., Pham, D. L., & Prince, J. L. (2000). Image segmentation using deformable models. *Handbook of medical imaging*, 2, 129-174.
- VI. Xu, C., & Prince, J. L. (1998). Snakes, shapes, and gradient vector flow. *IEEE Transactions on image processing*, 7(3), 359-369.
- VII. *Mathworks Documentation: Image Classification with Bag of Visual Words*. [ONLINE] Available at: <http://es.mathworks.com/help/vision/ug/image-classification-with-bag-of-visual-words.html>. [Last accessed 30 August 2016].
- VIII. *Code Project Recipe*. [ONLINE]. Available at: <http://www.codeproject.com/KB/recipes/619039/SIFT.JPG>. [Last accessed 31 August 2016].
- IX. Van De Sande, K. *ColorDescriptor software for object and scene categorization*. [ONLINE]. Available at: <http://www.colordescriptors.com>. [Last accessed 31 August 2016].
- X. Medialooks. *Advanced Chroma Key settings*. [ONLINE]. Available at: <https://www.medialooks.com/mformats/docs/CK%20Advanced.html>. [Last accessed 1 September 2016].
- XI. *Mathworks Documentation: Convert from HSV to RGB Color Space*. [ONLINE]. Available at: <http://es.mathworks.com/help/images/convert-from-hsv-to-rgb-color-space.html>. [Last accessed 1 September 2016].
- XII. Agudo, J. E., Pardo, P. J., Sánchez, H., Pérez, Á. L., & Suero, M. I. (2014). A low-cost real color picker based on Arduino. *Sensors*, 14(7), 11943-11956.
- XIII. Cid Sueiro, J., Mora Jiménez, I., Díaz de María, F. *Tratamiento Digital de la Imagen: Procesado morfológico*. Universidad Carlos III de Madrid.
- XIV. Menzies, S. W., Crotty, K. A., Ingvar, C., & McCarthy, W. H. (2009). *Dermoscopy: An Atlas*. Sydney, 3rd Edition. Australia: McGraw-Hill.
- XV. Celebi, M. E., Wen, Q., Iyatomi, H., Shimizu, K., Zhou, H., & Schaefer, G. (2015). A state-of-the-art survey on lesion border detection in dermoscopy images. *Dermoscopy Image Analysis*, 97-129.
- XVI. Smith, S. W. (1997). *The scientist and engineer's guide to digital signal processing*, Ch. 24, *Linear Image Processing: Convolution*.

- XVII. *VLFeat Tutorial: SIFT detector and descriptor*. [ONLINE]. Available at: <http://www.vlfeat.org/overview/sift.html>. [Last accessed 1 September 2016].
- XVIII. *VLFeat Tutorial: SLIC superpixels*. [ONLINE]. Available at: <http://www.vlfeat.org/overview/slic.html>. [Last accessed 1 September 2016].
- XIX. *Mathworks Documentation: Morphological structuring element*. [ONLINE]. Available at: <http://es.mathworks.com/help/images/ref/strel-class.html>. [Last accessed 3 September 2016].

Regulatory references

- A. *Spanish Constitution. Chapter 2: Rights and liberties. Section 18.1: Honour and personal reputation.* (BOE n. 311, 29 December 1978). [ONLINE]. Available at: http://www.congreso.es/portal/page/portal/Congreso/Congreso/Hist_Normas/Norm/const_espa_texto_ingles_0.pdf. [Last accessed 3 September 2016].
- B. *Ley Orgánica 1/1982, de 5 de mayo, de protección civil del derecho al honor, a la intimidad personal y familiar y a la propia imagen.* (BOE n. 115, 14 May 1982, pp. 12546-12548). [ONLINE] Available at: <https://www.boe.es/buscar/doc.php?id=BOE-A-1982-11196>. [Last accessed 3 September 2016].
- C. *Ley Orgánica 15/1999, de 13 de diciembre, de protección de Datos de character Personal.* (BOE n. 298, 14 December 1999). [ONLINE] Available at: <https://www.boe.es/buscar/act.php?id=BOE-A-1999-23750>. [Last accessed 3 September 2016].
- D. Primary Care Commissioning. (2013). *Quality Standards for Teledermatology, Standard 3 (Gaining the patient's informed consent), Standard 7 (Information governance and record-keeping).* United Kingdom.

Appendix A: Further studies on state of the art

A.1 Colour spaces

These three colour spaces have been selected for the implementation of some of the methods included in the project:

- **RGB.** The *RGB* colour space is most commonly used in computer systems, televisions, video, etc. It is easy to implement, but it is device dependent and non-linear with visual perception [13]. It is represented by red (*R*), blue (*B*) and green (*G*) chromaticities. The final colour is defined by the additive combination of those three primary colours [44].

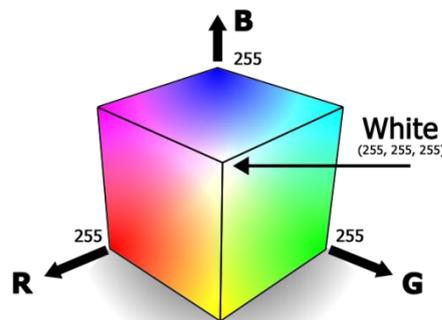


Figure 55: Representation of RGB colour space [X]

- **HSV.** The *HSV* colour space separates the intensity from the chromaticity and represents them independently. Hue (*H*) describes the position of colour in a 360° spectrum. Hue changes within angle variation. Saturation (*S*) describes pureness of colour. It is equivalent to the distance from the axis, i.e., it measures the difference between the colour and a greyscale value of equal intensity. The value (*V*) is the measurement of brightness and corresponds to distance along the axis [44].

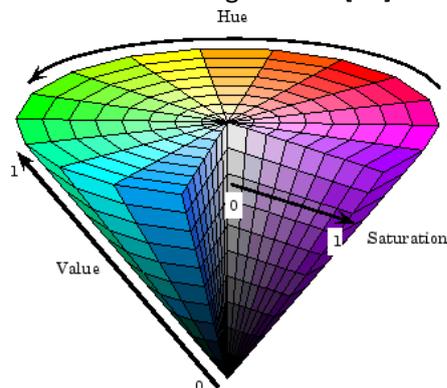


Figure 56: Representation of HSV colour space [XI]

- **Lab.** The *Lab* colour space is considered the most exact means of representing colour and is device independent. It includes all colours in the spectrum, as well as colours outside of human perception. It is a colour system based on three axes. The lightness dimension (*L*) represents the luminosity layer, while *a* and *b* are two colour-opponent

dimensions. The chromaticity layer a indicates where colour falls along the red-green axis, whereas b represents where colour falls along the blue-yellow axis [45].

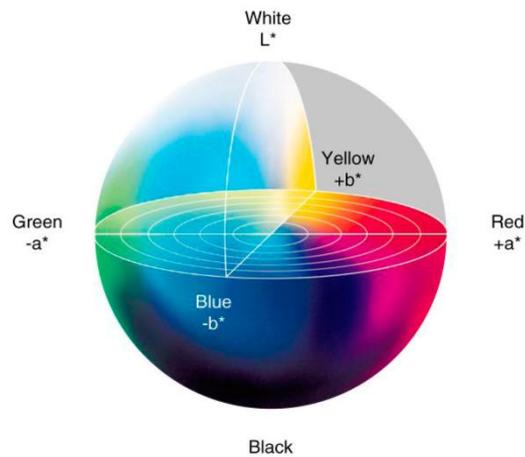


Figure 57: Representation of Lab colour space [XII]

A.2 Region-growing techniques

The goal of **region growing** is to use image features to map individual pixels to sets of pixels called regions. An image region might correspond to a real object or a meaningful part of one. The geometric characteristics of each region depend on the domain. Usually they are considered to be two-dimensional connected areas. But it is because the selected region-growing technique that regions can be disconnected, have smooth boundaries or be non-simply connected, that is, have holes. The simplest methods use only aggregates of properties of local groups of pixels to determine regions, whereas more sophisticated methods grow regions by merging more primitive regions [18].

Region growing can be implemented in four steps [19]:

1. A group of seed particles in the original image is selected.
2. A set of criteria for determining similar seeds and a stopping rule are established.
3. A region grows by adding to each seed those neighbouring pixels that have predefined properties which are similar to the seed pixel.
4. When there are no more pixels that satisfy the criterion for inclusion in the region, the region growth is stopped.

Automated seed selection does not always guarantee complete segmentation due to the unpredictability of seed point selection and might therefore require manual addition of extra seed points after initial segmentation [20]. This aspect could reduce the automation of the segmentation phase inside the dermoscopic analysis procedure.

Split and merge algorithms remove the need for subjective seed point selection. In these techniques, instead of selecting a set of initial seed points, the image is subdivided into a set of arbitrary unconnected regions and then these regions are merged or split according to the segmentation criteria [20].

A.3 Hierarchical clustering

Hierarchical clustering techniques are based on the use of a proximity matrix indicating the similarity between every pair of data points to be clustered. The result is a tree of clusters, a dendrogram, which represents the nested grouping of patterns and similarity levels at which groups change. The procedure of these methods are based both on merging smaller clusters into larger ones (bottom-up approach) or by splitting larger clusters (top-down approach). The differences among methods lay on how two small clusters are merged or a large cluster is split. One disadvantage of hierarchical clustering with respect to partitional clustering is that in the former one, once a cluster is assigned, it cannot be altered [14].

One interesting approach is **image segmentation by combining histogram thresholding and hierarchical clustering analysis** [17]. This method generates a dendrogram of gray levels, based on a similarity measure that involves the inter-class variance of the clusters to be merged and the intra-class variance of the merged new cluster. Therefore, it is bottom-up approach whose aim is a good separation of the clusters and a robust estimation of a threshold, so that a clear separation of object and background is achieved from unimodal to multimodal histograms. The estimated threshold for a usual two-level thresholding is achieved by iterating. In each iteration the similarities between adjacent clusters are computed and then the most similar pair is merged.

The hierarchical tree of this unification process can be viewed as a dendrogram. The height of the links between clusters represents the order of merging operations. The estimated thresholds for the t -level thresholding are obtained by dividing the dendrogram in t groups, cutting the branches. At two-level thresholding the threshold is obtained by cutting the highest branch of the dendrogram.

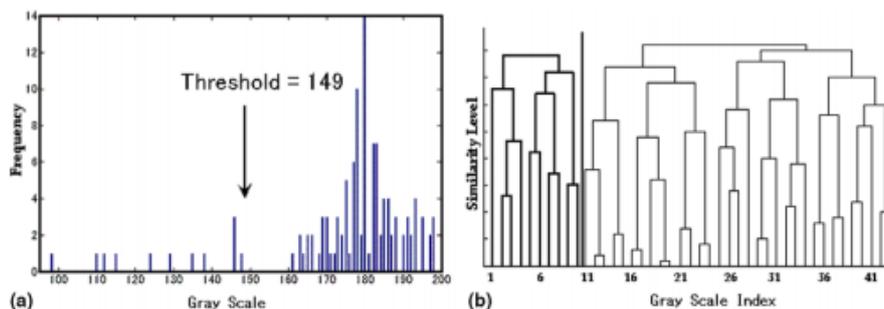


Figure 58: Histogram of image and the obtained dendrogram [III]

At the beginning of the merging process, each cluster is assigned to each original gray level in the histogram. Then the distance between adjacent clusters is computed and the pair of the smallest distance is found and unified into one cluster. The proposed distance definition is based on both the difference between the means of the two clusters and the variance of the resultant cluster after the merging.

Let $h(z)$, $z = 0, 1, \dots, L-1$, be the histogram of the image, where z indicates the gray level and L the number of available gray levels including the empty ones. Regarding the histogram as a

probability density function, $p(z)=h(z)/L$ is regarded as the probability of occurrence of one pixel with gray level z . The following function indicates the occurrence probability of pixels belonging to the cluster C_k :

$$P(C_k) = \sum_{z=T_{k-1}+1}^{T_k} p(z), \quad \sum_{k=1}^K P(C_i) = 1$$

The distance between clusters C_{k1} and C_{k2} is given by:

$$Dist(C_{k1}, C_{k2}) = \sigma_1^2(C_{k1} \cup C_{k2}) \sigma_A^2(C_{k1} \cup C_{k2})$$

where $\sigma_1^2(C_{k1}, C_{k2})$ and $\sigma_A^2(C_{k1}, C_{k2})$ are the inter-class variance and the intra-class variance, respectively, which are defined as:

$$\begin{aligned} \sigma_1^2(C_{k1} \cup C_{k2}) &= \frac{P(C_{k1})}{P(C_{k1}) + P(C_{k2})} [m(C_{k1}) - M(C_{k1} \cup C_{k2})]^2 \\ &+ \frac{P(C_{k2})}{P(C_{k1}) + P(C_{k2})} [m(C_{k2}) - M(C_{k1} \cup C_{k2})]^2 \\ &= \frac{P(C_{k1})P(C_{k2})}{(P(C_{k1}) + P(C_{k2}))^2} [m(C_{k1}) - m(C_{k2})]^2, \end{aligned}$$

$$\sigma_A^2(C_{k1} \cup C_{k2}) = \frac{P(C_{k1})}{P(C_{k1}) + P(C_{k2})} \times \sum_{z=T_{k1-1}+1}^{T_{k2}} [(z - M(C_{k1} \cup C_{k2}))^2 p(z)].$$

where $m(C_k)$ denotes the mean of cluster C_k and is defined as:

$$m(C_k) = \frac{1}{P(C_k)} \sum_{z=T_{k-1}+1}^{T_k} z p(z)$$

and $M(C_{k1} \cup C_{k2})$ denotes the global mean of the clusters C_{k1} and C_{k2} , given by:

$$M(C_{k1} \cup C_{k2}) = \frac{P(C_{k1})m(C_{k1}) + P(C_{k2})m(C_{k2})}{P(C_{k1}) + P(C_{k2})}$$

In the last iteration of the algorithm, if applied for t -level thresholding, C_1, C_2, \dots, C_t clusters are obtained and highest gray levels of the clusters are the estimated thresholds T_1, T_2, \dots, T_{t-1} .

In order to adapt this method to the analysis of dermoscopic images, it is necessary an effective convergence criterion for the algorithm, that is, a procedure to select conveniently for every lesion the number of final clusters in which the dendrogram must be divided. The histograms from dermoscopic images are rarely simply bimodal, so it is crucial to define when the algorithm must stop before the hierarchical clustering begins. This criterion should take into account how the algorithm computes the similarity between adjacent clusters. Another aspect to take into account is that this method lacks of spatial information of pixels because it generates the dendrogram only from the gray levels distribution. Perhaps a modification of the distance measurement by introducing average spatial information of the clusters could adapt the calculation of similarities between adjacent clusters to those lesions with inner colour-variegated areas.

A.4 Edge detection operators

The description of each detector is based on the exhaustive analysis included in [21].

A.4.1 Gradient operators

In one dimension, a *step* edge is associated with a local peak in the first derivative. The gradient is a measure of change in a function. Since an image can be considered as an array of samples of some continuous function of image intensity, significant changes in the gray values of the image can be detected by using a discrete approximation to the gradient. The gradient is the two-dimensional equivalent of the first derivative and is defined as the following vector:

$$G[f(x, y)] = \begin{bmatrix} G_z \\ G_y \end{bmatrix} = \begin{bmatrix} \frac{\delta f}{\delta z} \\ \frac{\delta f}{\delta y} \end{bmatrix}$$

There are two important properties associated with the gradient. First, the previous vector points in the direction of maximum rate of increase of the function $f(x, y)$. Second, the magnitude of the gradient equals the maximum rate of increase of $f(x, y)$ per unit distance in the direction G and it is common to approximate it by absolute values:

$$G[f(x, y)] = \sqrt{G_z^2 + G_y^2} \rightarrow G[f(x, y)] \approx |G_z| + |G_y|$$

For digital images, the simplest gradient approximation is approximated by differences:

$$G_z \cong f[i, j + 1] - f[i, j]$$

$$G_y \cong f[i, j] - f[i + 1, j]$$

The objective is to compute the gradient along the image and detect those points where the gradient acquires a relevant magnitude. An approach is to use a 3x3 neighbourhood and calculate the gradient in the center pixel. Several methods provide an approximation to the gradient magnitude by applying convolution mask around a central pixel in order to extract vertical and/or horizontal borders. It is the case of **Roberts**, **Sobel** and **Prewitt** operators. These detectors compute the first derivative and, if it is above a fixed threshold, the presence of an edge point is assumed. This results in the detection of too many edge points.

A.4.2 Second derivative operators

An approach to diminish the quantity of detected false edges is to consider as edge points only those points that have local maxima in gradient values, so that at edge points there is a peak in the first derivative and, equivalently, there is a zero crossing in the second derivative. Therefore, edge points may be detected by finding the zero crossings of the second derivative of the image intensity.

The most common operator that corresponds in two dimensions to the second derivative is the **Laplacian operator**. The Laplacian is the two-dimensional equivalent of the second derivative. For a function $f(x, y)$ the Laplacian is given by:

$$\nabla^2 f = \frac{\delta^2 f}{\delta x^2} + \frac{\delta^2 f}{\delta y^2}$$

The second derivatives along the x and y directions are approximated using difference equations and then centered about the pixel $[i, j]$:

$$\frac{\delta^2 f}{\delta x^2} = f[i, j + 1] - 2f[i, j] + f[i, j - 1]$$

$$\frac{\delta^2 f}{\delta y^2} = f[i + 1, j] - 2f[i, j] + f[i - 1, j]$$

By combining these two equations into a single operator, a 3x3 neighbourhood convolution mask can again be generated and applied at every central pixel, which usually is given more weight than the rest of pixels in the neighbourhood.

The main disadvantage of the second derivative operators is that edge points detected by finding the zero crossings of the second derivative of the image intensity are very sensitive to noise.

A.4.3 Laplacian of Gaussian

Due to the noise sensitivity of edge points detected by second derivative operators, it is crucial to filter the noise before the step of edge enhancement. To do this, the Laplacian of Gaussian, due to Marr and Hildreth, combines Gaussian filtering with the Laplacian for edge detection.

In this approach, the image should be convolved with a Gaussian filter. This step smoothes an image and reduces noise: isolated noise points and small structures are filtered out. Then the enhancement step is the Laplacian, approximation of the second derivative in two dimensions. The detection criterion is the presence of a zero crossing in the second derivative with a corresponding large peak in the first derivative. To avoid detection of insignificant edges, only the zero crossings whose corresponding first derivative is above some threshold are selected as edge points. Then the edge location can be estimated with subpixel resolution using linear interpolation.

A.4.4 Gaussian edge detection

As mentioned before, *step* edges in real images are not perfectly sharp due to the smoothing introduced by the low-pass filtering inherent in the optics of camera lenses and the bandwidth limitations in camera electronics. These aspects are added to the presence of noise in the image and result in a trade-off between noise suppression and edges localization at the implementation of detectors.

The type of linear operator that provides the best compromise between noise immunity and localization, while retaining the advantages of Gaussian filtering is the first derivative of a Gaussian. This operator is a combination of a Gaussian smoothing filter and a gradient approximation that can be numerically approximated using the standard finite-difference approximation for the first partial derivatives in the x and y directions, which has been previously listed in the Gradient operators section. The operator is symmetric along the edge

and antisymmetric perpendicular to the edge, that is, it is sensitive to the edge in the direction of steepest change, but is insensitive to the edge in the direction along the edge while acts as a smoothing operator.

The **Canny edge detector** is the first derivative of a Gaussian and closely approximates the operator that optimizes the product of signal-to-noise ratio and localization. The first step is smoothing the image with a Gaussian filter. The second step consists in computing the gradient magnitude and orientation using finite-difference approximations for the partial derivatives. The third step is the nonmaxima suppression to the gradient magnitude, which is a process that thins the ridges of the computed gradient magnitude by suppressing all values along the line of the gradient that are not peak values of a ridge. The four and last step consists in a double thresholding algorithm to detect and link edges.

The typical procedure to reduce the number of false edges is selecting an individual threshold, but it is difficult to identify the proper threshold and involves trial and error in order to accomplish a balance between the presence of false edges and the absence of relevant portions of contours. A more effective thresholding scheme uses two thresholds, τ_1 and τ_2 , with $\tau_2 \approx 2 \tau_1$ to produce two thresholded edge images $T_1[i, j]$ and $T_2[i, j]$. Image T_2 is formed with a higher threshold, so it contains fewer false edges but may have gaps in the contours due to the false negatives. The double thresholding links the edges in T_2 into contours. When it reaches the end of a contour, the algorithm looks in T_1 at the locations of the 8-neighbours for edges that can be linked to the contour. The algorithm continues to gather edges from T_1 until the gap is covered and jointed to an edge in T_2 . This procedure of edge linking resolves some of the problems with choosing a unique threshold.

A.5 Modifications of GVF algorithm

A proposed modification of GVF snakes in order to obtain a better adaptation to dermoscopic images is the one proposed in [27]. After generating an initial curve by thresholding technique and interpolation of points along the edge of the segmented lesion region, the GVF process is divided in two parts. First, a specific Gaussian filtering of the gradient map is performed and the curve starts to deform until a reduced number of iterations is completed. The deformation is then stopped and the snake is supposed to have moved closer to the lesion boundaries. Second, this time a smaller Gaussian kernel is used to filter the image, since there is less likelihood that the snake converges to noise points in the surrounding skin regions. The snake then deforms during a larger number of iterations until its final convergence.

Another alternative to the standard procedure of GVF snakes is the one proposed in [25]. The external force field is first computed using the global information of GVF. Then the initial contour is driven towards the boundary object gradually. When the active contour stops, the local information of it is employed to help determine the false part of the contour. If the false part exists, the magnetostatic force of it is computed and then the external force field is updated to move the active contour for further time. The magnetostatic active contour model hypothesizes that electric currents flow through both object boundary and active contour;

thus, the active contour is moved toward object boundary under the influence of the generated magnetic field. The described steps are repeated until the active contour stops and there is no false part on it.

Regarding the GVF algorithm's sensitivity to noise and difficulties to preserve weak edges, a neighbourhood-extending and noise-smoothing method is proposed in [26]. The proposed snake expresses the gradient vector flow as a convolution with a neighbourhood-extending Laplacian operator augmented by a noise-smoothing mask.

A.6 Active shape models

The study included in [30] proposes that flexible models, such as snakes, can be useful in situations where the objects of the same class are not identical. For instance, in medical applications, the shape of organs can vary considerably through the time of medical treatment and between individuals. It is also the case of melanomas, which evolve within time and their shape changes clearly from one patient to another.

The active shape models propose a method that by learning patterns of variability from a training set of annotated images is able to capture the natural variability within a class of shapes. The key difference with respect to the active contour models is that active shape models only allow deformations which are specific to the class of structures under study; the models can only deform in ways found in the training set.

The technique relies on a set of points that represent each object or the image structure. The points can represent the boundary, internal and external features. Points are placed in the same way on each example of the object within the train set. The sets of points are automatically aligned to minimize the variance in distance between equivalent points. The model provides then the average positions of the points and the main modes of variation found in the training set can be controlled by a set of parameters.

The adaptation of the active shape model to the segmentation of dermoscopic images would suppose the search of common patterns of shape among lesions in order to establish a set of points that would adapt the deformation of the model to the specific purpose. But lesions show very different combinations of shapes and textures. One possible approach would be the creation of different groups of lesions within the training set depending on some global shape features, so that a set of points is extracted for each group and the deformation process is adjusted to each shape pattern. Another approach would be extracting the set of points from intrinsic features of the lesions, since, as it has been mentioned, the points can also represent internal features and not only boundaries.

A.7 Modifications of SIFT descriptor

When SIFT is applied to medical image analysis, it is possible to combine contour and shape models with the extraction and description of features along the boundaries of the object, as proposed in [36]. Weak boundaries as well as strong edges in the neighbouring regions of the

true boundaries may difficult the computation of the exact main orientations of SIFT features. In addition, the surrounding anatomic structures of different contour parts can vary significantly. The inconsistency among areas may reduce the descriptive ability of the extracted features when being trained together. Therefore, eliminating the feature variation according to the anatomical structures surrounding each contour point is a meaningful task in order to allow the contour to evolve correctly. Instead of using the general gradient-based and region-based features, a new SIFT based feature called normal vector feature profile (NVFP) is proposed. This descriptor aims to describe the local image appearance of the contour points along its normal vector to the boundary. Thus, for building NVFP and eliminating the variations among features extracted on the contour, a proper number and locations of sample points should be determined.

Appendix B: Cross-validation procedure

Cross-validation is a technique whose aim is to guarantee that the obtained results from a procedure are independent from the partition of the database, i.e., the obtained optimal parameters are independent from the set of images that has been used for their determination and may lead to statistically valid conclusions.

The procedure is applied in different validation phases within the project and is consequently adapted to each validation. In this section the general implementation is described.

All the images and their manual masks within the database are randomly distributed in n equally-sized groups. Then, n train and n test sets are generated. Each test set is equivalent to one of the n groups that have been previously created, whereas each train set contains the remaining $n-1$ groups.

Therefore, there are n processing combinations. Each combination is based on one train set over which the validation is carried out and optimal parameters are obtained, and one test set that is subsequently processed with those obtained parameters.

At the validation process, several combinations of values for the validated parameters are assessed. A validation range of values is selected, as well as the increasing step between validated values. In each iteration, the specific operations that depend on the validated parameters are applied to each item within the train set. The accuracy of the resulting mask is analysed regarding the manual mask. The mean accuracy for every combination of values applied to all the automatic masks within the train set is stored inside a matrix.

When the last iteration is complete, the highest accuracy value within the matrix of mean accuracies is located along with its correspondent indexes, which are equivalent to the optimal parameters.

Appendix C: Summary

C.1 Introduction

Melanoma is the most frequent and one of the most lethal types of skin cancer, but it is also among the most curable cancers when it is early identified. The traditional diagnosis procedure bases its conclusions exclusively on observation and certain visual rules regarding lesion features such as colour or asymmetry. Two main handicaps are associated with this procedure: subjectivity, since different conclusions from the same lesion may be reached if different professionals carry out the study, and the application of intrusive measures due to errors in diagnosis.

In order to improve its detection and increase survival rates for malignant melanoma, dermoscopy is integrated in the diagnosis procedure. It is a non-invasive imaging technique that consists in placing oil, alcohol or water on the skin lesion. This fluid eliminates surface reflection and allows an enhanced visualization of skin lesions, so that significant patterns that cannot be detected to the naked eye become recognisable. There exists a regulatory framework around the dermoscopic images obtained from patients by means of the dermoscopy technique, mainly related to the informed consent, i.e., the patient must allow the taking and storage of images and, at the same time, his/her privacy and the protection of personal data must be assured.

The computer-aided diagnosis of skin lesions aims to support conventional diagnosis and reduce its drawbacks. The standard approach in dermoscopic image analysis has three steps: 1) image segmentation, 2) feature extraction and selection, 3) lesion classification. The first step of the process, i.e. the segmentation of the lesion, consists in the classification of all points in the image as part of the lesion or simply part of the surrounding non-lesional skin. It is crucial for image analysis due to two main requirements. First, the border structure serves as relevant information for diagnosis, since many clinical features, such as asymmetry, border irregularity or abrupt border cutoff are studied directly from the boundaries. Second, the extraction of other important clinical features, such as atypical pigment networks, globules, blue-white areas or regression structures, critically depends on the accuracy of the lesion segmentation.

Diverse tools and segmentation techniques are being developed over the last years in order to help the prevention of malignant melanoma. In our study, different methods from the broad state of the art were selected in order to be analysed, implemented and compared, along with introduced modifications and also proposed innovative techniques. The following division for the selected segmentation techniques is proposed:

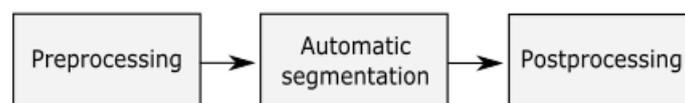
- 1. Threshold-based segmentation.** These algorithms determine one or more proper threshold values to divide the pixels of an image into several classes and separate objects from the background.

2. **Region-based segmentation.** These methods segment an image into regions that are similar according to a set of predefined criteria.
3. **Edge-based segmentation.** These methods involve the detection of edges, important local changes in the intensity of an image, between the regions using edge operators.
4. **Segmentation based on deformable models.** These algorithms involve the detection of object contours using curve evolution techniques.
5. **Segmentation based on the bag-of-words model.** These experimental segmentation methods, already applied in the extraction of image features and image classification, are based on the detection and description of keypoints within the image and the posterior generation of a visual codebook.

Many of the methods included in this section, regardless of the segmentation techniques they are based on, take advantage of the properties that different colour spaces may offer.

The objective is to implement a system that achieves the automatic segmentation of dermoscopic images, in order to separate the lesion region from the non-lesional areas in the image. The performance of the different implemented methods has been evaluated regarding the manual masks, which are considered as ground truth and denote the real boundaries for each lesion. A comparative study of the obtained results has been carried out, with the purpose of finding the most adequate method among the implemented ones.

The design of the implemented system follows the conventional structure of the automatic segmentation phase inside the computer-aided diagnosis of dermoscopic images. Three major steps constitute the system:



C.2 Preprocessing and postprocessing of dermoscopic images

The preprocessing and postprocessing phases are common to all the implemented automatic segmentation methods, with only specific exceptions.

The first step of the system consists in a preprocessing procedure, in order to reduce the impact of external artefacts within the dermoscopic image as well as intrinsic cutaneous features in the segmentation phase. It is based on two procedures. Firstly, a morphological preprocessing is applied to eliminate a common external artefact along the images within the database: black frames. Then, a Gaussian smoothing operation is applied in order to achieve a balance between blurring of undesirable features and conservation of edges.

In the last step before evaluation, once the automatic mask is generated, it enters the postprocessing step. The borders are smoothed and undesirable spurious lesion pixels or holes inside the lesion region are eliminated, by means of opening and closing morphological operators; the optimal size of the radiuses of both structural elements are conveniently cross-

validated. Then, an expansion of the lesion region is carried out, in order to reduce the discrepancy between the generated automatic masks and the ground truth. The outcome of the system is an automatic generated mask that can be evaluated regarding the manual mask considered as ground truth, so that the performance of the applied segmentation algorithm can be analysed.

C.3 Automatic segmentation of dermoscopic images

The second and decisive step of the system is the automatic segmentation phase, which is specifically designed for each implemented method. Several algorithms based on different state-of-the-art segmentation techniques have been implemented in order to achieve an efficient segmentation of dermoscopic images, taking into account the global and specific limitations for the implementation. Every algorithm includes the introduced modifications for its better adaptation to the specific segmentation problem of dermoscopic images.

C.3.1 Threshold-based methods

The first approach is the application of the **traditional Otsu method**, which is a point-dependent thresholding technique, since a threshold is determined from the gray level of each pixel. The traditional version of the method consists in global thresholding because the entire image is thresholded by one threshold value. The optimal threshold value that is computed maximizes the separability of the resultant classes in gray levels.

Compact lesions with high contrast between lesion and skin achieve good segmentation results. However, when the luminance histogram shows several peaks with several valleys in between, which correspond to colour-variegated lesion areas, this distribution complicates the segmentation of the lesion region by computing a unique threshold and a more complex method might be required. In those lesions which contain either inner regression or blue-white veil areas, several zones are excluded from the lesion region too.

The second approach is based on the **local Otsu method**, in which the image is divided into smaller sections and a threshold is computed for each partition individually. This results in gray level discontinuities at the boundaries of each two partitions. Thus, a smoothing algorithm is required in order to reduce these discontinuities. The local application of the Otsu method on rectangular concentric partitions was selected over a grid approach, since the smoothing of discontinuities between partitions was facilitated by the number of image divisions and their distribution. Each image is divided in three concentric rectangular areas and the Otsu method is applied on each one separately. As a result, there is a binarization of the external, middle and inner areas, which are then concatenated.

The local implementation of the Otsu method does not show appreciable improvements regarding the performance of the traditional approach. Segmentation problems related to low contrast between lesion and skin areas, as well as the presence of cutaneous features such as regression or external features such as hairs, are not solved. Besides, additional segmentation

irregularities that were not present before have emerged for those images with strong skin patterns and illumination inconsistency.

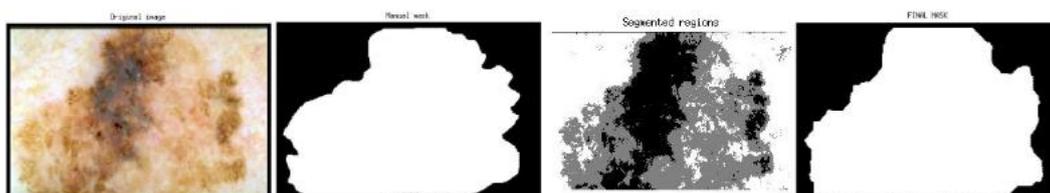
The third and last approach within this category is the **multithresholding Otsu**. The main advantage is that it allows a more accurate segmentation of those lesions that contain several colour/texture-variegated areas. Those areas can be individually detected by extracting several threshold values from the luminance histogram. This way the segmentation of complex images that cannot be easily divided into two classes is more complete and exact.

The suitable number N of regions within the image must be previously defined in order to compute $N-1$ threshold values and quantize the image regarding those values. The method proposed in this project to determinate the number of regions into which the image must be segmented is based on the analysis of local maxima in the histogram of luminance image and is described below. After image quantization, it is necessary to merge all regions considered as lesion into one.

- **Method to determine the number of regions within a dermoscopic image based on the analysis of local maxima in the luminance histogram**

Optimal values for the minimum peak distance and the minimum peak height along the dermoscopic images within the database were empirically determined. In a normalized luminance histogram the height of a maximum is equivalent to the percentage of occurrence of a certain gray level, so a minimum percentage of occurrence and a minimum difference in luminance are established in order to consider peaks as relevant local maxima. If these minimum values are not surpassed due to dispersed levels of luminance, the minimum height is decremented. After relevant local maxima in the normalized luminance histogram are extracted regarding the previous parameters, an extra limitation between close adjacent maxima is imposed in order to reduce the influence of non-representative maxima and a posterior oversegmentation of the image.

The performance of the multithresholding approach surpasses the previous results for both expanded and colour-variegated lesions.



Multithresholding Otsu method. Expanded, colour-variegated lesion. From left to right: original image, manual mask, segmented regions, final mask.

Nevertheless, the problem for images with low contrast between lesion and skin is still present. Regarding the performance at the presence of artefacts, this adaptation of the algorithm is largely resistant to cutaneous features and external features. The main handicap of this algorithm, as a threshold-based technique, is the lack of spatial information when the optimal thresholds are computed. Therefore, the segmentation of those images which present illumination inconsistency may be still unstable.

C.3.2 Region-based methods

The selected technique for region-based methods, k-means algorithm, is a partitioning clustering technique that requires a priori specification of the number of clusters, that is, it is necessary to introduce the number of final regions into which the image is divided. If the accordant number of regions for each image is not known, the first and most general approach to dermoscopic image segmentation is the **two-region k-means clustering**, based on the determination of two initial classes: one for lesion regions and one for skin regions.

The first step consists in the generation of a matrix that contains the features of each pixel in the image, which will be analysed by the algorithm in order to assign each pixel its nearest centroid. Therefore, regarding the values of its features, a pixel will join the cluster associated with lesion regions or the one associated with non-lesional skin. This design contemplates two main pixel features: colour components and location within the image.

Different colour spaces have been considered and three different versions of the algorithm were implemented, each one with a specific combination of colour components:

- The *R*, *G* and *B* components from the *RGB* colour space.
- The *S* and *V* components from the *HSV* colour space.
- The *L*, *a* and *b* components from the Lab colour space.

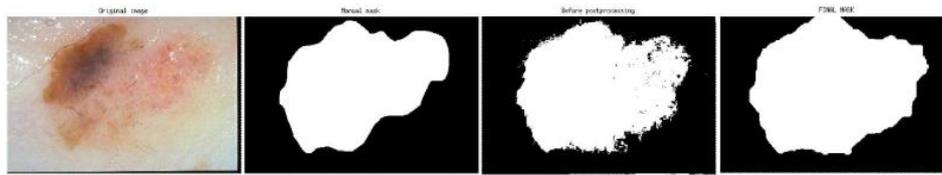
In dermoscopic images the particular location of pixels within the image may contain significant information about the distribution of lesion and skin regions. Regarding the importance of spatial information, the location of each pixel is an additional feature included in the matrix that is common to all versions of the algorithm. Since the prevailing distribution of the lesion area in dermoscopic images goes from the center towards the borders of the image, the location assigned to every pixel is based on the modulus of radial distance from the pixel to the center of the image.

The next step is the creation of the initialization matrix. K-means algorithm is highly sensitive to initialization. The initialization matrix contains the data of the two initial centroids: centered in the image and middle values for the colour components in the case of the lesion cluster centroid; located on the image corner and also middle values for the colour components in the case of the skin cluster centroid.

The third step is the application of the k-means algorithm, introducing the number of final clusters (one for lesion and one for skin), the feature matrix and the initialization matrix. After the iterative procedure of the algorithm is finished because the convergence condition has been reached, the final centroids and a binary mask are obtained.

The most noticeable improvement of k-means clustering with respect to threshold-based segmentation is the inclusion of lesion areas with low pigmentation inside the final lesion region, due to the spatial information used by the algorithm.

The best results among the two-region k-means method were obtained by the approach based on radial distance and *S*, *V* components. The exclusion of the *H* component is due to the distorted information it provides.



Two-region k-means clustering (radial distance, S , V). Low contrast between lesion and skin. From left to right: original image, manual mask, mask before postprocessing, final mask.

However, it is still necessary the division of the image into more than two regions, since the resultant masks of many lesions, especially those that contain several colour-variegated areas and complex features, do not change with respect to those obtained from direct thresholding binarization. Therefore, the second approach within region-based segmentation is based on **multi-region k-means clustering**. The method proposed in this project for the determination of a suitable number of regions within the image is based on the analysis of local maxima in the histogram of luminance image, which it is also used before the application of multithresholding algorithms and is described above.

Once the method is applied, the total number of representative local maxima is considered as the number of clusters to be introduced in the k-means algorithm. If only one representative maximum has been detected, the number of clusters is automatically two. Because of the obtained results, the implementation whose feature matrix contains the radial distance of each pixel to the center of the image along with the components S and V from the HSV colour space is the one applied.

The initialization matrix must be now adapted to the determined number of clusters. The values of S and V for each initial cluster centroid are again 0.5, but the value of their radial distance is now dynamically applied regarding the number of clusters, i.e., each centroid is assigned a value at a proportional distance from the center.

After the k-means algorithm converges, a mask with several classes is obtained. The applied initialization of the clusters guarantees a stable distribution of classes and makes possible to merge all regions that belong to the lesion into one.

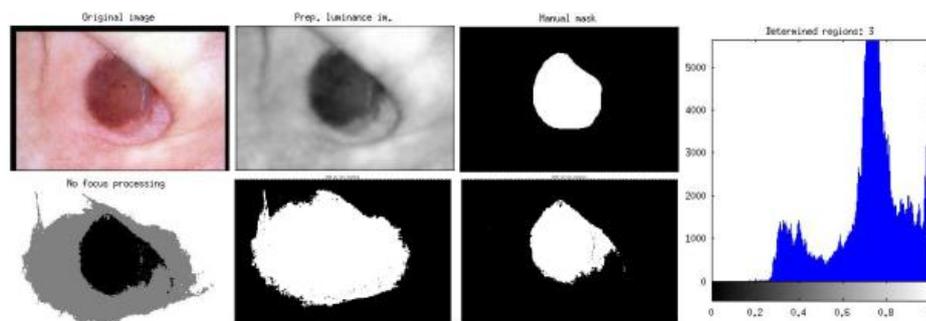
Since dermoscopic images depend on the device that was used to take the image, it was observed that changes in illumination may be a common issue that needs specific processing in the application of this approach, which is addressed in the last step of the procedure.

- **Method to address illumination inconsistency in dermoscopic images**

The first step of the method is called focus analysis and has two branches. Since the most common consequence of illumination inconsistency is the apparition in the generated mask of an ellipsoidal region around the real lesion, one branch is based on the analysis of the shape of the lesion region, by means of extracting its eccentricity, solidity and orientation of horizontal axis properties. If certain values, according to a standard ellipse shape, for the three parameters are satisfied, then the focus condition is positive for this branch. An additional condition is imposed on the other branch: the highest local maximum belongs to the skin region, since the focus area covers a compact

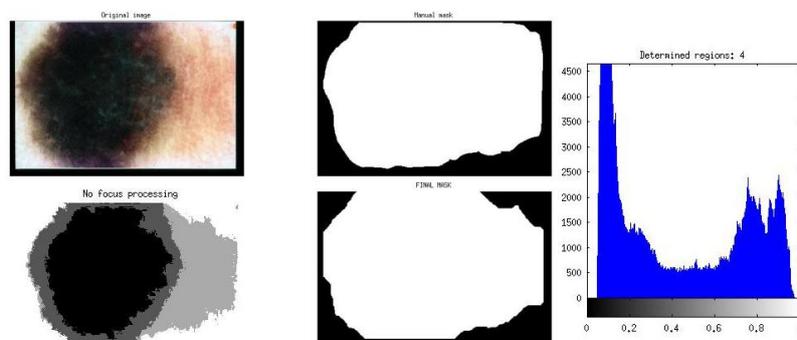
lesion in the center and a big area of skin in the surroundings. If the two branches satisfy the imposed focus conditions, the focus processing is applied.

In the focus processing, a bounding-box which only contains the focus area is extracted from the original image. Then, within the bounding-box, the local maxima are extracted following the previously described procedure and the multi-region k-means algorithm is applied, as well as the merging of the obtained lesion regions. After the focus processing, the ellipsoidal focus within the mask is expected to have disappeared and the new mask is considered as the definitive one. If not, the ellipsoidal area is still present in the mask but its ellipsoidal shape is likely to be smoother. Thus, a second focus analysis is carried out, but this time the imposed conditions on eccentricity, orientation and solidity are less restrictive. The same conditions on local maxima are imposed. Then a second and last focus processing is carried out and the generated mask is now the definitive one.



Multi-region k-means. On top: original image (left), preprocessed luminance image and manual mask (center), luminance histogram (number of determined regions: 3) (right). On bottom: mask before region merging and focus processing (left), mask after first (center) and second (right) focus processing.

Besides the addressing of illumination inconsistency, the multi-region k-means approach maintains the good segmentation performance with those images which contain lesion areas with low pigmentation. Additionally, it introduces important advances regarding the segmentation of colour-variegated lesions and considerable progress is achieved respecting the presence of both external and cutaneous features within the image. The accurate separation of inner lesion regions allows the inclusion of lesional areas that were not detected as lesion in any of the previous methods.



Multi-region k-means clustering. Colour-variegated lesion. On top: original image (left), manual mask (center) luminance histogram (number of determined regions: 4) (right). On bottom: segmented regions (left), final mask (right).

C.3.3 Segmentation based on deformable models

The **gradient vector flow (GVF) snakes** were selected over edge-based methods and traditional snakes, due to their limitations. GVF snakes are a type of parametric deformable models and are based on the energy minimization formulation. In order to find the boundary of the object to be segmented, these curves are initialized within the image domain and are forced to move toward the potential energy minima under the influence of two groups of forces: internal forces, associated with the elasticity and rigidity of the snake, and external forces, which drive the contour toward the edges. GVF snakes employ the gradient vector flow as an additional external force.

The first step is the definition of an edge map derived from the image, having the property of being larger near the image edges. The blue channel (B) from the RGB colour space was selected due to its high entropy value and its good discrimination between lesion and skin. The next step consists in the computation of the GVF field. The GVF field aims to keep the desirable property of the gradients near the edges and at the same time extend the gradient map farther away from the edges and into homogeneous regions. The objective is to increase the capture range of the snake and to achieve a better adaptation to concavities within the boundaries.

The next step within the procedure is the generation of the initial curve. In order to avoid sensitivity to initialization to the extent possible and better adaptation for each lesion, the initial curve is based on the mask obtained from the multi-region k-means algorithm, which provides the highest segmentation performance. Additionally, the lesion contour is then expanded and the number of points that constitute it is decremented, so that the snake is simplified and the computational load is diminished.

The following phase is the deformation of the curve. The curve deformation depends on several parameters, apart from the number of iterations. The main parameters, α and β , regulate the internal forces by controlling the tension and the rigidity of the snake, respectively, and a cross-validation procedure was carried out in order to find their optimal values.

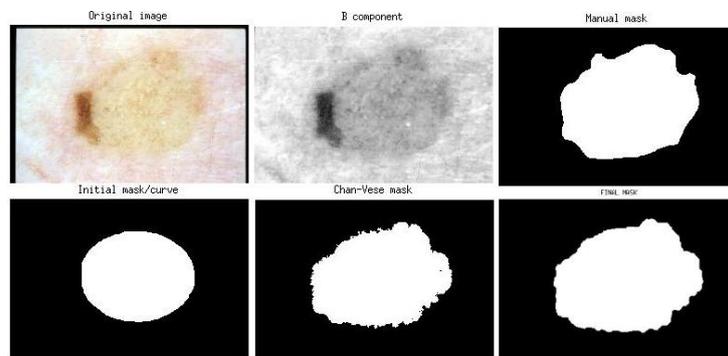
The main handicap observed in the resulting masks from the application of this deformable model is that the snakes do not adjust to the real boundaries precisely. Two particular cases can be differentiated: when the initial curve is not near the real boundaries and when the initial curve is already sufficiently adjusted.

Therefore, a more advanced deformable model was implemented. The **Chan-Vese method**, also referred to as active contours model without edges, is a geometric deformable model, since it is based on the theory of curve evolution and the level set method; the model is supposed to detect boundaries not necessarily defined by gradient.

The B channel was again selected as the definitive input due to its high entropy value and good discrimination between lesion and skin areas. Since the Chan-Vese method stopping term is based on Mumford-Shah segmentation techniques, the model is supposed to detect better the real boundaries. Consequently, no smoothing is applied this time to the gray-level component

before the deformation. The selected initialization for the Chan-Vese method was also the obtained segmentation from the multi-region k-means clustering due to its high performance. Two different approaches of the method were implemented regarding the initial curve: for one of them the original contour was maintained, whereas the number of points within the curve was reduced for the second approach. Since the speed of computation of the algorithm was faster than the GVF model, both approaches were finally considered.

The contours of this model perform without requiring edges information; therefore, a better adjustment of the final curve to the real boundaries can be globally observed, even at the presence of intrinsic features, such as blue-white veil, and external features, such as traces of liquid. When the initial curve is sufficiently close to the real boundaries and its borders are not necessarily well defined, the adjustment is well performed and a more accurate segmentation is achieved, even if there is low contrast between lesion areas and skin.



Chan-Vese method. Final contour well adjusted. On top: original image (left), B component (center), manual mask (right). On bottom: initial mask (left), mask before postprocessing (center), final mask (left).

Both implementations of the Chan-Vese method (initialization with reduced points and the approach that does not reduce the points within the initial curve) show a global improvement of the segmentation performance with respect to the results obtained from the multi-region k-means clustering. Therefore, the Chan-Vese method can be considered as a good option for segmentation refinement.

C.3.4 Segmentation based on the bag-of-words model

The extraction of local features combined with the bag-of-words model supposes powerful means for pattern recognition and image classification. Due to their proper functioning in these two fields, it was decided to introduce as part of the system an experimental segmentation method based on these two techniques. It is considered to be experimental because, to our knowledge, the specialised literature does not include applications of the extraction of local descriptors and bag-words model to image segmentation. Therefore, the conventional process was followed and adapted to the segmentation of dermoscopic images.

The detection of relevant local features is performed on a set of reference images for the posterior generation of a visual vocabulary that allows the recognition of patterns and their classification. When adapted to the segmentation of dermoscopic images, the classification is reduced to two classes: lesion and skin. Thus, the visual codebook contains words for these

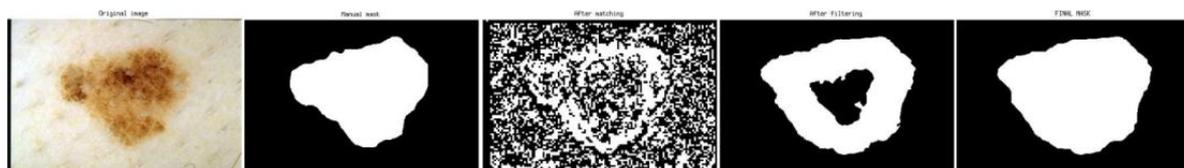
two classes in order to achieve the segmentation of an image that enters the system by matching every keypoints extracted from the image and the closest word inside the vocabulary, which corresponds to lesion region or non-lesional skin.

There is an additional approach for the matching phase and, therefore, the generation of the final mask, that has been implemented for this system, a superpixel-level approach, but it was discarded from a final implementation due to its low performance.

Different results were obtained from different combinations of keypoints detectors (including the Harris Laplace detector), keypoints descriptors (including colour-SIFT descriptors) and visual vocabulary properties (diverse sizes and proportions of lesion/skin words). The definitive implementation of this experimental segmentation method is based on the following combination of steps:

- Dense sampling for the extraction of local features from both the set of reference images and the images which enter the system to be segmented.
- SIFT descriptor for the description of all the detected keypoints.
- K-means clustering for the generation of a visual vocabulary for lesion regions that contains 32 lesion words and a visual vocabulary for skin regions that contains 32 skin words.
- Individual matching of each keypoint with one of the 64 visual words within the codebook (the centroid that minimizes the Euclidean distance), so that it is classified as lesion or skin.
- Creation of binary mask by nearest-neighbour interpolation and application of pillbox filter in order to eliminate the noise on skin areas and compact the lesion region.

The performance of the method is high for centered and compact lesions, whose lesion borders are well discerned from the rest of the image. Some of them are already compact before the postprocessing phase, whereas other lesions need the postprocessing to a greater or lesser extent, mainly for the filling of holes inside the lesion region. Those compact lesions with colour-variegated areas or surrounding lesion areas with low pigmentation can also be segmented due to the strong filtering process. The generated masks are rough and certainly expanded. Despite the elimination of the expansion process inside the postprocessing phase, this causes the specificity value to increase due to the inclusion of healthy skin areas inside the final lesion region.



Segmentation based on dense sampling, SIFT, 32/32 vocabulary. Compact lesion that requires postprocessing to fill holes. From left to right: original image, manual mask, mask after matching, mask after filtering, final mask.

The main problem of this method is associated with the connectivity of lesion borders. The pillbox filter may facilitate the connection of borders in those compact lesions which contain at least some areas with high contrast between lesion and skin. However, for those lesions whose

entire lesion region has low pigmentation, the closing of lesion borders is not accomplished by the filtering process neither the postprocessing phase. Nonetheless, it has been observed that the connectivity of lesion borders improves with the increment of iterations of the k-means clustering employed for the generation of both lesion and skin vocabularies.

C.4 Conclusions and future work

After the corresponding preprocessing phases and the automatic segmentation procedure, the output of each automatic segmentation method was compared with the corresponding manual segmentation in order to evaluate the performance, once the postprocessing phase was completed. The following table summarizes the global average results obtained from the application of each implemented method to the 724 images within the database and facilitates the comparison between techniques regarding the selected evaluation metrics: sensitivity, specificity, accuracy and balanced accuracy.

METHOD	AVERAGE EVALUATION MEASURES			
	SE (%)	SP (%)	AC (%)	BAC (%)
Traditional Otsu	79,64	97,30	90,62	88,47
Local Otsu	81,97	94,23	88,63	88,10
Multithresholding Otsu	85,08	94,44	91,03	89,76
Two-region k-means (RGB)	87,73	95,88	91,51	91,80
Two-region k-means (Lab)	88,59	93,86	90,57	91,23
Two-region k-means (SV)	89,93	95,79	92,03	92,86
Multi-region k-means	91,71	93,35	92,39	92,53
GVF snakes	91,70	90,84	90,96	91,27
Chan-Vese method	93,05	93,18	92,86	93,12
Chan-Vese method (red. points)	95,17	90,86	92,41	93,01
Local features + BoW model	86,47	83,46	82,37	84,97

As it may be observed, the performance of clustering-based method surpasses the results obtained by thresholding-based methods. This reflects the paramount importance of spatial information at segmentation of dermoscopic images. Thus, the multi-region adaptation of the k-means algorithm leads to better performance.

The best results were obtained by the combination of the multi-region k-means clustering and the Chan-Vese method as a refinement tool. As it may be observed in the previous table, if the original lesion contour from the multi-region k-means segmentation is employed as initialization for the Chan-Vese deformation, there is a general increase of the evaluation metrics, achieving the highest global performance among all the implemented methods. When the points within the original curve are reduced and a simpler initial contour is generated, the highest sensitivity value is obtained to the detriment of the specificity and accuracy values. Since the sensitivity, also referred as to true positive rate, is clinically the most significant value and a satisfactory trade-off between it and specificity is maintained, this solution can also be considered to have a high performance.

Regarding the lines of future work, it is important to take into account the main identified handicaps that have hindered the segmentation of dermoscopic images, regardless of the implemented method, and that should be especially addressed. It is the case of illumination

inconsistency and images which contain areas or entire lesions with low contrast between lesional and non-lesional skin. In addition to general modifications focused on the solution of the main segmentation handicaps, particular modifications of the implemented methods may also be proposed, such as adjustments for clustering-based methods, for the first approach of the segmentation based on the bag-of-words model and its superpixel-level approach. Finally, also the addition of procedures within the system and the evaluation of new algorithms are crucial in order to accomplish optimal segmentations.

** All the references related to the contents of Appendix C are included along the different sections within the document.*