



Universidade de Coimbra

Faculdade de Ciências e Tecnologia

Departamento de Engenharia Eletrotécnica e de Computadores

Diogo Martins Duarte Sereno

Automated evaluation of gamete and embryo quality for assisted reproduction

Junho de 2013



UNIVERSIDADE DE COIMBRA



Universidade de Coimbra
Faculdade de Ciências e Tecnologia
Departamento de Engenharia Eletrotécnica e Computadores

Mestrado Integrado em Engenharia Eletrotécnica e Computadores

Automated evaluation of gamete and embryo quality for assisted reproduction

Diogo Martins Duarte Sereno

Júri:

Professor Doutor Paulo José Monteiro Peixoto (Orientador)

Professor Doutor Paulo Jorge Carvalho Menezes (Vogal)

Professor Doutor Jaime Batista dos Santos (Presidente)

Coimbra – Portugal

Junho de 2013

Agradecimentos

Não desejando toldar este testemunho com o tom excessivamente saudosista que sempre tende a acompanhar o final de um ciclo da vida de cada um de nós, não posso contudo deixar de reservar esta breve secção para deixar os meus mais sinceros agradecimentos a quem estes são devidos. Procurarei não enumerar exaustivamente os nomes de todos aqueles que me acompanharam e que contribuíram de uma forma mais ou menos marcante para o percurso que realizei até ao momento, pois condensar a gratidão por todo um conjunto de eventos decorrido ao longo de mais de duas décadas de existência não é tarefa fácil.

Assim sendo, em primeiro lugar deixo o meu reconhecimento a todos os amigos mais próximos e sobretudo aos colegas que ao longo destes últimos cinco anos me seguiram através das venturas e desventuras que foram surgindo.

De seguida estendo os meus mais cordiais agradecimentos ao meu orientador, o Professor Doutor Paulo José Monteiro Peixoto, que foi uma figura de eminente importância no desenvolvimento deste projecto, sendo que o seu suporte foi fulcral para a sua concretização.

Por último, deixo a minha mais profunda gratidão à minha família, sem os esforços da qual eu não estaria hoje a redigir este documento.

Diogo Martins Duarte Sereno

Resumo

A presente dissertação surge como uma abordagem de carácter exploratório face a um projecto de parceria entre o Centro de Neurociências e Biologia Molecular (CNC), os Hospitais da Universidade (HUC) e o Instituto de Sistemas e Robótica (ISR), cujo principal objectivo é o de estabelecer um método de inspecção automática com base em métodos da área de Visão por Computador que permita proceder a uma avaliação mais rigorosa e objectiva de gâmetas e embriões, visando auxiliar as correntes Técnicas de Reprodução Assistida.

Este trabalho consiste num conjunto de etapas sequenciais que permitem com algum rigor fazer uma detecção e classificação de espermatozóides em imagens microscópicas. Procede-se inicialmente a um pré-processamento da imagem alvo de forma a reduzir o ruído indesejado presente nas imagens e a realçar os contornos das estruturas a segmentar. Esta etapa é seguida de um processo de binarização através do cálculo de um limiar simples. São de seguida extraídas as regiões segmentadas através da obtenção das suas caixas envolventes.

Os objectos resultantes são multiplicados por um fator de escala que resulta do quociente da sua área com a de um espermatozóide de referência e manipulados geometricamente através do recurso a um método de enquadramento de elipses de forma a facilitar a sua posterior classificação. É extraída a secção intermédia desta região por intermédio da sua segmentação pelo método de Otsu, removendo o fragmento do objecto posterior à região mais escura obtida pela aplicação deste método.

De seguida é utilizada uma abordagem com base na recolha de parâmetros morfológicos e num sistema de classificação como base em Máquina de Vetores de Suporte (SVM) aplicada sobre a assinatura dos contornos da cabeça com a finalidade de classificá-la morfolologicamente. Este procedimento é seguido por a recolha de um conjunto de parâmetros de forma da secção intermédia para determinar a sua espessura e a possível presença de gotas citoplasmáticas. Procede-se ainda a uma nova segmentação pelo método de Otsu de uma faixa de pixéis em torno do eixo maior da cabeça na imagem processado por forma a determinar a proporção do acrossoma face à zona nuclear.

Para efeitos de seguimento da cauda é utilizado um algoritmo com base em binarização local, interpolação quadrática e cálculo do Índice de Semelhança Estrutural (SSIM) de um conjunto de pixéis face a um pixel central.

São apresentados os resultados para cada um dos métodos da implementação descritos na dissertação. Conclui-se que o sistema sugerido é uma proposta viável para a detecção e classificação de espermatozóides em imagens microscópicas.

Palavras-Chave: Técnicas de Reprodução Assistida, classificação de espermatozoides, segmentação de imagens, SVM, SSIM

Abstract

The current thesis emerges as the result of an exploratory project whose purpose is to provide an efficient, robust and more objective solution for the automated evaluation of gamete and embryo quality to be used along the currently practiced Assisted Reproduction Techniques, a joint project between the Neurosciences and Cellular Biology Center (CNC), the Hospitals of the University of Coimbra and the Institute of Systems and Robotics (ISR).

This work consists of a number of steps that resort to methods based of Computer Vision to segment and classify Spermatozoon present in microscopic images: first we go through a preprocessing stage so as to reduce undesired noise while largely preserving fine detail such as sharp corners, thin lines and curved structures. After this step we proceed by binarizing the image by applying a simple threshold based on the brightness of the pixels of the processed image. After eliminating the smaller objects still present in the binary image, we extract the bounding boxes which contain the head and mid-piece of the various spermatozoa present in the image.

The next step is to scale the various objects as a function of the quotient between the object's and a reference spermatozoon's areas and then readjust their orientation by performing an ellipse fitting of each binary object so as to have them centered and horizontally aligned for an easier comparison. The mid-piece is then cropped through a process based on a segmentation of the region through Otsu's method and the contour signature and additional morphological features are then calculated for each of the extracted heads. The contour signature vectors are then evaluated using an approach based on Support Vector Machine (SVM) so as to classify them morphologically. We then proceed to analyze a number of features including the column histogram of the several mid-pieces to ascertain whether or not there are defects present such as cytoplasm droplets or any sort of abnormal thickness. To finalize this step we retrieve a strip of pixels situated around the major axis of the head and once again apply Otsu's algorithm so as to separate the pixels in two distinct sets that allow us to verify the existence of the acrosome in regards to a specific head and its relative dimension when compared to the nucleus of the spermatozoon.

To identify the points that constitute the sperm's tail we resort to an iterative method based on a quadratic interpolation of the results of a local segmentation and by calculating the SSIM index of a group of potential tail points relatively to the currently selected pixel. Results to each of the methods presented in this thesis are shown at the end of the document in order to

demonstrate their performance. We conclude that the current solution is a viable method for the detection and classification of sperm cells through the use of microscopic digital images.

Keywords: Assisted Reproduction Techniques, sperm cells evaluation, image segmentation, SVM, SSIM

Índice

Agradecimentos.....	i
Resumo	iii
Abstract.....	v
Índice.....	vii
Lista de Tabelas	ix
Lista de Abreviaturas	x
Capítulo 1 – Introdução	1
1.1 - <i>Contextualização</i>	1
1.2 - <i>Objetivos do Trabalho</i>	3
Capítulo 2 – Estado da Arte.....	6
Capítulo 3 – Implementação do Sistema	15
3.1 – <i>Etapa 1 – Tratamento da Imagem</i>	15
3.2 – <i>Etapa 2 – Segmentação</i>	18
3.3 – <i>Etapa 3 – Classificação Morfológica da Cabeça</i>	19
3.3.1 – <i>Obtenção da Região da Cabeça</i>	19
3.3.2 – <i>Análise dos Parâmetros Morfológicos</i>	20
3.3.3 – <i>Sistema de Análise por SVM</i>	22
3.4 – <i>Etapa 4 – Análise do Acrossoma e Secção Intermédia</i>	26
3.5 – <i>Etapa 5 – Identificação e Caracterização da Cauda</i>	28
Capítulo 4 – Análise dos Resultados.....	36
4.1 - <i>Deteção e Classificação das Cabeças</i>	36
4.2 - <i>Cálculo da Dimensão do Acrossoma</i>	41
4.3 - <i>Análise da Secção Intermédia</i>	42
4.4 - <i>Deteção dos pontos da Cauda</i>	45
4.5 – <i>Resultados Gerais</i>	47
Capítulo 5 – Conclusão e Trabalho Futuro	49
Bibliografia	51
Anexo I - Resultados da classificação SVM	57
<i>Imagens Spz Diff-Quik</i>	57
<i>Hemacolor</i>	59
<i>Fotos Hemacolor 15-5-07</i>	61
<i>Diff-Quik</i>	64

Lista de Figuras

Figura 1.1.1- Esquema Representativo de um espermatozóide normal.....	2
Figura 1.1.2- Diferentes anomalias morfológicas.....	2
Figura 1.2.1- Esquema das etapas do sistema de classificação	4
Figura 2.1- Comparação dos resultados de diferentes métodos de segmentação .	9
Figura 2.2- Análise de métodos de extração propostos na literatura.....	10
Figura 2.3- Resultados da segmentação por <i>nth-fusion</i>	11
Figura 2.4- Estrutura e parâmetros de forma de um espermatozóide	11
Figura 2.5- Resultado de um método de pesquisa iterativo de caudas	13
Figura 3.1.1- Exemplos de Imagens Fornecidas	15
Figura 3.1.2- Resultados da Etapa de Tratamento de Imagens	17
Figura 3.2.1- Resultados da Etapa de Segmentação.....	18
Figura 3.3.1.1- Regiões Segmentadas	19
Figura 3.3.1.2- Resultado da Extração das Cabeças.....	20
Figura 3.3.2.1- Imagens de Referência.....	21
Figura 3.3.3.1- Representação do processo de obtenção de um classificador ..	24
Figura 3.3.3.2- Esquema de funcionamento de um sistema SVM Linear.....	24
Figura 3.3.3.3- SVM Não Linear	25
Figura 3.3.3.4- Classificação morfológica com base em SVM.....	26
Figura 3.4.1- Representação do processo de extração da secção intermédia e classificação do acrossoma.....	27
Figura 3.5.1- Algoritmo com base em SSIM para deteção de caudas.....	31
Figura 3.5.2- Resultados esperados da aplicação do algoritmo de deteção.....	33
Figura 3.5.3-Resultados práticos da implementação do algoritmo de referência	34
Figura 3.5.4- Resultados do algoritmo alternativo proposto.....	35
Figura 4.1.1- Imagens a analisar e respetiva segmentação	39
Figura 4.4.1- Resultados da aplicação do algoritmo de deteção de caudas	46

Lista de Tabelas

Tabela 4.1.1- Parâmetros Morfológicos das Cabeças Analisadas	37
Tabela 4.1.2- Resultados da Classificação SVM das Cabeças.....	41
Tabela 4.2.1- Rácio do Acrossoma das Cabeças Analisadas	41
Tabela 4.3.1- Parâmetros Morfológicos das Seções Intermédias Analisadas	42
Tabela 4.4.1- Resultados do Algoritmo de Detecção das Caudas	45
Tabela 4.5.1- Resultados Gerais do Sistema	47
Tabela IV.1- Resultados da Classificação SVM para ‘Spz Diff-Quik’	57
Tabela IV.2- Resultados da Classificação SVM para ‘Hemacolor’	59
Tabela IV.3- Resultados da Classificação SVM para ‘Fotos Hemacolor’	61
Tabela IV.4- Resultados da Classificação SVM para ‘Diff-Quik’	64

Lista de Abreviaturas

<i>HUC</i>	Hospitais da Universidade de Coimbra
<i>CNC</i>	Centro de Neurociências e Biologia Celular
<i>ISR</i>	Instituto de Sistemas e Robótica
<i>TRA</i>	Técnicas de Reprodução Assistida
<i>PMA</i>	Procriação Medicamente Assistida
<i>WHO</i>	World Health Organization (Organização Mundial de Saúde)
<i>pH</i>	Potencial Hidrogeniónico
<i>RGB</i>	Red Green Blue (Vermelho Verde Azul)
<i>PIP</i>	Ponto Inicial de Procura
<i>CASA</i>	Computer-Assisted Sperm Analysis (Análise de Esperma Ass. por Comp.)
<i>ASMA</i>	Automated Sperm Morphology Analyzer (Analisador Automático de Morfologia Espermática)
<i>SQA</i>	Sperm Quality Analyzer (Analisador de Qualidade Espermática)
<i>SCA</i>	Sperm Class Analyzer (Analisador de Classe Espermática)
<i>IOVS</i>	Integrated Visual Optical Systems (Sistemas Integrados Visuais Óticos)
<i>MIR</i>	Maximum Intensity Region (Região de Máxima Intensidade)
<i>EM</i>	Expectation Maximization (Maximização de Expectativas)
<i>MRF</i>	Markov Random Field (Campo Aleatório de Markov)
<i>SSIM</i>	Structural Similarity (Semelhança Estrutural)
<i>SVM</i>	Support Vector Machine (Máquina de Vetores de Suporte)
<i>PDE</i>	Partial Differential Equation (Equação Diferencial Parcial)
<i>RM</i>	Relaxed Median (Mediana Relaxada)
<i>RF</i>	Radial-Basis Function (Função de Base Radial)

Capítulo 1 – Introdução

1.1 - Contextualização

A infertilidade é um grave problema que tem vindo a afetar inúmeros casais nos últimos anos e que continua desafortunadamente a aumentar. A aceção que lhe é atribuída pela Associação Portuguesa de Fertilidade denota que “a infertilidade é o resultado de uma falência orgânica devida à disfunção dos órgãos reprodutores, dos gâmetas ou do conceito” [1], sendo um casal considerado infértil “quando, após um ano de relações sexuais, sem qualquer tipo de método contraceptivo não consegue procriar”, estimando-se que 15% dos casais são actualmente afligidos por este problema [1][2].

Muitos destes casais recorrem frequentemente a clínicas de fertilidade, onde médicos e técnicos especializados recorrem a um conjunto de técnicas avançadas denominadas Técnicas de Reprodução Assistida (TRA) para determinar a origem do problema de forma a solucioná-lo, procurando através delas auxiliar a fertilização e reprodução humana. Estes métodos têm por base uma recolha e observação criteriosa dos gâmetas (células reprodutoras) masculinos e femininos de forma a detetar defeitos que possam advir de problemas fisiológicos e determinar eventualmente as amostras com maior probabilidade de fertilização.

No caso dos gâmetas masculinos, também conhecidos como espermatozóides, esta análise é tipicamente realizada através de um conjunto de diferentes exames clínicos, dos quais se destaca o espermograma. Segundo a Organização Mundial de Saúde (OMS), a qualidade de uma amostra de esperma é avaliada de acordo com um conjunto de parâmetros que podem ser observados num espermograma, dos quais se realçam como tendo maior importância na avaliação da qualidade espermática: a mobilidade, a concentração e a morfologia, segundo relatam Renata Tavares e Sofia Morais em [2] e de acordo com as indicações da OMS [3]. O primeiro refere-se à aptidão que o espermatozóide apresenta para se movimentar. Distinguem-se sob a alçada deste conceito a noção de progressividade e o conceito de densidade móvel que descrevem a capacidade de percorrer uma dada trajectória, e a de se deslocar desde o local onde são depositados até à zona de fertilização, respectivamente. A concentração refere-se ao número de espermatozóides em milhões presentes numa amostra por cada ml de sémen, sendo que para ser considerada viável uma amostra deverá apresentar mais de 20 milhões por cada ml, de acordo com os artigos [3], [4], [5], [6]. Quanto à morfologia, um espermatozóide normal (como representado na **Figura 1.1.1**) deverá apresentar: uma cabeça oval, constituída por uma zona anterior mais clara denominada acrossoma, que deverá ocupar 40% a 70% da cabeça, e uma

região posterior mais escura correspondente ao núcleo; uma peça intermédia de largura superior à da cauda, sem presença de gotas citoplasmáticas e unida axialmente à cabeça; a cauda por sua vez deverá ser aproximadamente recta, uniforme em termos de largura e não enrolada. Segundo o critério de Tygerberg [7], uma amostra é considerada normal quando mais de 15% dos espermatozóides possam ser classificados como normais, caso contrário encontramos perante a presença de um caso de teratozoospermia, uma condição fisiológica que afeta negativamente a fertilidade masculina.

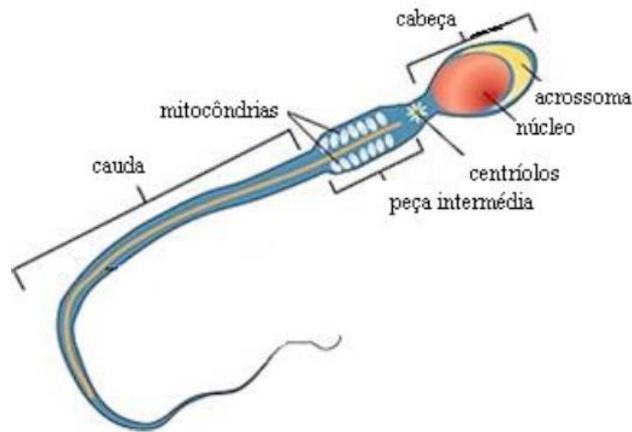
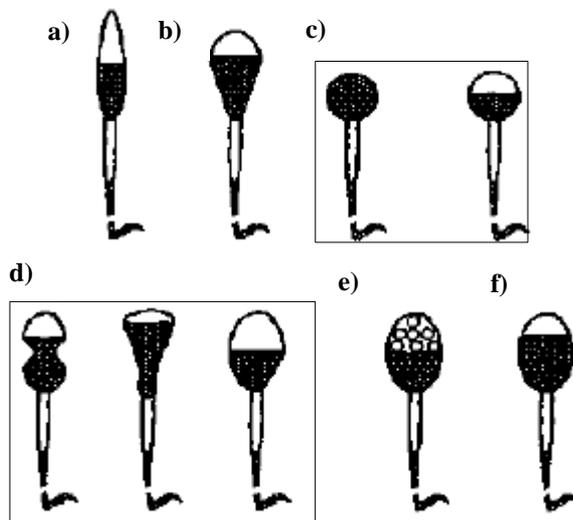


Figura 1.1.1 - Esquema representativo de um espermatozóide normal

Um espermatozóide classificado como anormal poderá apresentar, entre outras, as anomalias demonstradas na **Figura 1.1.2**:

A:



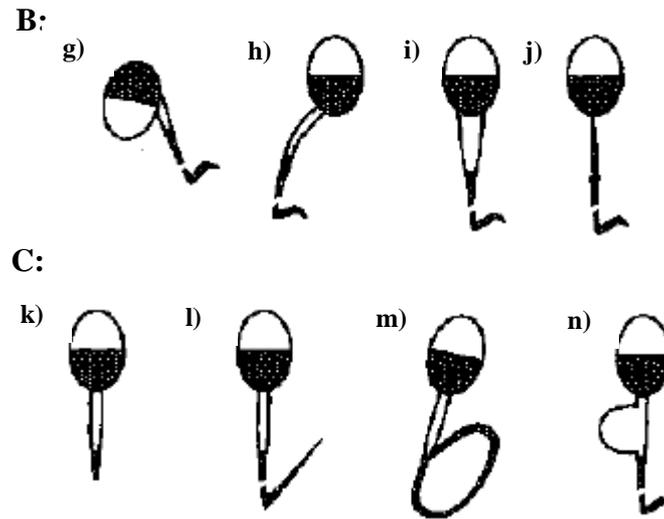


Figura 1.1.2 - Diferentes anomalias morfológicas observadas em espermatozóides humanos

A: a. Cabeça alongada; b. Cabeça piriforme; c. Cabeças arredondadas (Sem acrossoma e pequena); d. Cabeças amorfas; e. Presença de vacúolos; f. Área do acrossoma pequena; **B:** g. Pescoço torcido; h. Inserção Assimétrica; i. Secção intermédia espessa; j. Secção intermédia fina; **C:** k. Cauda curta; l. Cauda dobrada; m. Cauda enrolada; n. Gota Citoplasmática [2]

1.2 - Objetivos do Trabalho

Feita uma contextualização inicial do trabalho, esta secção focar-se-á na enumeração e descrição dos objectivos do projecto e da actual tese.

Este projecto surge de uma colaboração entre o Centro de Neurociências e Biologia Celular (CNC), os Hospitais da Universidade de Coimbra (HUC) e o Instituto de Sistemas e Robótica (ISR) para desenvolver um método de avaliação automática de gametas e embriões, denomina-se “Automated evaluation of gamete and embryo quality for assisted reproduction” e visa auxiliar os atuais métodos de TRA utilizados por forma a aumentar a percentagem de casos de sucesso. A sua finalidade e principal inovação científica é a de procurar criar um sistema de avaliação automático, eficiente e simples para análise gametas e embriões através de imagens microscópicas digitais recolhidas em tempo real. Este sistema pressupõe a formulação de índices de referência para a classificação de múltiplos parâmetros obtidos através do cruzamento dos dados adquiridos com resultados clínicos que dêem indicação das características dos embriões e gametas associados a gravidezes de sucesso.

A corrente dissertação surge como uma abordagem preliminar de carácter exploratório ao problema, focando-se exclusivamente na detecção e classificação de espermatozóides presentes em imagens microscópicas digitais fornecidas pelo CNC, dando assim continuidade ao trabalho desenvolvido por R. Sampaio em [56]. O seu principal propósito é o de propor um método para

contagem e análise das propriedades morfológicas dos espermatozóides presentes nas imagens de amostras disponibilizadas.

Este processo é nuclearmente constituído por 4 etapas distintas conforme apresentado na **Figura 1.2.1**. Estas etapas foram inspiradas pela morfologia típica de um espermatozóide de forma a permitir uma classificação completa dos objectos de análise. A classificação é assim realizada de forma hierárquica, isto é, eliminando das restantes etapas espermatozóides que revelem não reunir as características morfológicamente desejadas de acordo com a avaliação realizada na etapa anterior, por forma a tornar mais célere o processo de classificação:

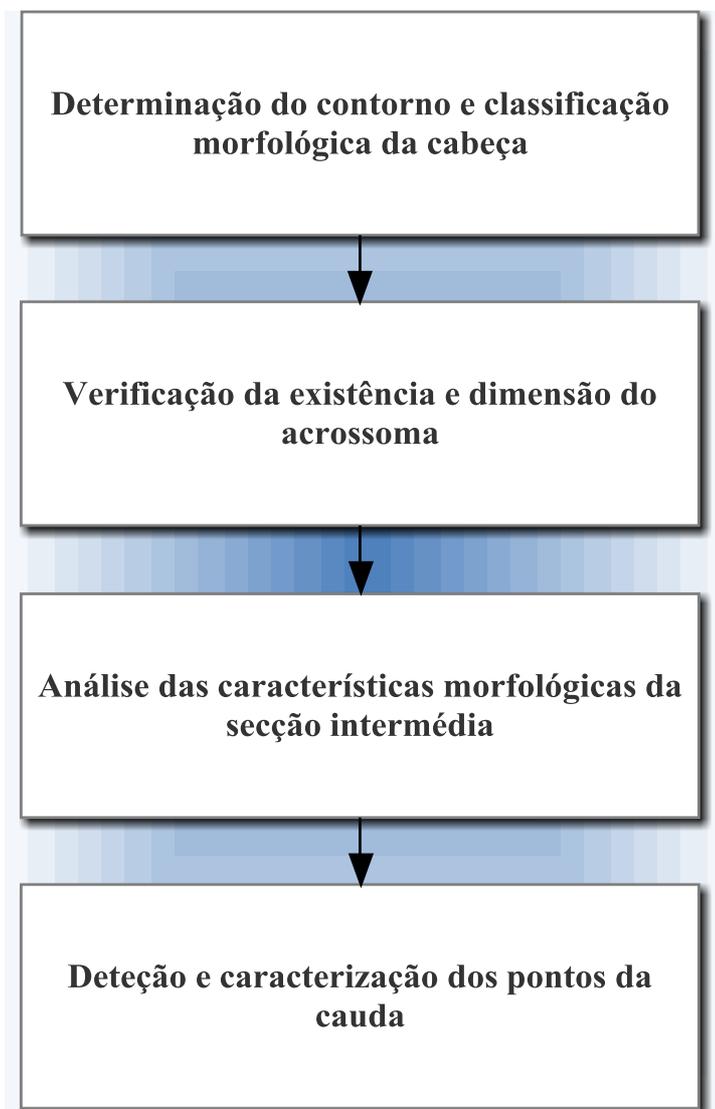


Figura 1.2.1 – Esquema das etapas do sistema de classificação

Esta abordagem é precedida de um passo de pré-processamento cujo propósito é reduzir o ruído indesejado presente na imagem, desprezar objectos de menores dimensões quando comparados com a cabeça dos espermatozóides (manchas de coloração, restos citoplasmáticos,

etc.) e realçar as estruturas na imagem que correspondem à cabeça e secção intermédia de forma a proceder à sua correta segmentação e extração.

Após a etapa de segmentação procede-se a uma transformação da escala das dimensões dos objectos extraídos tomando como referência a área de um espermatozóide de referência de forma a permitir uma comparação mais pronta e eficiente das diferentes características recolhidas. Procura-se ainda centrá-los e orientá-los horizontalmente, aproveitando para isto a forma aproximadamente elipsoidal dos espermatozoides através do uso de um método de enquadramento de elipses. Este passo deverá ser sucedido por uma etapa de separação da cabeça da respectiva secção intermédia, o qual por sua vez dará lugar a uma comparação das características e contorno do espermatozóide em análise com outros de morfologia conhecida.

Se o espermatozóide apresentar uma cabeça de morfologia classificada como normal, procede-se de seguida ao cálculo das dimensões do seu acrossoma. Este procedimento é feito através da segmentação em duas regiões distintas da faixa em torno do eixo maior do espermatozóide onde o acrossoma se encontra geralmente mais realçado. Desta forma, e assumindo que a cabeça tem uma forma regular, é possível de uma forma simples extrapolar se o acrossoma possui as dimensões (40 a 70% da área total da cabeça) desejadas.

Após verificação das dimensões do acrossoma, a atenção é voltada para as características da secção intermédia. Nesta etapa pretende-se determinar genericamente os defeitos associados à secção intermédia representados na **Figura 1.1.2 B e D**. A selecção dos limiares a utilizar para a análise de algumas destas características é feita com base num modelo de aprendizagem extraído de espermatozoides cuja morfologia se encontra previamente estabelecida.

Por último, é feita a detecção e caracterização dos pontos da cauda. Este procedimento parte de um ponto inicial de procura (PIP), tipicamente o ponto distal da secção intermédia segmentada, e procura determinar um conjunto de pontos que constituem a cauda do espermatozóide a analisar. Devido ao facto de ser não existir uma calibração prévia da câmara utilizada na captura das imagens, não é possível caracterizar com muito rigor os pontos da cauda recolhidos, restando assim apenas a possibilidade de proceder a uma análise qualitativa dos resultados.

Capítulo 2 – Estado da Arte

Tendo exposto o problema subjacente ao projeto e delineado genericamente a abordagem utilizada para a sua resolução, dedica-se esta secção ao levantamento e revisão da literatura relevante ao tema, dando particular destaque aos artigos que motivaram os procedimentos descritos ao longo deste documento.

Segundo relatam Y. Wang *et al* em [7], apesar de todo um vasto conjunto de critérios especificados pela Organização de Mundial de Saúde (OMS) para avaliação da morfologia de espermatozóides [3], dos quais alguns foram previamente expostos neste documento, ainda existe uma grande controvérsia científica acerca do tópico devido à falta de validação clínica precisa. É de notar que mesmo antes de se proceder à classificação é necessário recorrer a um conjunto de procedimentos que incluem a amostragem, liquefacção, lavagem e coloração do esperma para permitir a sua avaliação. De entre este conjunto destaca-se a coloração, já que os diferentes métodos afetam substancialmente os resultados obtidos [8]. Com o progressivo desenvolvimento tecnológico, têm surgido métodos de coloração que permitem a utilização mais eficiente de sistemas CASA (Computer-Assisted Sperm Analysis) Estes sistemas surgem como uma forma de colmatar os problemas que emergem das divergências na avaliação como resultado natural da subjectividade dos especialistas responsáveis pela classificação morfológica dos espermatozóides presentes nas amostras e visam providenciar melhores resultados através de uma metodologia mais objectiva e mais precisa.

Em [7] encontram-se descritos genericamente as etapas do típico processo CASA:

- 1) *Aquisição de Imagens*: Este processo centra-se na aquisição de imagens através de uma câmara digital com a lente do microscópio, sendo que tanto a magnificação do microscópio como a resolução da câmara utilizada para a captura são factores chave na qualidade da imagem obtida. Obter imagens de boa qualidade é um detalhe crucial ao bom funcionamento das restantes etapas. Além disto é importante estabelecer o factor de conversão de pixéis para micrómetros [10], cálculo normalmente feito através de um objecto de referência presente na imagem cujas dimensões são conhecidas *a priori*. Adicionalmente é de relevar que sendo os espermatozóides objectos tridimensionais a imagem poderá não reflectir de forma precisa as suas características [11];
- 2) *Melhoria da qualidade da imagem*: O propósito desta etapa é melhorar o contraste entre os objectos a segmentar e o fundo em que estes se encontram. O método a utilizar deve-se-á adequar ao método de captura utilizado. Em [9] e [10], por exemplo, é aplicada transformação por lei de potência a cada um dos componentes do espaço RGB de forma a

realçar as partes constituintes do espermatozóide, enquanto que em [11] R. Rahatabad *et al.* propõem a utilização de operações de abertura e fecho para melhorar as imagens. Esta etapa é fundamental para obter os melhores resultados possíveis nas que a seguem;

- 3) *Deteção de arestas*: Em muitos casos é utilizada a detecção de arestas para destacar contornos na imagem. Este processo implica a conversão para escala de cinzentos e segmentação binária da imagem – em [9], [10] e [12] é descrita a utilização do método de Otsu descrito em [13] para calcular o limiar a utilizar neste processo. Em alguns casos recorre-se a algum pré-processamento para remover objectos espúrios e ruído indesejável – em [14], por exemplo, é utilizado um filtro de mediana 3x3 para reduzir o ruído presente nas imagens.
- 4) *Segmentação*: Muitas vezes surge a necessidade de segmentar os espermatozóides nos seus elementos constituintes, como a cabeça, a secção intermédia e a cauda. Em [9] e [10] é utilizado um método denominado *nth-Fusion* e em [14] um algoritmo de Watershed controlado que derivam dos métodos mais convencionais.
- 5) *Extração de características*: O propósito desta etapa é o de medir os parâmetros do espermatozóide, tais como, por exemplo, a área da cabeça, o seu comprimento e a sua largura, bem como a largura da secção intermédia, o comprimento da cauda e sua curvatura e ainda área do acrossoma [14]. A análise de alguns destes parâmetros permite-nos obter as características morfológicas dos diferentes espermatozóides;
- 6) *Classificação*: Por último, os espermatozóides são classificados de acordo com as suas características morfológicas em diferentes tipos, seguindo os critérios padrão definidos em [3].

Embora seja possível de uma forma geral seguir este tipo de estrutura para o sistema a desenvolver é de notar no entanto que as imagens do CNBC encontram-se agrupadas em múltiplas colorações e apresentam entre si variadas resoluções e mesmo diferentes parâmetros de captura digital, o que limita à partida a efectividade dos algoritmos a implementar. Tendo isto em mente, passa-se à análise individual de soluções já existentes e dos resultados por elas produzidos.

Em [7] discorre-se ainda acerca dos três principais sistemas CASA comerciais utilizados atualmente e a forma como estes lidam com a questão da análise morfológica do esperma humano, processo normalmente conhecido como ASMA (*Automated Sperm Morphology Analyzer*). O sistema SQA (*Sperm Quality Analyzer*) é um sistema simples e pouco dispendioso [15][16] que implementa um método digital para contagem de espermatozóides com mobilidade. Contudo o sistema como é apresentado em [17] não quantifica anomalias específicas. Por outro lado, os sistemas SCA (*Sperm Class Analyzer*) são utilizados para comparação dos diferentes

métodos de coloração, lavagem e preparação das amostras, devido à sua capacidade de analisar 9 parâmetros morfológicos primários da cabeça e secção intermédia tais como a área e o comprimento de ambas, a distância e ângulo de divergência entre os seus eixos maiores e dimensões do acrossoma, bem como como o cálculo de 4 parâmetros derivados da forma da cabeça [18]. Apesar de obter estas medições, o sistema não permite a quantificação de anomalias específicas tais como cabeças alongadas, arredondadas ou amorfas. Os sistemas IOVS (*Integrated Visual Optical Systems*) classificam os espermatozóides em uma de três categorias: normal, subnormal ou anormal, sendo este últimos subclassificados de acordo com um número de anomalias específicas [19].

Existem dois pontos fulcrais a destacar da análise destes sistemas: é fundamental providenciar um sistema que em caso de anomalia seja capaz de classificá-la através das características medidas e que nenhum destes 3 sistemas comerciais, apesar de tornarem mais eficiente e objectiva a detecção e classificação de espermatozóides, proporciona uma solução para detecção e classificação das caudas dos mesmos.

Ao longo dos últimos anos muitos têm sido os métodos sugeridos para detecção, seguimento e classificação de esperma humano. Apesar do seguimento de espermatozóides móveis se encontrar fora do âmbito do corrente projecto de tese, visto que este lida apenas com imagens digitais ao invés de vídeo, [12], [20], [21] e [22] sugerem algumas metodologias para este propósito, sendo interessante destacar o método proposto em [20] devido à sua capacidade de determinar e fazer o seguimento a pontos da cauda através de dois algoritmos distintos, um baseado em MIR e outro no fluxo ótico. O método sugerido em [22] merece também realce devido à proposta de uso de um algoritmo Watershed combinado com filtro de partículas de forma a permitir lidar com situações em que existe colisão de espermatozóides.

Relativamente à segmentação das amostras existe também um conjunto de soluções propostas. Em [23] é comparado um método de segmentação simples por cálculo de entropia máxima em duas dimensões com base no operador de Canny, que utiliza múltiplos limiares para lidar com os típicos problemas de baixo contraste e presença, com outros métodos convencionalmente utilizados. Na **Figura 2.1** são apresentados os resultados da aplicação dos diversos métodos testados no artigo para segmentação de uma imagem com múltiplos espermatozóides:

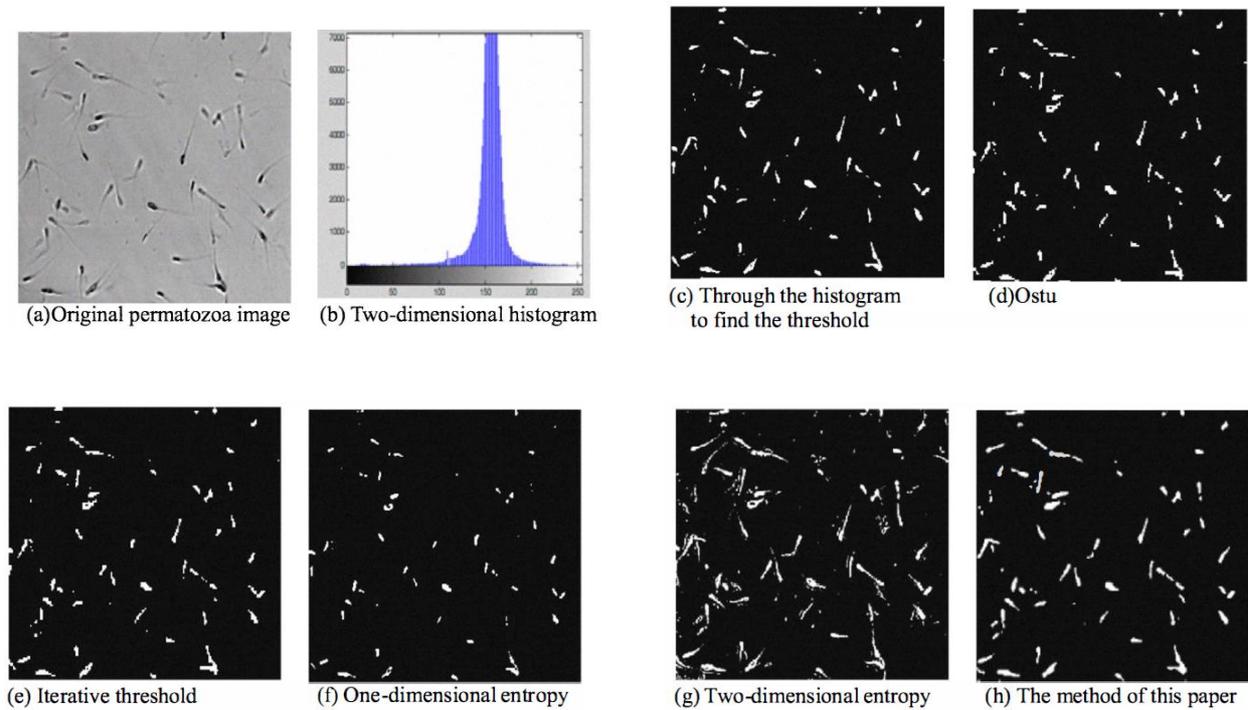


Figura 2.1 - Comparação dos métodos de segmentação testados em [23]

Em [14] é por sua vez sugerida, tal como já mencionado, a utilização de um filtro de mediana 3x3 para reduzir o ruído indesejado, seguido de uma detecção e segmentação feita através do uso do algoritmo de detecção de Sobel e cálculo dos componentes ligados da imagem binária resultante. Este processo é seguido de uma versão modificada de um algoritmo de segmentação baseado em Watershed controlado por marcadores [24] de forma a segmentar a secção intermédia da restante região da cabeça, fazendo posteriormente a recolha de medidas referentes a ambas as partes (em pixéis) que são então utilizadas para proceder à classificação dos diferentes espermatozóides.

Nos artigos [9] e [10], Henry Carrillo e a sua equipa propõem um método que se inicia pela captura e conversão de imagens digitais do espaço RGB para escala de cinzentos e posterior binarização através da utilização do método de Otsu [13]. Este procedimento é seguido por uma remoção dos objectos de menor área presentes na imagem que se esperam ser artefactos presentes no plasma seminal. Os restantes objectos são submetidos a uma análise dos seus histogramas de forma a detetar manchas de dimensões mais próximas dos espermatozóides mas que não apresentam o número de pixéis escuros que tipicamente se associaria ao núcleo destes. Os espermatozóides detetados são então encapsulados na sua caixa envolvente e procede-se à extracção da região correspondente da imagem original no espaço RGB para posterior análise.

Estas regiões da imagem inicial que contêm o núcleo, acrossoma e secção intermédia vêm de seguida o seu contraste aumentado através do uso de uma transformação por lei de potência sobre cada uma das componentes do cubo RGB. É destacada a importância das componentes Vermelha, pois contém a maior parte da informação relativa à cor mais escura que

delineia a cabeça e permite uma maior diferenciação do núcleo das restantes partes da cabeça, e Verde, que dá uniformidade ao fundo permitindo uma mais fácil segmentação da cabeça do contorno relativamente ao fundo (algo que também se verifica nas imagens disponibilizadas pelo CNC). Os resultados apresentados no artigo encontram-se ilustrados na **Figura 2.2**:

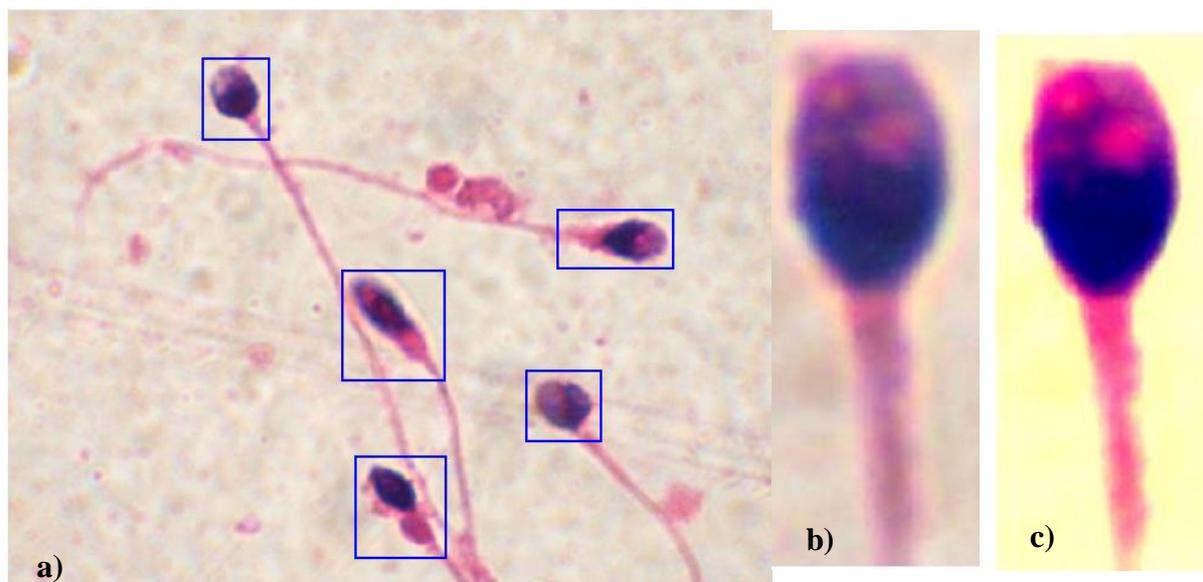


Figura 2.2 - Análise dos resultados dos métodos propostos em [9] e [10]: a) Resultado da fase de deteção e extração; b) Imagem do espermatozóide antes da melhoria de contraste; c) Resultado da melhoria de contraste

Para segmentar as diferentes partes da cabeça, um processo que, devido à proximidade em termos de brilho e textura da secção intermédia e do acrossoma, se revela tendencialmente complexo, é sugerido um método alternativo denominado *nth fusion*. Este método consiste na determinação de regiões isoladas de interesse através do cálculo na binarização por *nth-level* da imagem seguida de uma intersecção com *n* máscaras especiais. As regiões isoladas são depois combinadas para obter os resultados pretendidos, o que requiere *a priori* informação sobre o modelo morfológico em questão [26]. De uma forma resumida, o método selecciona os dois extremos do eixo maior, procede a uma segmentação no domínio RGB do núcleo e estabelece duas máscaras crescentes [27], uma sobre o núcleo já segmentado, e outra sobre a secção da cabeça que se estende do ponto de referência superior da cabeça ao seu centro de massa. Aplica-se então o algoritmo de segmentação previamente descrito, e da intersecção da cabeça segmentada com as áreas de interesse isoladas resultam um conjunto de objetos que definem 3 regiões distintas: acrossoma, núcleo e secção intermédia. Os resultados deste procedimento encontram-se representados na **Figura 2.3**.

Testes feitos sobre 216 imagens de amostras experimentais apresentam uma precisão na detecção de espermatozóides de 95.78% (quociente entre o número de espermatozóides detetados pelo sistema e o número total de gâmetas presentes nas mesmas).

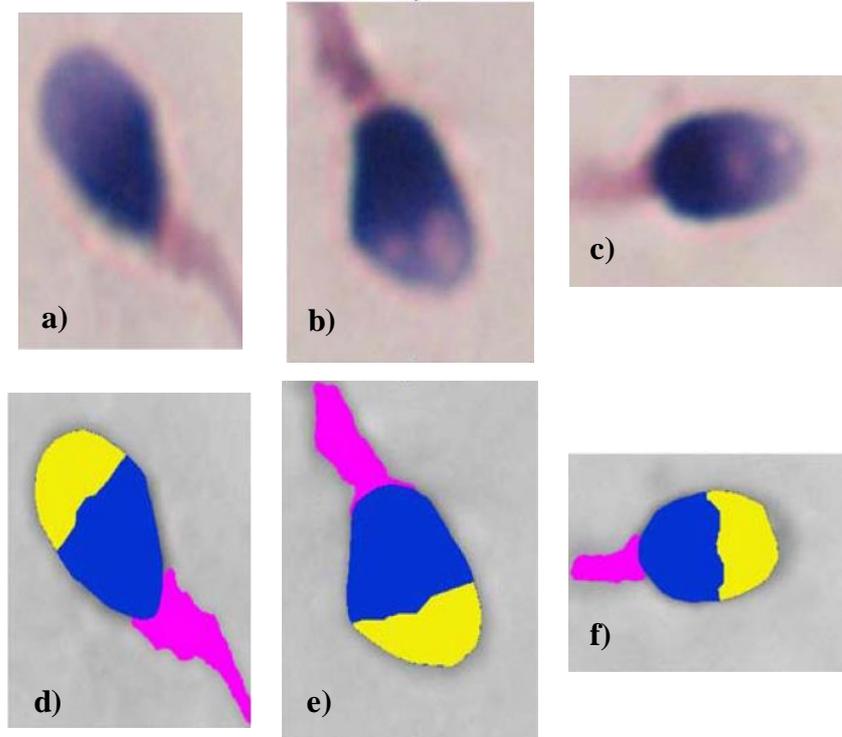


Figura 2.3 - Resultados da etapa de segmentação por nth-fusion presentes em [10]: a), b) e c) Imagens originais; d), e) e f) Resultados da segmentação

Para propósitos de classificação seguiu-se uma abordagem semelhante à sugerida em [14]: foram recolhidas e convertidas por um fator de escala um conjunto de medidas representadas na **Figura 2.4** que permitem determinar a existência de alguma anomalia no espermatozóide a ser analisado, sendo estes parâmetros comparados com os valores estabelecidos em [3]. Se qualquer dos parâmetros estiver fora dos intervalos definidos pela OMS, o espermatozóide é considerado inviável. Terão sido realizados testes sobre 1000 regiões extraídas sobre a supervisão clínica verificando-se uma precisão na classificação estimada em torno dos 85%.

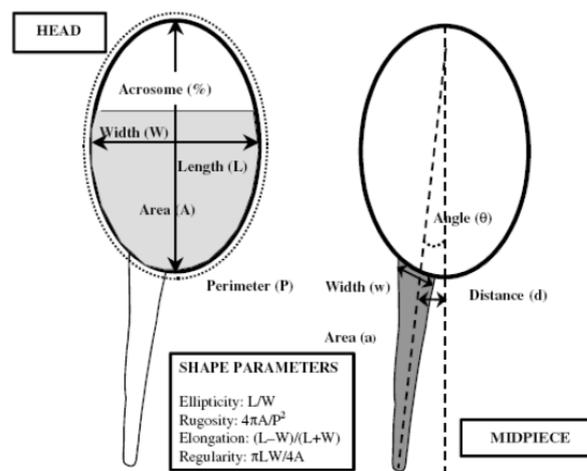


Figura 2.4 - Estrutura e parâmetros de forma de um espermatozóide humana utilizados numa análise objetiva de morfologia de acordo com [18]

Na **Figura 2.4** encontram-se os parâmetros morfológicos tipicamente recolhidos com a finalidade de proceder à análise morfológica da cabeça e secção intermédia de um dado espermatozóide.

Finalmente apresenta-se o método proposto por A. Bijar *et al* em [29] para lidar com o problema. Apesar de não se ocupar do processo de classificação, o sistema descrito incorpora etapas de pré-processamento, detecção e segmentação da cabeça e partes constituintes e, de particular interesse, também apresenta uma solução para detecção da cauda originalmente sugerida em [28].

O processo começa pela aplicação de um método híbrido melhorado sugerido por Raja *et al.* em [30] para remover o ruído da imagem original. De seguida a imagem é binarizada utilizando um limiar simples e são eliminados os objectos de área menor de forma a criar uma máscara contendo toda as cabeças dos espermatozoides presentes na imagem. As regiões segmentadas são encapsuladas em caixas envolventes determinadas através do método de Pinças Rotativas (*Rotating Calipers*) de acordo com os métodos sugeridos em [31] e [32].

Segmentadas as regiões, modela-se o brilho dos pixéis presentes nas caixas envolventes utilizando um modelo de mistura gaussiana que consiste em três *kernels* diferentes: fundo, núcleo, e uma classe que engloba o acrossoma e a secção intermédia. Esta etapa utiliza um algoritmo EM (Expectation Maximization) baseado no cálculo de entropia para estimar de forma automática uma primeira aproximação das três classes mencionadas sem necessidade de estipular parâmetros iniciais. Segue-se o cálculo das probabilidades *a priori* e a função de distribuição de probabilidade condicional de cada classe, através de um modelo baseado em MRF (*Markov Random Fields*) e maximização de expectativas. Por último, os valores calculados são utilizados para classificar as amostras através de critérios Bayesianos [33] [34].

Feita esta classificação, é seleccionado o ponto distal da secção intermédia para servir como ponto inicial para a pesquisa dos pontos da cauda. O método proposto recorre ao SSIM (*Structural Similarity Index*) [35] e entropia de Rényi [36] num esquema iterativo de forma a determinar os pontos que pertencem à cauda de uma dado espermatozóide [29]. Os resultados obtidos podem ser observados na **Figura 2.5**, onde podemos visualizar não só a segmentação das regiões correspondentes à cabeça e respectiva secção intermédia dos vários espermatozoides, mas também os pontos da cauda detetados (é de notar contudo que ao longo do artigo apresentam-se apenas resultados do algoritmo de detecção de caudas para espermatozoides isolados).

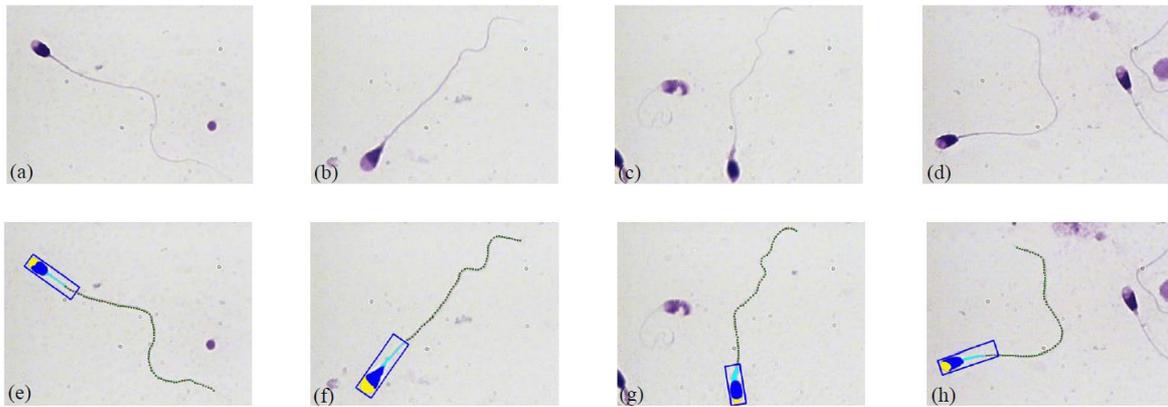


Figura 2.5 - Resultados do método de detecção da cauda proposto na literatura: a) - d) Imagens Originais; e) - h) Segmentação das partes constituintes da cabeça e pontos estimados da cauda [28][29]

Revista a principal literatura acerca da tónica em questão, chegou o momento de ponderar em que medida os contributos científicos feitos através desta poderão servir como base e inspiração do trabalho a realizar.

À partida, e visto tratar-se de um dos artigos recentes mais promissores acerca da temática, optou-se seguir um esquema semelhante ao de [28] [29] de A. Bijar *et al* procedendo às alterações necessárias para adequar o sistema proposto às necessidades do projecto. Esta decisão deve-se ao facto do artigo não só apresentar uma sugestão de segmentação com resultados geralmente satisfatórios para imagens que utilizam métodos de coloração semelhantes aos disponibilizados para este trabalho, mas também porque se trata do único método que inclui um algoritmo para detecção dos pontos da cauda.

Devido ao facto do método proposto não apresentar nenhuma sugestão concreta relativamente à caracterização e avaliação dos parâmetros de forma dos espermatozoides, optou-se por uma metodologia semelhante à proposta em [9] e [10] por Henry Carrillo *et al*. Surge contudo o problema, como foi previamente referido, da impossibilidade de determinar uma fator de escala que permita converter as medidas de píxéis para micrómetros para poder comparar os dados recolhidos directamente com os intervalos especificados pela OMS em [3]. Assim optou-se por redimensionar todas as cabeças a uma escala resultante do quociente da sua área face à área de um espermatozoide de referência. Assim, após segmentação seria possível utilizar uma metodologia assente sobre um modelo de aprendizagem com base na comparação de parâmetros de forma com espermatozoides de morfologia previamente identificada.

Para colmatar situações em que não é possível distinguir alguns tipos de cabeças amorfas através da mera análise dos parâmetros recolhidos é também utilizado um método que parte do cálculo da assinatura de contornos [37] da cabeça para proceder à sua classificação através de um sistema SVM [38], permitindo assim fazer uma triagem mais eficiente.

Ao longo deste processo, procede-se também a uma transformação geométrica de forma a alinhar horizontalmente o eixo maior da cabeça. Este detalhe permite, entre outras coisas, a simplificação a segmentação/extracção da secção intermédia. Ao invés de recorrer a um procedimento computacionalmente mais exigente, considerou-se uma abordagem simplificada que consiste na extração de todos os pixéis da região encontrados para além do ponto de intersecção do eixo intermédio com a região do núcleo segmentado através da utilização do método Otsu [13]. Foi também ponderada a possibilidade de acelerar a caracterização do acrossoma extrapolando a sua proporção face ao núcleo ao longo do eixo maior da cabeça após remoção da secção intermédia, visto que os espermatozóides que possuem uma cabeça de morfologia normal apresentam um acrossoma tipicamente distribuído de forma uniforme.

Capítulo 3 – Implementação do Sistema

Este capítulo é dedicado à descrição dos métodos implementados para atingir os objectivos previamente delineados. Procura-se assim fazer uma breve fundamentação das escolhas feitas e uma enumeração dos passos tomados até à sua concretização.

3.1 – Etapa 1 – Tratamento da Imagem

O objectivo desta primeira etapa é descrever o método utilizado para o pré-processamento de cada imagem de forma a melhorar o seu contraste, diminuir o ruído indesejável e realçar os objectos que pretendemos extrair posteriormente numa etapa de segmentação. Na **Figura 3.1.1** são incluídas algumas das 89 imagens de amostras (contendo um número variável de espermatozóides) fornecidas pelo CNC:

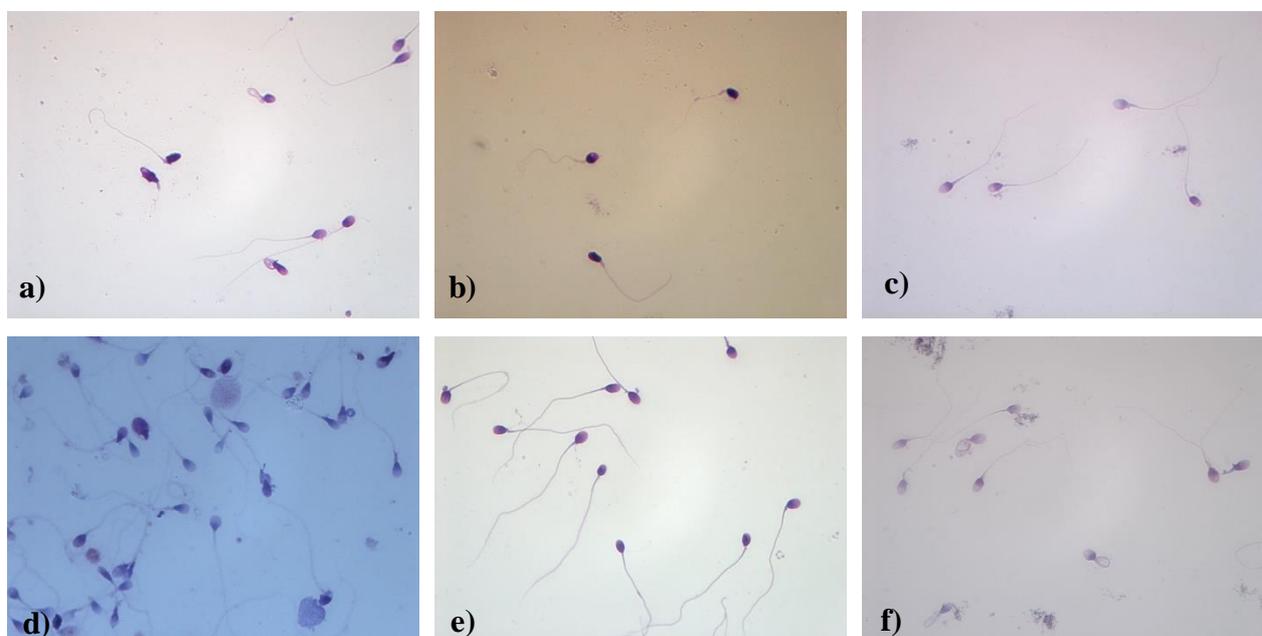


Figura 3.1.1 - Exemplos de Imagens fornecidas demonstrando as diferentes colorações utilizadas: a) Diff-Quik; b) Hemacolor; c-f) Giemsa

Como é possível observar, estas apresentam uma grande variação de brilho e contraste entre si. Isto deve-se ao facto de estas se encontrarem organizadas por 3 métodos diferentes de coloração (Diff-Quick, Hemacolor e Giemsa). É de notar também que mesmo as imagens que utilizam o mesmo método de coloração apresentam variações, por vezes não insignificantes, das suas diferentes características, facto que de acordo com [8] poderá afetar de forma significativa os resultados dos algoritmos que venham a ser aplicados. Contudo, de uma forma geral estas apresentam genericamente a mesma configuração: espermatozóides mais escuros sobrepostos

sobre um fundo mais claro e geralmente uniforme, pelo que é razoável assumir à partida que os mesmos pressupostos assumidos em [28] e [29] se aplicariam a estas imagens.

Tendo isto em consideração, procedeu-se à implementação do método híbrido melhorado descrito em [30] por Rajan *et al.*. Este método consiste em duas etapas aplicadas sobre a componente vermelha do espaço RGB da imagem original: um primeiro estágio consiste na aplicação de uma PDE (Equação Diferencial Parcial) de 4ª Ordem, enquanto o segundo resulta da aplicação de um filtro de mediana relaxado à imagem resultante do estágio anterior. É utilizado um filtro de mediana relaxado pois este consegue preservar mais detalhes que um filtro de mediana típico. Este método mostra-se capaz de preservar detalhes finos, cantos, linhas com pouca espessura e estruturas curvas de uma forma geral. Para este propósito é utilizado o método de fluxo de gradiente de curvatura-L2 de You *et al.* [39]:

$$\frac{\partial u}{\partial t} = -\nabla^2 [c(|\nabla^2 u|) \nabla^2 u] \quad (1)$$

onde $\nabla^2 u$ representa o laplaciano da imagem u . Já que o laplaciano da imagem num dado pixel é zero se a imagem for planar na sua vizinhança, a PDE tenta remover o ruído e preservar arestas, aproximando a imagem observada por uma imagem planar por partes. O coeficiente de difusão c deverá ser tal que (1) resulte numa maior difusão em áreas homogéneas e menor em regiões de transição de intensidade, de forma a reduzir pequenas variações no brilho como ruído e texturas indesejadas e a preservar as arestas. Outro objectivo na selecção do coeficiente c é que ocorra uma difusão inversa em torno das regiões de transição para que as arestas sejam realçadas, e o oposto nas áreas mais homogéneas para que haja remoção do ruído. É assim considerada a função de difusidade de *Perona-Malik* [40] na implementação do método:

$$c(s) = \frac{1}{1 + \left(\frac{s}{k}\right)^2} \quad (2)$$

Segundo [30] a forma discreta da PDE descrita em (1) é a seguinte:

$$u_{i,j}^{n+1} = u_{i,j}^n - \Delta t \nabla^2 g_{i,j}^n \quad (3)$$

onde

$$\nabla^2 g_{i,j}^n = \frac{g_{i+1,j}^n + g_{i-1,j}^n + g_{i,j+1}^n + g_{i,j-1}^n - 4g_{i,j}^n}{h^2} \quad (4)$$

$$g_{i,j}^n = g(\nabla^2 u_{i,j}^n)$$

$$\nabla^2 u_{i,j}^n = \frac{u_{i+1,j}^n + u_{i-1,j}^n + u_{i,j+1}^n + u_{i,j-1}^n - 4u_{i,j}^n}{h^2} \quad (5)$$

Δt é o tamanho do passo temporal e h é o tamanho da grelha espacial (que por padrão é 1). O filtro de mediana relaxado [41] [42] é usado em combinação com a equação 1 para remover picos de ruído, sendo a definição do método proposto:

$$u_{i,j}^{n+1} = RM_{\alpha\omega}(u_{i,j}^n - \Delta t \nabla^2 g_{i,j}^n) \quad (6)$$

em que RM é o filtro de mediana relaxado com limite inferior α e limite superior ω . Sendo $Y(i,j)$ o resultado do filtro de mediana relaxado, podemos escrever $Y(i,j)$ como

$$Y(i,j) = RM_{\alpha\omega}\{W(i,j)\} = \begin{cases} X(i,j) & X(i,j) \in [[W(i,j)]_{\alpha}[W(i,j)]_{\omega}] \\ [W(i,j)]_m, & \text{caso contrário} \end{cases} \quad (7)$$

onde $[W(i,j)]_m$ é o valor da mediana dos pixels dentro da janela centrada no pixel de coordenadas (i,j) e de dimensão m , em que $\alpha \leq m \leq \omega \leq 2N + 1$, ou seja, os dois limites estabelecem uma sublista dentro de:

$$W(i,j) = \{X(i+r, j+r), r \in W\} \quad (8)$$

que define os valores que assumimos serem bons o suficiente para não ser filtrados. De uma forma resumida, o que isto significa é que se o valor do pixel a ser analisado se encontrar nesse sublista o seu valor de brilho permanecerá inalterado, caso contrário será substituído pelo valor do filtro de mediana padrão. Os valores seleccionados para os limites inferior e superior para esta implementação foram 3 e 5, de acordo com o sugerido em [30].

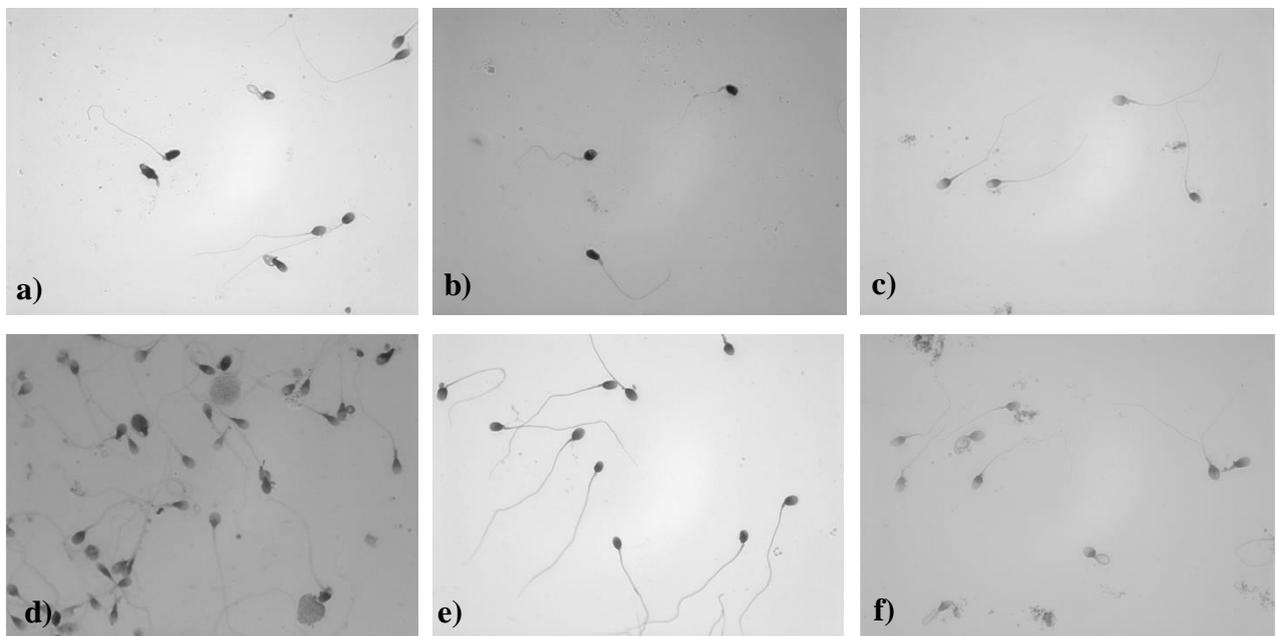


Figura 3.1.2 - Resultados da etapa de tratamento das imagens através do método sugerido por Rajan em [30]

Os resultados apresentados na **Figura 3.1.2** foram obtidos com a execução do algoritmo proposto durante 3 iterações e com $\Delta t = 0.9$ e $h = 1$, valores que revelaram experimentalmente permitir obter um bom equilíbrio entre o tempo computacional utilizado e as características das imagens resultantes. Quando comparadas com as imagens originais representadas na **Figura 3.1.1**, verifica-se geralmente um realce notório dos espermatozoides face ao fundo que, graças à remoção de ruído, se torna também mais homogéneo. É de notar contudo que a qualidade da imagem original não deixa de ser determinante para a qualidade dos resultados. Basta para isto

verificar que imagens com pouco contraste e presença de muitos artefactos no plasma seminal, como é o caso na **Figura 3.1.2 d)**, resultam obviamente em imagens de mais difícil segmentação apesar da melhoria resultante do procedimento.

3.2 – Etapa 2 – Segmentação

A próxima etapa centra-se na criação de uma máscara primária que contenha os diferentes elementos constituintes da cabeça do espermatozóide. Para isso é executado um procedimento de binarização simples, semelhante ao sugerido em [29], com base num limiar $thr = \mu - \sigma$, em que μ e σ representam respetivamente a média e o desvio padrão da imagem tratada $f(x,y)$. Empiricamente verificou-se que os resultados obtidos eram geralmente melhores quando multiplicando σ por 2 para a vasta maioria das imagens, reduzindo o número de artefactos ainda recolhidos, pelo que foi considerada esta adição. Assim, a máscara primária $g(x,y)$ é definida como:

$$g(x,y) = \begin{cases} 1 & \text{se } f(x,y) < thr \\ 0 & \text{se } f(x,y) > thr \end{cases} \quad (9)$$

Este procedimento é seguido de operações morfológicas com o objectivo de remover os objectos com área menor a 1/3 do maior objecto da máscara primária, retirando assim todo um conjunto de artefactos de menor dimensão.

Para terminar esta etapa procurou-se definir as caixas envolventes que contêm as regiões isoladas. Por questões de simplicidade de manipulação dos objectos após extracção, optou-se por determinar apenas as caixas envolventes de menor área cujos lados são paralelos aos eixos do sistema de coordenadas da imagem, sendo para isso necessário determinar o envelope convexo de cada região [43] e os seus limites máximos em função de x e y.

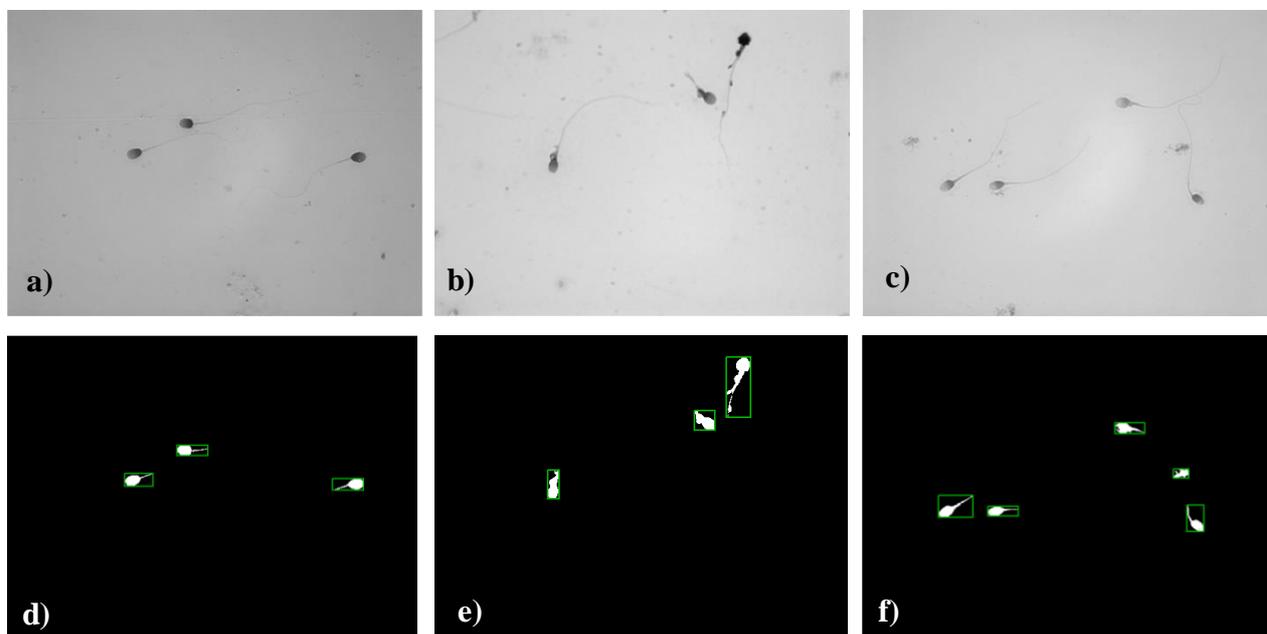


Figura 3.2.1 – Resultados da Etapa de Segmentação: a) - c) Imagens após tratamento; d) - f) Imagens Segmentadas

Pode-se verificar através da **Figura 3.2.1** que o método permite extrair as regiões correspondentes à cabeça e secção intermédia dos espermatozóides nas imagens tratadas.

Nesta etapa procede-se ainda à verificação do histograma da região delimitada pelas caixas envolventes de forma a determinar a existência de manchas de dimensões próximas às das cabeças segmentadas, através da análise da percentagem de pixels escuros existentes que corresponderiam a região do núcleo, para que estas possam ser desprezadas nas etapas seguintes.

3.3 – Etapa 3 – Classificação Morfológica da Cabeça

3.3.1 – Obtenção da Região da Cabeça

O primeiro passo nesta etapa referente ao processamento da cabeça e sua análise morfológica é obter uma imagem com uma resolução 500*500 (dimensão escolhida experimentalmente de forma a permitir a inclusão das regiões segmentadas) com o espermatozóide detetado centrado, orientado horizontalmente.

Para proceder a esta manipulação recorreu-se ao algoritmo sugerido em [44] por A. Fitzgibbon *et al.*, cuja implementação em MATLAB se encontra disponível em [45], que determina através de um algoritmo não iterativo a elipse que melhor se enquadra a uma dada configuração de pontos segundo os desvios mínimos quadrados. Esta implementação do algoritmo devolve os valores correspondentes às dimensões do eixo maior e menor da elipse, as coordenadas do seu centro e o ângulo que esta apresenta face ao eixo das coordenadas x da imagem. Utilizando estes parâmetros é possível assegurar que o objecto é separado e realinhado horizontalmente como se verifica na **Figura 3.3.1.1**. Esta figura apresenta a cabeça e secção intermédia de 3 espermatozóides diferentes extraídos de imagens distintas após a etapa de segmentação.

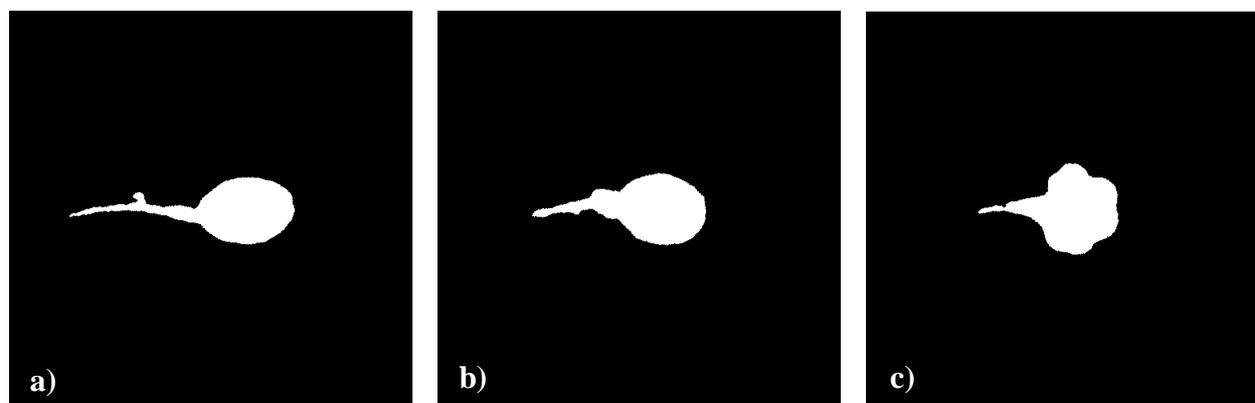


Figura 3.3.1.1 - Regiões Segmentadas contendo a cabeça e secção intermédia

De seguida, procede-se à separação da secção intermédia da cabeça para possibilitar uma recolha eficiente do conjunto de parâmetros utilizados para caracterizar a cabeça. Como já foi

referido anteriormente, optou-se por simplificar este processo através do uso de um método de binarização com base num limiar calculado pelo método de Otsu. Aproveitando o facto das cabeças se encontrarem orientadas na horizontal, é feita uma primeira segmentação dos pixéis da região de forma a delinear a região mais escura que corresponderia ao núcleo. De seguida, procura-se determinar o ponto contido no eixo maior da região que se encontra na fronteira da região que corresponderia ao colo espermático onde se inseriria a secção intermédia. Para este efeito determina-se canto da caixa envolvente que apresenta a menor distância euclidiana ao centro de massa da região do núcleo, procurando de seguida obter as coordenadas do ponto de fronteira que se encontra sobre o eixo maior mais distante do canto determinado previamente. Determinadas as coordenadas deste ponto é possível proceder à extracção da secção intermédia. Esta aproximação tem as suas limitações no facto de não permitir fazer uma completa diferenciação dos pontos da secção intermédia que possam pertencer ao colo espermático (região de inserção da secção intermédia na cabeça), mas revela-se, na maioria dos casos, uma abordagem que permite solucionar expeditamente a separação entre os dois elementos sem comprometer a sua devida classificação. Na **Figura 3.3.1.2** podemos verificar que foi possível isolar a região aproximada da cabeça dos espermatozóides presentes nas imagens da **Figura 3.3.1.1**.

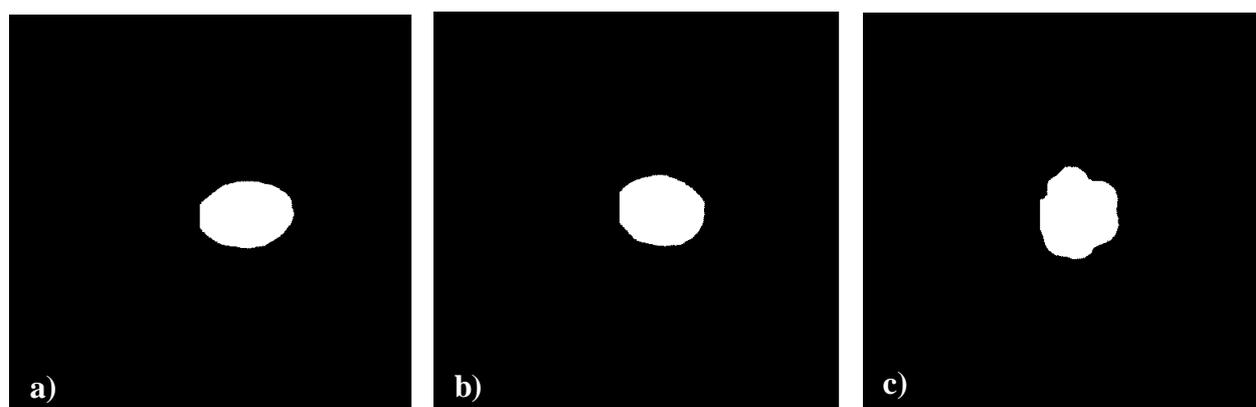


Figura 3.3.1.2 - Regiões Segmentadas após extracção da secção intermédia

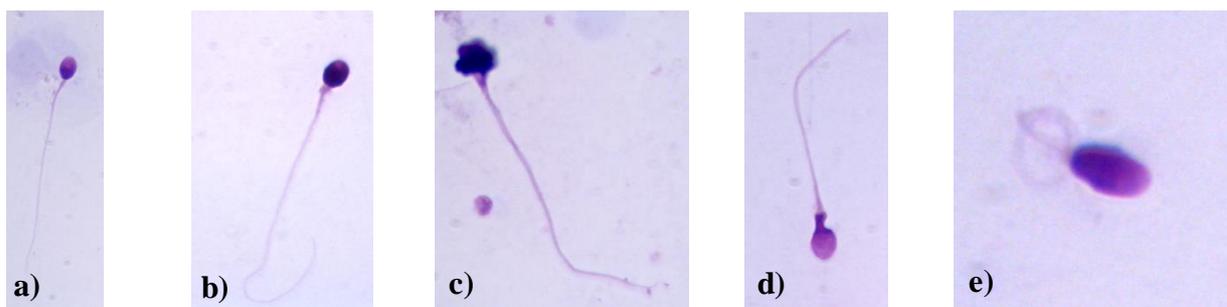
3.3.2 – Análise dos Parâmetros Morfológicos

O primeiro passo desta classificação, como já previamente descrito, centra-se na recolha de um conjunto de parâmetros descritos em [18] e por C. Soler *et al* e surgem representados na **Figura 2.4** com o propósito de fazer uma avaliação das características da cabeça, e que se encontram enumerados na seguinte lista:

- *Comprimento da cabeça*: Este parâmetro é definido como o segmento de reta que contém o centro de massa da cabeça e tem declive igual ao da reta definida pelo pixel de maior valor numa transformada de distância euclidiana;

- *Largura da cabeça:* Calculada como o comprimento mas relativamente ao eixo menor perpendicular ao eixo maior;
- *Perímetro:* Número de pixéis contidos na fronteira da cabeça;
- *Área da Cabeça:* É o número de pixéis contidos na região da cabeça;
- *Ângulo de Inserção da Secção Intermédia:* É o ângulo existente entre o eixo maior da cabeça e o eixo maior da região segmentada (cabeça juntamente com a secção intermédia).
- *Distância da Secção Intermédia:* Separação existente entre a extremidade do eixo maior da cabeça e da região segmentada no início da secção intermédia.
- *Largura da secção intermédia:* Magnitude do eixo menor da secção intermédia;
- *Achatamento, Rugosidade, Elongação e Regularidade:* Parâmetros secundários calculados de acordo com as fórmulas apresentadas na **Figura 2.4**.

Como não houve uma calibração prévia da câmara utilizada na captura e não é possível determinar um fator de escala para efetuar a conversão de pixéis para μm , esta abordagem fica dependente de um modelo de aprendizagem que apresenta um conjunto de limitações. Estas limitações devem-se maioritariamente ao facto de apenas se encontrarem disponíveis 10 imagens com espermatozoides isolados com classificação morfológica conhecida *a priori* o que obviamente restringe severamente a precisão da classificação. Ainda assim esta abordagem, assumindo um muito maior número de amostras, teoricamente permitiria fazer uma primeira distinção morfológica eficiente das cabeças, já que com os parâmetros utilizados é possível detetar facilmente cabeças alongadas, particularmente arredondadas, piriformes ou de dimensões anómalas. Para o caso específico, e visto que as únicas cabeças classificadas previamente como anómalas são alongadas ou amorfas, este método só permite fazer uma distinção das cabeças com um eixo maior que se apresente claramente superior às restantes, sendo estas classificadas como alongadas.



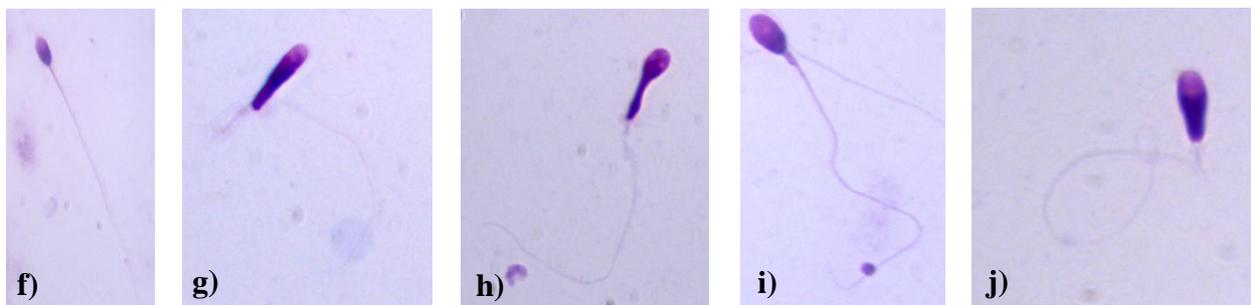


Figura 3.3.2.1 - Imagens de referência (com morfologia previamente identificada pelo CNC): a) Normal; b) Gota Citoplasmática; c) Cabeça Amorfa 3; d) Cabeça Amorfa 1; e) Cauda Enrolada; f) Inserção Assimétrica; g) Cabeça Alongada; h) Cabeça Amorfa 2; i) Secção Intermédia Espessa; j) Cauda Dobrada

Na **Figura 3.3.2.1** representada acima é possível verificar as 10 imagens com espermatozoides isolados cuja morfologia foi previamente identificado por técnicos do CNC. É de relevar que apesar de existirem 10 imagens, apenas 5 têm a morfologia da sua cabeça identificada. As restantes 5 têm designações apenas referentes à secção referente ou cauda.

3.3.3 – *Sistema de Análise por SVM*

Surge porém a questão do que fazer relativamente aos casos em que não é possível discriminar formas anómalas somente através destes parâmetros. Isto é particularmente prevalente quando lidamos com cabeças amorfas que englobam todo um conjunto de deformidades que vão desde da presença de pequenas reentrâncias em cabeças de outra forma normais a formas completamente irregulares. Veja-se, por exemplo, a cabeça do espermatozóide c) da **Figura 3.3.2.1**: uma rápida inspeção visual revela que esta cabeça é inquestionavelmente morfológicamente anómala, tratando-se de facto de uma cabeça amorfa. No entanto o facto de esta apresentar um contorno bastante irregular poderá não ser completamente óbvio se considerarmos apenas características tais como o seu comprimento, largura, área e perímetro. Para distinguir este tipo de cabeças amorfas de entre as classes mais abrangentes – redondas, piriformes, com dimensões anómalas, alongadas ou mesmo, neste caso particular, normais – cuja distinção seria feita na primeira fase do procedimento, foi decidido recorrer a um método alternativo. Decidiu-se proceder à obtenção da assinatura de contornos como base de uma segunda etapa de procedimento, visto tratar-se de uma técnica tipicamente utilizada para resolução de problemas semelhantes, geralmente com bons resultados.

Com base no exposto por P. Peixoto, J. Gonçalves, e H. Araújo em [46] e por F. Schmidberger e F. Stolzenburg em [47], foi desenvolvido um método para a extracção da assinatura de contornos das cabeças. Embora neste caso a base do projeto não sejam vídeos mas sim apenas imagens digitais, não sendo possível definir imagens de assinatura como em [46], procedeu-se a um método semelhante ao proposto para obtenção de dois vectores ρ e θ , que

correspondem respectivamente à distância entre centro de massa e cada ponto do contorno da cabeça e ao ângulo existente entre esse mesmo ponto e o eixo de coordenadas x da imagem. Estes vetores são depois amostrados de forma a ficarem com um comprimento $N = 100$, sendo ρ também normalizado de forma a permitir proceder à eventual comparação das diferentes assinaturas.

O passo seguinte prende-se com a necessidade de encontrar um método eficiente para proceder à comparação dos vetores de dimensão 100×2 obtidos. Apesar do número bastante limitado de amostras classificadas existente para uso como referência, procurou-se desenvolver um sistema com base em Máquinas de Vetores de Suporte (SVM, do inglês *Support Vector Machines*) [48].

As técnicas com base em SVM fazem parte de um conjunto de métodos de *Machine Learning* que tem vindo a receber cada vez mais atenção por parte da comunidade científica, sendo que o seu uso em variados domínios, tais como a Bioinformática e o processamento digital de imagens, especialmente no reconhecimento de padrões, é bem conhecido. O método tem por base a teoria de aprendizagem estatística desenvolvida por Vapnik, teoria esta que estabelece uma série de princípios relevantes para a obtenção de bons classificadores, caracterizados pela sua capacidade de prever correctamente a classe de novos dados do mesmo domínio em que ocorreu a aprendizagem.

A técnica utilizada tem por base critérios indutivos definidos por um treino prévio do classificador. A ideia é, neste caso específico, providenciar vetores de assinatura devidamente identificados como pertencendo a cabeças morfologicamente normais e amorfas de forma a obter um classificador que consiga binariamente determinar a qual das duas classes pertencem os pontos correspondentes a uma nova cabeça cuja classe é ainda desconhecida. O propósito do método que procuramos desenvolver é providenciar uma descrição compacta e robusta (é importante que esta seja capaz de lidar com dados imperfeitos, como a presença de *outliers*) dos dados fornecidos, ajustando-se o melhor possível aos dados disponíveis de forma a maximizar o número de predições corretas. A **Figura 3.3.3.1** contém a representação de obtenção de um classificador através destas técnicas supervisionadas, onde são utilizados dados de treino com múltiplos atributos distribuídos entre classes distintas. Segue-se uma muito breve explicação da teoria que se encontra na base do método utilizado.

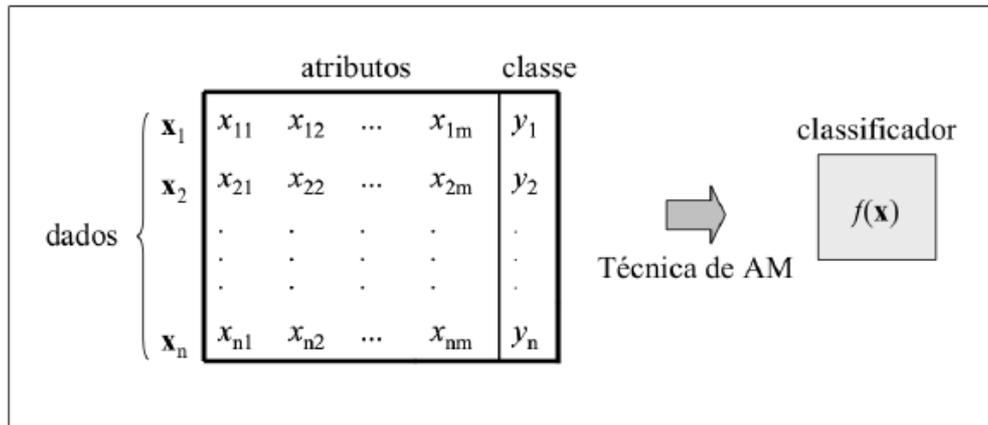


Figura 3.3.3.1 - Representação esquemática do processo de obtenção de um classificador através de uma técnica de Machine Learning supervisionada genérica [48]

Tipicamente um sistema SVM linear simples classifica dados procurando encontrar o melhor hiperplano que consiga separar os pontos de um classe dos pontos da outra, ou seja, o hiperplano com maior margem entre as duas classes. A margem representa a largura máxima da secção paralela ao hiperplano que não inclui quaisquer pontos interiores. Os vetores de suporte são os pontos que se encontram mais próximos do hiperplano de separação, encontrando-se sobre os limites da vizinhança da secção definida pela margem, como é observável na **Figura 3.3.5**.

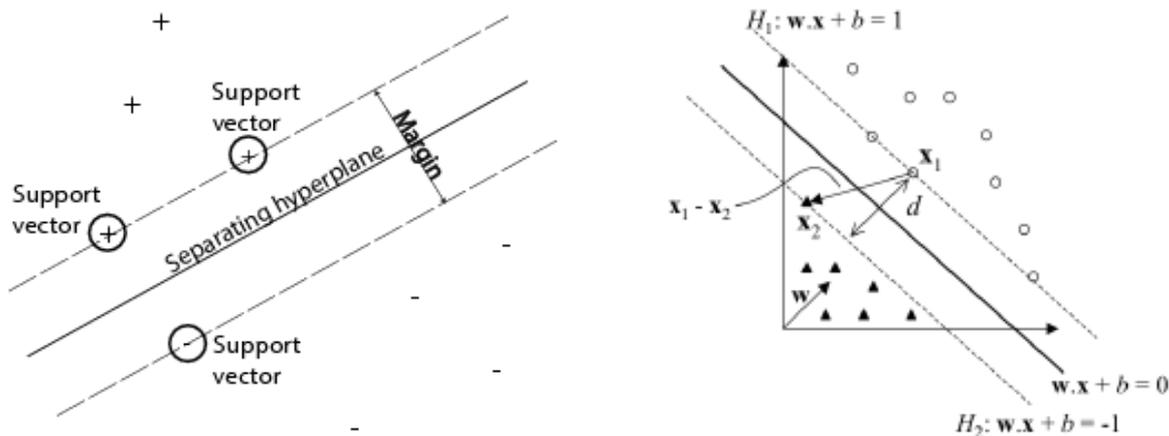


Figura 3.3.3.2 - Esquema de funcionamento de um sistema baseado em SVM linear (cálculo da distância de separação dos dois hiperplanos)[38] [48]

Contudo para o caso particular dos dados que pretendemos classificar a utilização de SVMs lineares não é suficiente, visto que não estamos a lidar com dados que sejam linearmente separáveis. Não sendo possível separar os dados de treino satisfatoriamente por um hiperplano, passa-se a considerar uma fronteira curva. A forma como as técnicas com base em SVM lidam com este tipo de problemas não lineares é através do mapeamento do conjunto de treino do seu espaço original, denominado espaço de entradas, para um espaço de maiores dimensões denominado espaço de características (*feature space*). Procura-se assim obter um mapeamento

$\Phi : X \rightarrow \mathfrak{S}$, sendo X o espaço de entradas e \mathfrak{S} o espaço de características, tal que Φ possibilite que a separação do conjunto de pontos de treino mapeado em \mathfrak{S} através de uma SVM linear. Isto baseia-se no teorema de Cover [50], segundo o qual um conjunto de dados não linear em X pode ser transformado para um espaço de características \mathfrak{S} no qual existe uma alta probabilidade destes serem linearmente separáveis. Isto contudo implica que a transformação seja não linear e que a dimensão do espaço de características seja suficientemente elevada. O racional por detrás deste método encontra-se representado na **Figura 3.3.6**, onde verificamos que apesar de no espaço de entradas apresentado em a) e b) ser impossível separar as classes de treino através de uma fronteira linear, ao fazermos o mapeamento para um espaço de características de maior ordem em c), passa a ser possível fazer uma separação linear.

No entanto como \mathfrak{S} pode ter dimensão muito alta ou mesmo infinita, a computação de Φ poderá ser inviável. Para simplificar o processo recorre-se ao uso de funções denominadas Kernels. Um Kernel K não é mais do que uma função que recebe dois pontos x_i e x_j do espaço de entradas e computa o produto escalar no espaço de características. Estes são muitas vezes empregues sem conhecimento do mapeamento Φ , que é gerado implicitamente, facilitando assim os cálculos e permitindo representar espaços abstractos..

Alguns dos Kernels mais utilizados são os Polinomiais, os Gaussianos ou RBF (*Radial-Basis Function*) e os Sigmoidais. Cada um deles apresenta um conjunto de parâmetros a serem determinados pelo utilizador de forma a obter resultados óptimos.

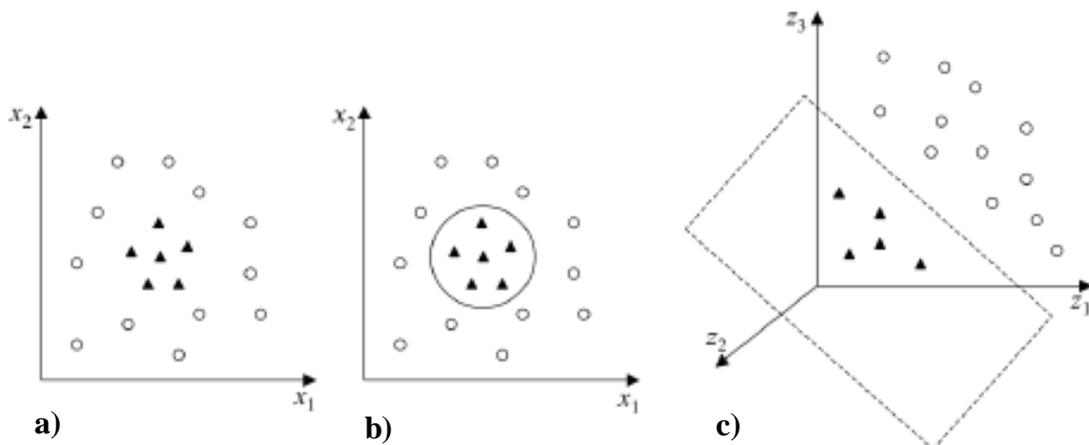


Figura 3.3.3.3 - SVM Não Linear: a) Conjunto de Dados não linear;b) Fronteira não linear no Espaço de Entradas; c) Fronteira linear no Espaço de Características [48]

Para o problema em questão foram obtidos experimentalmente melhores resultados através do uso de um Kernel RBF com $\sigma = 0.2$. O treino do classificador SVM foi feito a partir de uma classe constituída por um conjunto de pontos correspondentes a assinaturas de cabeças assumidas como tendo à partida uma forma normal e uma outra classe correspondente aos pontos de assinaturas de diversas cabeças amorfas, encontrando-se estes dados em ficheiros do tipo *.xls*

guardados com os nomes N.xls e A.xls, respectivamente, que são carregados aquando da execução do módulo. Depois desta etapa de treino, o classificador é então confrontado com os pontos de teste da nova assinatura a analisar, classificando-os como normais ou anómalos. No caso em que se verifique que os pontos classificados como anómalos excede os que são classificados como normais, a cabeça do espermatozóide é classificada como amorfa e este é desprezado nas restantes etapas do sistema. Apesar do número bastante restrito de amostras disponíveis, os resultados gerais apresentaram-se promissores, permitindo distinguir geralmente as cabeças amorfas que não fossem detectadas durante a primeira comparação com as distribuições de parâmetros morfológicos das cabeças previamente classificadas.

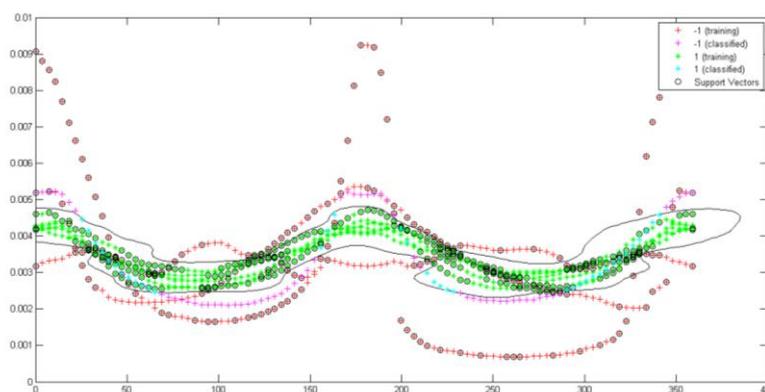


Figura 3.3.3.4 - Classificação com base SVM de um espermatozóide de cabeça com morfologia amorfa: A verde encontram-se os pontos do conjunto de treino classificados como normais; a vermelho os pontos do conjunto de treino classificados como anómalos; a azul os pontos do conjunto de teste normais; e, por último, a magenta os pontos de teste anómalos

Observando os resultados na **Figura 3.3.7**, referentes aos pontos da assinatura de contornos da cabeça extraída da imagem presente na **Figura 3.3.3 j)**, é possível verificar que apesar de esta possuir uma forma com algumas semelhanças em termos de características a uma cabeça de formato elipsoidal normal, a maioria dos 100 pontos (76, para ser mais preciso) que compõem o vetor de assinatura foram classificados como anómalos.

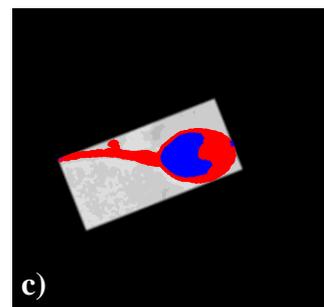
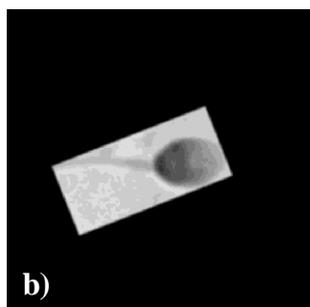
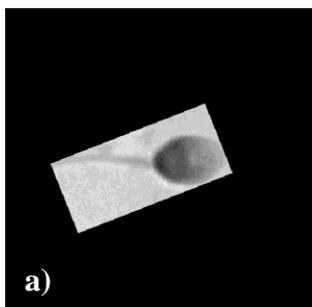
3.4 – Etapa 4 – Análise do Acrossoma e Secção Intermédia

Como algumas das questões abordadas nesta secção já foram tocadas anteriormente, optou-se por limitar a discussão acerca das mesmas a uma única secção.

Tal como foi referido aquando da revisão da literatura, antes de se proceder à análise deste elementos constituintes do espermatozóide tipicamente recorre-se a um conjunto de processos para segmentar os diferentes componentes da cabeça.

Houve uma tentativa inicial para implementar um algoritmo semelhante ao sugerido por A. Bijar *et al* em [29] com base em inferência Bayesiana. Este método baseia-se na modelação da distribuição de brilho dos pixéis da componente Vermelha da imagem (após aplicação de um filtro Gaussiano) das regiões inicialmente extraídas num modelo de mistura gaussiana. Os pixéis no interior das caixas envolventes seriam então segmentados em três kernels referentes ao fundo, ao núcleo e acrossoma/secção intermédia, cujas funções de densidade de probabilidade condicional e probabilidades *a priori* de cada classe seriam obtidas através do uso de um algoritmo com base em EM e MRF. Concluiu-se porém que num conjunto de imagens era difícil lidar com os pixéis da fronteira da região segmentada, que eram muitas vezes segmentados inesperadamente como fazendo parte do acrossoma/secção intermédia.

Como o principal objectivo a atingir com a segmentação dos pixéis pertencentes ao acrossoma é verificar a percentagem da área da cabeça que este constitui, ponderou-se uma solução alternativa para a resolução deste problema. Sabendo a cabeça classificada como normal se encontraria já orientada horizontalmente, considerou-se que o processo poderia eventualmente ser simplificado. Constatou-se que em muito casos esta percentagem pode ser extrapolada fazendo uma única segmentação de uma faixa de pixéis em torno do eixo maior da cabeça (de uma forma geral verifica-se que área do acrossoma tende a distribuir-se uniformemente pela cabeça). Sendo que o ponto terminal contido no eixo maior da região mais escura da cabeça, correspondente ao núcleo, já teria sido determinado para extracção da seção intermédia, decidiu aproveitar-se também esse facto para esta análise. É extraída uma faixa com uma largura de 20 pixéis em torno do eixo maior da cabeça que se estende desde o ponto inicial do eixo localizado no topo da cabeça até ao final da região do núcleo previamente segmentada. Esta faixa é por sua vez segmentada através do cálculo do limiar de binarização pelo método de Otsu, sendo os pixéis mais claros considerados como fazendo parte do acrossoma e os mais escuros como fazendo parte do núcleo. Calcula-se de seguida o rácio entre o total de pixéis presentes na faixa e os pixéis associados ao acrossoma. Se esta percentagem exceder os 40 % e for simultaneamente inferior a 70% o acrossoma será considerado existente e de dimensões adequadas, caso contrário o espermatozóide em questão será descartado da etapa seguinte.



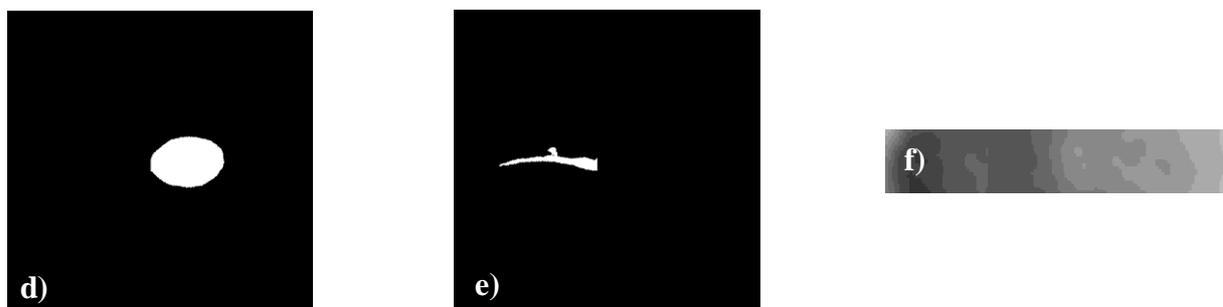


Figura 3.4.1 - Representação do processo de extracção da secção intermédia e classificação do acrossoma: a) Componente Vermelha da Imagem Original orientada horizontalmente; b) Imagem após aplicação de um filtro de média 3x3; c) Segmentação da região para detecção da região correspondente ao núcleo; d) Cabeça extraída; e) Secção intermédia extraída; f) Faixa em torno do eixo maior utilizada para análise das dimensões do acrossoma

A análise da secção intermédia baseia-se principalmente nalguns dos parâmetros já descritos na secção anterior, nomeadamente a sua largura, o seu ângulo de inserção e a distância entre o limite do seu eixo maior em relação ao limite correspondente do eixo maior da cabeça. Muito à semelhança do procedimento tomado para a primeira fase da análise da cabeça pretende-se comparar estes parâmetros com os de alguns espermatozóides de morfologia conhecida de forma a determinar a classificação mais apropriada. Mais especificamente, a intenção é determinar se a espessura se aproxima da esperada para um espermatozóide normal e se existe inserção assimétrica na cabeça ou torção do pescoço. Adicionalmente procura-se determinar a presença de gotas citoplasmáticas distais através da análise do histograma de colunas (número de pixéis a branco em cada coluna da imagem binarizada da secção) de forma a verificar se existe alguma irregularidade na espessura da secção intermédia. Tal como nas etapas anteriores, qualquer espermatozóide que apresente defeitos morfológicos a este nível não deverá ser alvo da análise da sua cauda.

Os detalhes principais desta etapa são ilustrados na **Figura 3.4.1** onde se demonstram a imagem do espermatozóide original, a sua segmentação para obtenção da região aproximada do núcleo, a separação da cabeça e secção intermédia, e ainda a faixa extraída para análise da dimensão do acrossoma.

3.5 – Etapa 5 – Identificação e Caracterização da Cauda

Depois de determinar quaisquer anomalias existentes na região segmentada originalmente, resta ainda a necessidade de identificar os pontos da cauda dos espermatozóides considerados normais até esta etapa de forma a determinar se esta se encontra desdobrada, desenrolada e com as dimensões apropriadas.

De seguida será feito um resumo breve do esquema proposto por A. Bijar *et al* em [28] e [29] utilizado como base desta etapa do processo e dos conceitos teóricos subjacentes ao mesmo.

O método proposto utiliza o Índice de Semelhança Estrutural (SSIM, do inglês *Structural Similarity Index*) e o cálculo da entropia de Rényi de forma iterativa para identificar a cauda, estimando alguns dos seus pontos de forma precisa. Estes pontos poderão eventualmente utilizadas para caracterizar a cauda quanto ao seu comprimento, forma, etc.

O Índice de Semelhança Estrutural [35] é uma medida de qualidade da perspectiva da formação de imagem em função da sua luminância, contraste e estrutura. A principal vantagem desta medida é o facto de geralmente se aproximar bem do critério subjetivo humano. Em particular, tal como o sistema visual humano o SSIM é bastante sensível à degradação na estrutura espacial da luminância da imagem. A luminância da superfície de um objecto observado é o produto da iluminação e da refletância, mas as estruturas dos objectos que constituem a cena são independentes da iluminação. A informação estrutural é definida numa imagem como sendo os atributos que representam a estrutura dos objectos presentes na cena, independentemente da luminância e contraste médios. Já que a luminância e o contraste podem variar ao longo da cena, utilizam-se a luminância e contraste locais.

Suponhamos que $X = \{x_i | i = 1, 2, \dots, N\}$ e $Y = \{y_i | i = 1, 2, \dots, N\}$ são dois sinais de imagem. Consideremos ainda μ_x , σ_x^2 e σ_{xy} a média de X, a variância de X e a covariância de X e Y, respectivamente. Pode-se afirmar que μ_x e σ_x^2 podem ser vistas como estimativas da luminância e o contraste de X, e que σ_{xy} mede a tendência de X e Y variarem em conjunto, servindo assim como uma indicação de semelhança estrutural. A forma geral do Índice de Semelhança Estrutural entre os sinais x e y é assim definido como:

$$SSIM(x, y) = [l(x, y)]^\alpha \cdot [c(x, y)]^\beta \cdot [s(x, y)]^\gamma \quad (10)$$

onde:

- $l(x, y)$ é a medida de comparação de Luminância. A luminosidade deriva da comparação dos valores de média de cada uma das imagens:

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \quad (11)$$

A constante C_1 é incluída para evitar instabilidade quando $\mu_x^2 + \mu_y^2$ se aproxima de 0, e

$$C_1 = (K_1L)^2 \quad (12)$$

onde $K_1 \ll 1$ e a gama dinâmica dos elementos de x e y é representada por L.

- $c(x, y)$ é a medida de comparação do Contraste e é estimada através do desvio padrão e variância de cada uma das imagens:

$$l(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad (13)$$

Muito à semelhança do caso anterior, C_2 é uma constante adicionada para evitar instabilidade quando $\sigma_x^2 + \sigma_y^2$ tendem para um valor próximo de zero, e

$$C_2 = (K_2L)^2$$

onde $K_2 \ll 1$ e L representa a gama dinâmica dos elementos de x e y .

- $s(x, y)$ é a componente que representa a comparação da Estrutura e é estimada após a subtracção da média local e normalização pelo desvio padrão:

$$s(x, y) = \frac{2\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \quad (15)$$

onde $C_3 = C_2/2$.

- α, β e γ são utilizados tipicamente como pesos que denotam a importância relativa das três componentes. Em [29] é sugerido por análise experimental $\alpha = 0.1, \beta = 0.5$ e $\gamma = 0.9$.

A função de semelhança estrutural (SIIM) constituída por $l(x, y)$, $c(x, y)$ e $s(x, y)$ deve satisfazer as seguintes condições:

- 1) Simetria: $S(x, y) = S(y, x)$;
- 2) Limite: $S(x, y) \leq 1$;
- 3) Máximo Único: $S(x, y) = 1$ se e só se $x = y$ (em representações discretas, $x_i = y_i$, para $i = 1, 2, \dots, N$).

Partindo uma imagem digital I de tamanho $(M \times N)$ que contenha um espermatozóide, o propósito do algoritmo é detetar um ponto inicial o mais próximo da cauda possível de forma a determinar qual a janela w_2 : $k \times k$ numa vizinhança determinada por uma janela maior w_1 : $p \times p$, $p > k$, tem maior índice de semelhança estrutural (SSIM) face ao ponto inicial A . Este ponto inicial A é determinado através do cálculo do ponto pertencente à região que apresenta maior distância euclidiana ao canto da *bounding box* mais próximo do centro de massa da mesma, correspondendo este ao ponto distal da secção intermédia. w_2 é então centrada em A e limita a sua vizinhança enquanto que w_1 funciona com uma janela deslizante que se move sobre os limites de w_2 e permite calcular para cada um dos pontos da fronteira o seu valor SSIM relativamente ao ponto A . Este valores são guardados e convertidos para o intervalo $[0, 255]$. O ponto com maior SSIM relativamente a A , identificado como B , é então seleccionado como novo ponto central da janela w_1 de forma a iniciar uma nova iteração do algoritmo. No entanto, se esta selecção fosse executada sem qualquer tipo de cuidado adicional existiria a possibilidade do ponto inicial A ser novamente seleccionado numa próxima iteração do algoritmo. Para evitar isto considera-se uma terceira janela w_3 : $l \times l$, $k < l < p$ que define uma vizinhança em torno do ponto actual a ser colocada a um valor que se distinga apropriadamente da vizinhança, neste caso 0.

O algoritmo é suposto funcionar num esquema iterativo, sendo que para cada novo pixel seleccionado é também computada a entropia local de forma a determinar se o algoritmo já determinou o final da cauda. Com esta finalidade, é utilizada uma janela de dimensão ímpar w_4 : $h \times h$ centrada no novo pixel e as componentes Verde e Vermelha de todos os pixéis no interior

desta janela são extraídas como uma característica de duas dimensões e utilizadas para computar a entropia local correspondente. Na **Figura 3.5.1** encontra-se demonstrado a dinâmica geral do algoritmo descrito em [28] e [29].

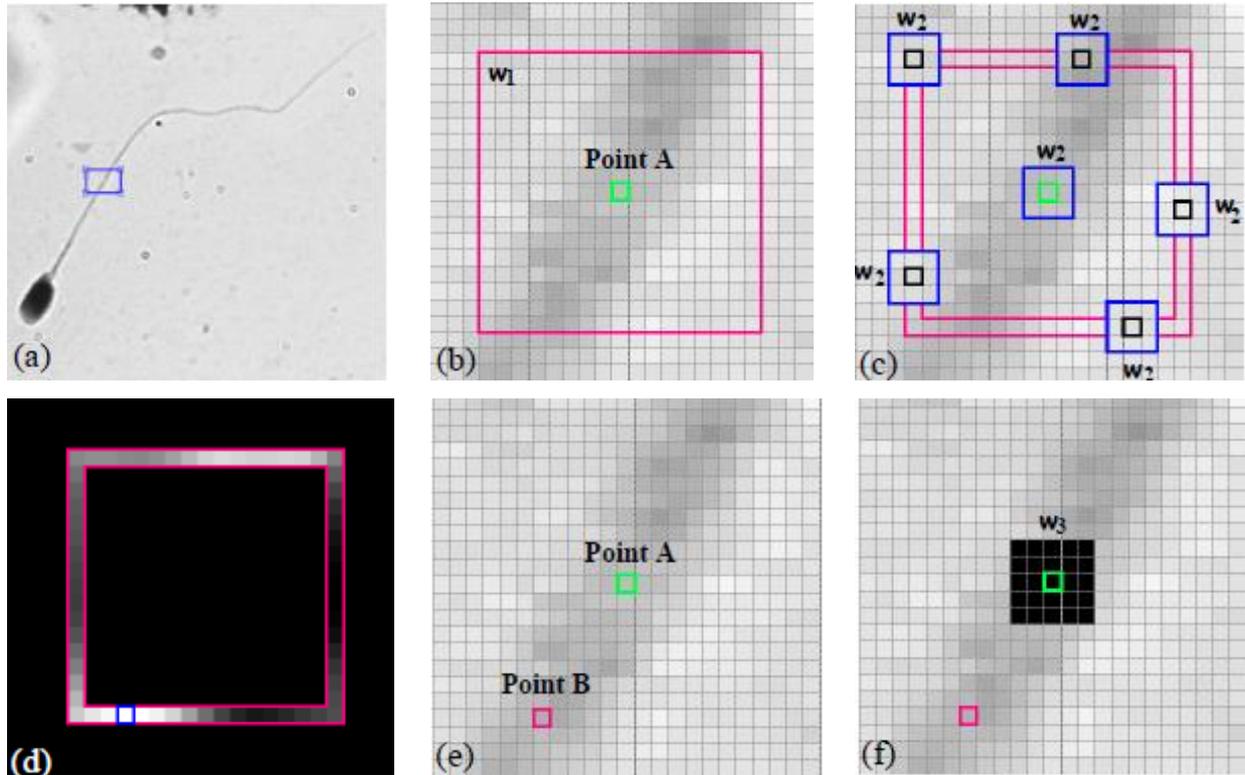


Figura 3.5.1 - Uso do índice SSIM para encontrar os pixels que apresentam maior semelhança: a) Componente vermelha de uma imagem típica de uma amostra; b) Pixel inicial A (verde) e janela w1 (magenta); c) w2 (janela deslizante, a azul), que serve para calcular os valores SSIM para cada um dos pixels sobre os limites de w1; d) Valores SSIM sobre os limites de w1; e) Novo ponto seleccionado B com maior índice SSIM (magenta) e ponto inicial A (verde); f) Uma janela w3 em torno de A é colocada a zero para evitar que o algoritmo regresse a pontos anteriores [28][29]

Na teoria da informação, a entropia é uma medida do grau de incerteza do sistema. O termo entropia foi introduzido neste contexto pela primeira vez por Shannon em 1948 [52].

Seja $X = \{x_1, \dots, x_n\}$ uma variável discreta aleatória com valores em S e função de probabilidade $p(x)$, a informação (ou incerteza) de cada evento possível é

$$I(x_i) = -\log p(x_i) \quad (16)$$

A média de $I(X)$ é introduzida como entropia de Shannon e é representada por $H(X)$

$$H(X) = E(I(X)) = -\sum_{x \in S} p(x) \log p(x) \quad (17)$$

$H(X)$ varia de zero até $\log(|S|)$, com o zero a significar que não existe incerteza e $\log(|S|)$ a apresentar uma situação em que todos os elementos de X tem iguais probabilidades sendo, neste caso, a incerteza máxima [53].

Rényi expandiu este conceito ao definir a entropia de ordem q de uma função densidade de probabilidade $p(x)$ como

$$H_q(X) = \frac{1}{1-q} \log \sum_{x \in S} p(x)^q, q \neq 1 \quad (18)$$

A entropia de Rényi é uma função não crescente de q e pode-se provar, aplicando a regra de l'Hôpital que a entropia de Shannon é a entropia de Rényi de grau 1.

Para estimar este último tipo de entropia foi utilizado no algoritmo sugerido o método desenvolvido por Leonenko *et al.* em [54]. Seja $X \in \mathbb{R}^m$ um vetor aleatório com medida de probabilidade μ e densidade p , o método proposto por Leonenko estima H_q de uma amostra de N variáveis aleatórias, independentes e igualmente distribuídas $X_1, \dots, X_N, N \geq 2$ baseado nas distâncias ao vizinho mais próximo da amostra. Este método surge na continuação do trabalho desenvolvido para calcular H_1 de Kozachenko e Leonenko em [55]. O método de Leonenko estima I_q através da computação dos momentos condicionais das distâncias dos vizinhos mais próximos.

$$I_q = I_q(p) = \mathbb{E}\{p^{q-1}(x)\} = \sum_{x \in S} p(x)^q, q \neq 1 \quad (19)$$

Seja $\rho(x, y)$ a distância euclidiana entre dois pontos x, y de \mathbb{R}^m . Para uma dada amostra $X_1, \dots, X_N, N \geq 2$ com as características já enunciadas e dado um X_i da amostra, das $N-1$ distâncias $\rho(X_i, X_j), j = 1, \dots, N, j \neq i$, são formadas as estatísticas de ordem $\rho_{1,N-1}^{(i)} \leq \rho_{2,N-1}^{(i)} \leq \dots \leq \rho_{N-1,N-1}^{(i)}$, $\rho_{k,N-1}^{(i)}$ é o k -ésima distância de vizinho mais próximo de X_i a outro X_j na amostra, em que $i \neq j$. I_q na equação anterior é estimado para $q \neq 1$, por

$$\hat{I}_{N,k,q} = \frac{1}{N} (\zeta_{N,i,k})^{1-q}, \quad (20)$$

com

$$\zeta_{N,i,k} = (N-1) C_k V_m (\rho_{k,N-1}^{(i)})^m \quad (21)$$

onde $V_m = \pi^{m/2} / \Gamma(\frac{m}{2} + 1)$ é o volume de uma esfera unitária $\beta(0,1)$ em \mathbb{R}^m e

$$C_k = \left[\frac{\Gamma(k)}{\Gamma(k+1-q)} \right]^{1/(1-q)} \quad (22)$$

Assim H_q pode ser estimado da seguinte forma

$$\hat{H}_{N,k,q} = \log(\hat{I}_{N,k,q}) / (1-q) \quad (23)$$

Conforme o algoritmo se aproxima do final da cauda a entropia local tenderá a diminuir. Assim ao determinar um limiar apropriado o algoritmo é parado automaticamente. Segundo [29] foi observado experimentalmente que as componentes Verde e Vermelha permitiam obter um limiar mais rigoroso que as restantes características daí a sua utilização.

Para o limiar determinado, se valor de entropia local estimado para um novo ponto for superior a este, o novo ponto é aceite como fazendo parte da cauda. Contudo determinar este limiar não é uma tarefa fácil, razão pela qual se utiliza uma abordagem com base no cálculo da

média da entropia local de 5 em 5 pixéis comparando-a com um limiar denominado limiar-janela. Este valor ajustado a todos os espermatozóides permite saber se o algoritmo se encontra ou não próximo dos pontos distais da cauda. Se o valor médio calculado estiver acima deste limiar, isto significará que o algoritmo ainda não se encontra perto do final da cauda, caso contrário o limiar individual de cada pixel seleccionado terá de ser comparado com o limiar determinado para o algoritmo em cada uma das próximas 5 iterações. O valor utilizado para determinar este limiar individual é o valor de entropia mínimo (MLE) de entre os últimos 5 pixéis calculada da seguinte forma:

$$\text{limiar} = \delta * MLE$$

com δ definido como 0.9 experimentalmente para o melhor resultado.

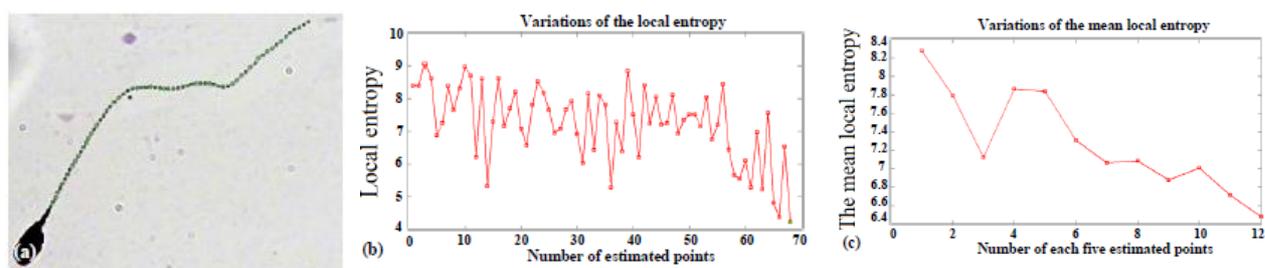


Figura 3.5.2 – Resultados apresentados em [29] para o algoritmo descrito: a) Representação dos pontos estimados pelo algoritmo; b) Variações da entropia local para os pontos da cauda estimados; c) Valor médio das entropias locais, a cada 5 pixéis seleccionados

Foi feita uma primeira abordagem com base neste algoritmo de forma a determinar a sua eficácia face às imagens disponíveis. Apesar de um conjunto de tentativas feito para determinar os valores mais adequados para p , k , l e h , correspondentes às diversas janelas de análise a considerar no algoritmo e restantes parâmetros (α , β e γ que representam os pesos atribuídos a cada uma das componentes consideradas no cálculo do índice SSIM), os resultados apresentaram-se bastante variáveis. Isto ficou-se a dever maioritariamente a dois factores: por um lado, o algoritmo de pesquisa, em muitas circunstâncias e para diferentes parâmetros, era desviado pontualmente da cauda aquando do cálculo do Índice de Semelhança Estrutural para a selecção do novo ponto, geralmente devido à existência de pequenas estruturas na imagem com características semelhantes às dos pixéis da cauda; por outro, a entropia local (calculada através do uso da Toolbox TIM disponível para MATLAB em [57]) para além de geralmente não se revelar um elemento suficientemente discriminante para a deteção do ponto terminal da cauda, apresentava também valores bastante distintos para as diversas imagens entre as diferentes colorações, impossibilitando a determinação experimental de um limiar que permitisse a paragem do algoritmo de forma automática. Foram testadas um conjunto de abordagens para minorar os efeitos do algoritmo, como delimitar as seções da janela w_1 a considerar no cálculo

dos valores SSIM numa nova iteração através do gradiente direccional ou o redimensionamento dinâmica da mesma, que embora tenham resultado em melhorias apreciáveis em algumas das imagens testadas, continuavam a não permitir fazer a paragem automática do algoritmo. Na **Figura 3.5.3** comparam-se os resultados das duas abordagens e verifica-se que, para um mesmo conjunto de parâmetros, os resultados obtidos são melhores quando se considera o gradiente direccional para delimitar o espaço de pesquisa. Em **e)** podemos ainda visualizar que a variação da entropia local ao longo do algoritmo é extremamente reduzida quando comparada com os resultados apresentados em [29] que se encontram representados na **Figura 3.5.2 b)**.

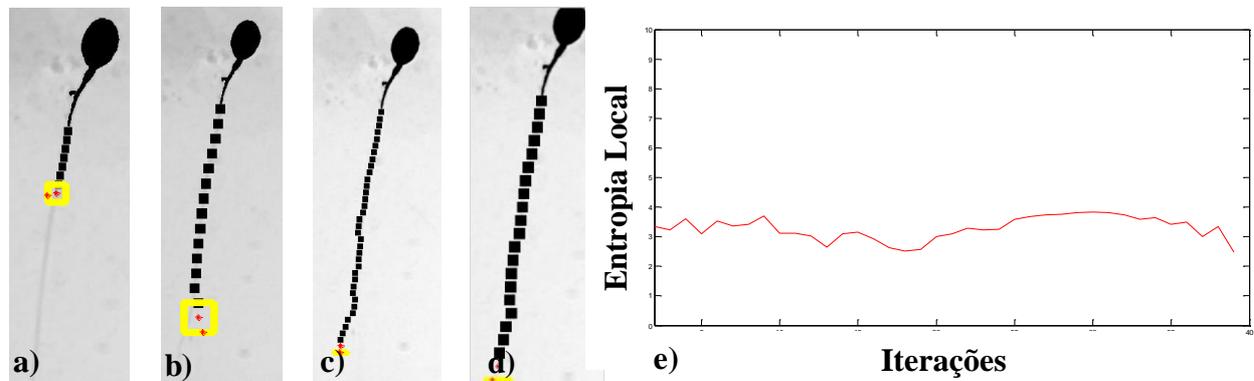


Figura 3.5.3 - Resultados da implementação do algoritmo na imagem de referência: a)-b) Instabilidade demonstrada pelo algoritmo inicial para diferentes valores de k e l ; c)-d) - Melhorias verificadas após restrição da direcção de pesquisa através da consideração do gradiente direccional; e) Gráfico típico da entropia local calculada ao longo das diversas iterações

Para colmatar ambos os problemas optou-se por uma abordagem com base na segmentação local da cauda pelo método de Otsu combinada com o uso do Índice de Semelhança Estrutural com o propósito de determinar os pontos pertencentes a esta. Este algoritmo baseia-se numa interpolação quadrática (seleccionada por se aproximar geralmente bem da morfologia local da cauda) dos pontos da cauda segmentados no fragmento da imagem, de forma a determinar a vizinhança dos pontos limite da curva obtida que pertenceriam à cauda. O Índice de Semelhança Estrutural utilizado para seleccionar desta vizinhança o ponto que mais se aproxima das características esperadas para a cauda. As principais vantagens deste algoritmo baseiam-se no facto de não permitir ao algoritmo seleccionar pontos da vizinhança da cauda que se distanciem notavelmente do seu perfil e de simultaneamente permitir o seu seguimento enquanto exista uma porção de cauda segmentável dentro da janela. Todavia verificam-se problemas em situações em que exista cruzamento de caudas - por haver selecção de pontos da cauda errada (problema também respeitante ao algoritmo proposto em [29]) ou bloqueio num ponto de intersecção devido ao resultado da interpolação podendo o algoritmo desviar-se permanentemente do seu curso devido à selecção de um ponto limite que não permita a sua continuidade.

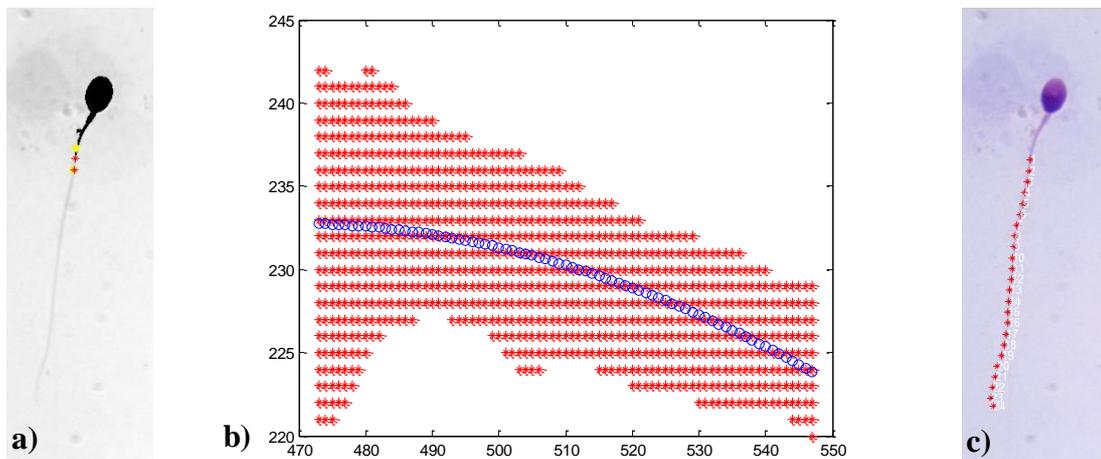


Figura 3.5.4 - Representação dos resultados do algoritmo alternativo proposto: a) Iteração inicial do algoritmo com os pontos inicial e seleccionado a vermelho e a região de pesquisa a amarelo; b) Interpolação quadrática dos pontos da janela local segmentada; c) Pontos detetados como resultado do algoritmo

Os parâmetros seleccionados experimentalmente para obter bons resultados nas diversas imagens disponíveis foram: $p = 49$, $k = 11$ (seleccionado de acordo com [35]), $l = 21$ e α , β e γ iguais aos previamente descritos como tendo sido utilizados em [29]. O algoritmo é dado como terminado quando ao longo das 50 iterações para o qual está configurado para funcionar para uma cauda específica o valor da soma do módulo das derivadas parciais dos seus pontos numa dada iteração atinja um valor inferior a 20, ou seja, quando o algoritmo atinge o ponto terminal da cauda ou um qualquer ponto que impeça o seu progresso.

Devido ao facto de não existir um fator de escala utilizável para obter um valor aproximado dos pixéis em micrómetros ou uma forma eficaz de proceder a uma conversão que permita uma fácil comparação, a classificação das caudas encontra-se um tanto limitada. Foram tomados em consideração para análise das caudas que tenham sido completamente detectadas parâmetros tais como: o número de pontos recolhidos, a distância euclidiana entre o seu ponto inicial e final e as maiores variações das derivadas parciais das coordenadas dos seus pontos, bem como o número de pontos com coordenadas x e y repetidas (para verificar a possível existência de caudas enroladas), embora estes permitam, neste caso, apenas uma análise muito circunstancial da sua morfologia.

A secção seguinte será dedicada a analisar os resultados obtidos nesta e nas restantes etapas de forma a verificar a viabilidade do sistema considerado.

Capítulo 4 – Análise dos Resultados

4.1 - Detecção e Classificação das Cabeças

O primeiro objetivo desta secção é verificar a contagem e classificação de cabeças feita pelo algoritmo de forma a determinar a sua eficiência face às 89 de imagens disponíveis para análise.

Tal como já havia sido mencionado previamente este processo tem início após a extracção da secção correspondente à cabeça das regiões segmentadas da imagem original e encontra-se dividida em duas etapas distintas: uma primeira avaliação das cabeças com base nos parâmetros apresentados na **Figura 2.4**, e uma segunda com base na classificação do vetor de assinatura de contornos recolhido através do uso de um sistema SVM.

Como já foi referido, um dos principais problemas com a primeira etapa deste processo deve-se largamente ao facto de se encontrarem disponíveis apenas cinco imagens cuja classificação da cabeça do único espermatozóide presente se encontra identificada. Estes cinco espermatozóides subdividem-se em apenas três grupos: um único espermatozóide de morfologia normal, 2 espermatozóide com cabeça alongada, um dele com deformidades que o levam a ser classificado simultaneamente como amorfo, e dois espermatozóides amorfos adicionais. Isto restringe tremendamente o que é possível fazer nesta primeira abordagem, pois devido ao facto de os espermatozóide amorfos não apresentarem parâmetros consistentes entre si, isto apenas nos permite distinguir as cabeças alongadas das restantes através da dimensão do seu eixo maior.

Apesar de devido às limitações deste modelo face aos recursos disponíveis, foi considerado de interesse fazer uma breve análise aos parâmetros descritos (em pixéis) para todos os espermatozóides da pasta ‘Morfologia’. Convém realçar novamente que apenas os 5 primeiros espermatozóides se encontravam identificados quanto à morfologia da sua cabeça, sendo estes os principais elementos de base do sistema de classificação.

Nome	Área	Perímetro	Eixo Maior	Eixo Menor	Achatamento	Class. SVM
Normal	7822	361	120	83	1,4	N(100)
Alongada	9566	525	235	54	4,4	A(86)

Amorfa_1	8386	428	153	76	2,0	A(76)
Amorfa_2	9530	581	258	59	4,4	A(94)
Amorfa_3	8589	382	110	101	1,1	A(66)
Cauda_Dobrada	9376	433	172	71	2,4	A(66)
Cauda_Enrolada	8416	401	144	83	1,7	N(100)
Gota_Citoplasmatica	7671	353	111	88	1,3	N(92)
Insercao_Assimetrica	8336	382	138	77	1,8	N(100)
Peca_Intermedia_Espessa	8555	382	142	77	1,8	N(93)

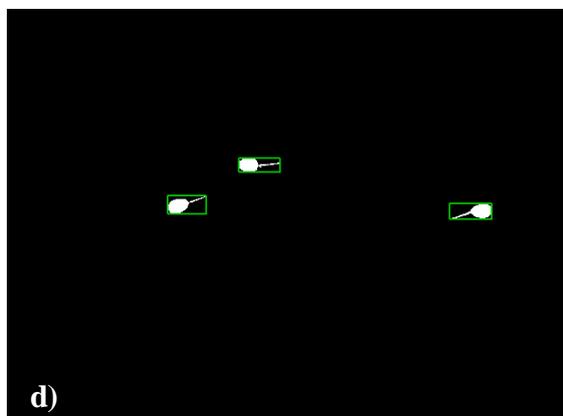
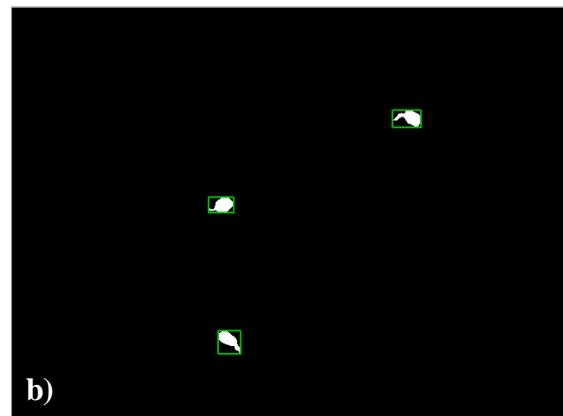
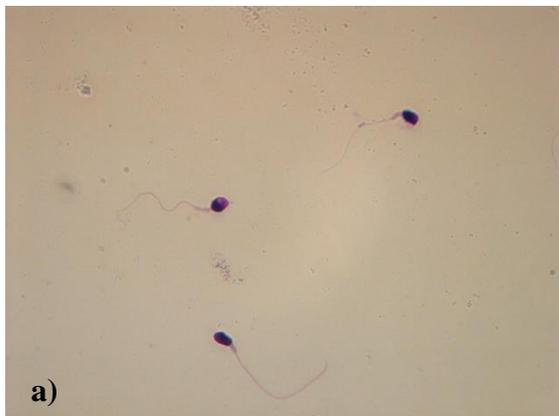
Tabela 4.1.1 - Recolha de parâmetros morfológicos para os espermatozóides presentes nas 10 imagens de referência

Na **Tabela 4.1.1** podemos observar os valores para os diversos parâmetros previamente enunciados para todos os espermatozóides contidos pelas imagens de referência. Foram ignorados todos os parâmetros secundários que caracterizam a cabeça com exceção do achatamento, pois conclui-se que de uma forma geral não serviam como fator discriminante entre os diversos espermatozóides presentes. Na coluna mais à direita verifica-se a classificação atribuída pelo classificador SVM a cada uma das cabeças analisadas, com N representando uma cabeça classificada como Normal, A as cabeças anómalas e os valores entre parênteses o número de pontos de entre os 100 da assinatura de contornos assim classificados. É de recordar que o sistema foi treinado utilizando algumas das assinaturas de contornos destas imagens como conjunto de treino.

O primeiro ponto a destacar é o facto de tanto o espermatozóide contido na imagem ‘Cabeça_Alongada’ como o pertencente a imagem ‘Cabeça_Amorfa_2’ revelarem um achatamento (quociente entre o eixo maior e menor) próximo de 4.4 e valores para o seu eixo maior acima dos 500 pixéis. Apesar de se tratar de uma abordagem grosseira devido à existência de apenas duas amostras, decidiu ter-se ao em conta a média e desvio padrão das dimensões dos seus eixos maiores, verificando se o valor obtido para a nova cabeça a analisar se encontra no

intervalo $[\mu - \sigma, \mu + \sigma]$. Considerou-se também um limiar inferior de 4 para o valor de achatamento, de forma a fazer uma avaliação inicial do alongamento das cabeças.

As restantes cabeças são consideradas por classificar até a sua assinatura de contornos ser classificada através do sistema SVM, momento a partir do qual serão determinadas binariamente como Normais ou Anómalas. Veremos de seguida que, apesar de limitada, esta abordagem permite fazer uma verificação rápida da morfologia da cabeça de múltiplos espermatozóides numa imagem. Para este efeito, e para simplificar a apresentação dos resultados, será utilizado um conjunto de 5 imagens retiradas de entre as diversas colorações, que serão sujeitas a uma contagem dos espermatozóides nelas presentes e classificação das suas cabeças. Em circunstâncias de operação normal do sistema apenas os espermatozóides cuja cabeça fosse classificada como Normal nesta primeira etapa continuariam a ser sujeitos a uma análise dos resultados obtidos nas etapas subsequentes mas, para efeitos de análise, serão consideradas nas diversas etapas todos os espermatozóides da imagem, excetuando circunstâncias em que todos os espermatozóides de uma dada imagem sejam classificados como anómalos numa dada etapa.



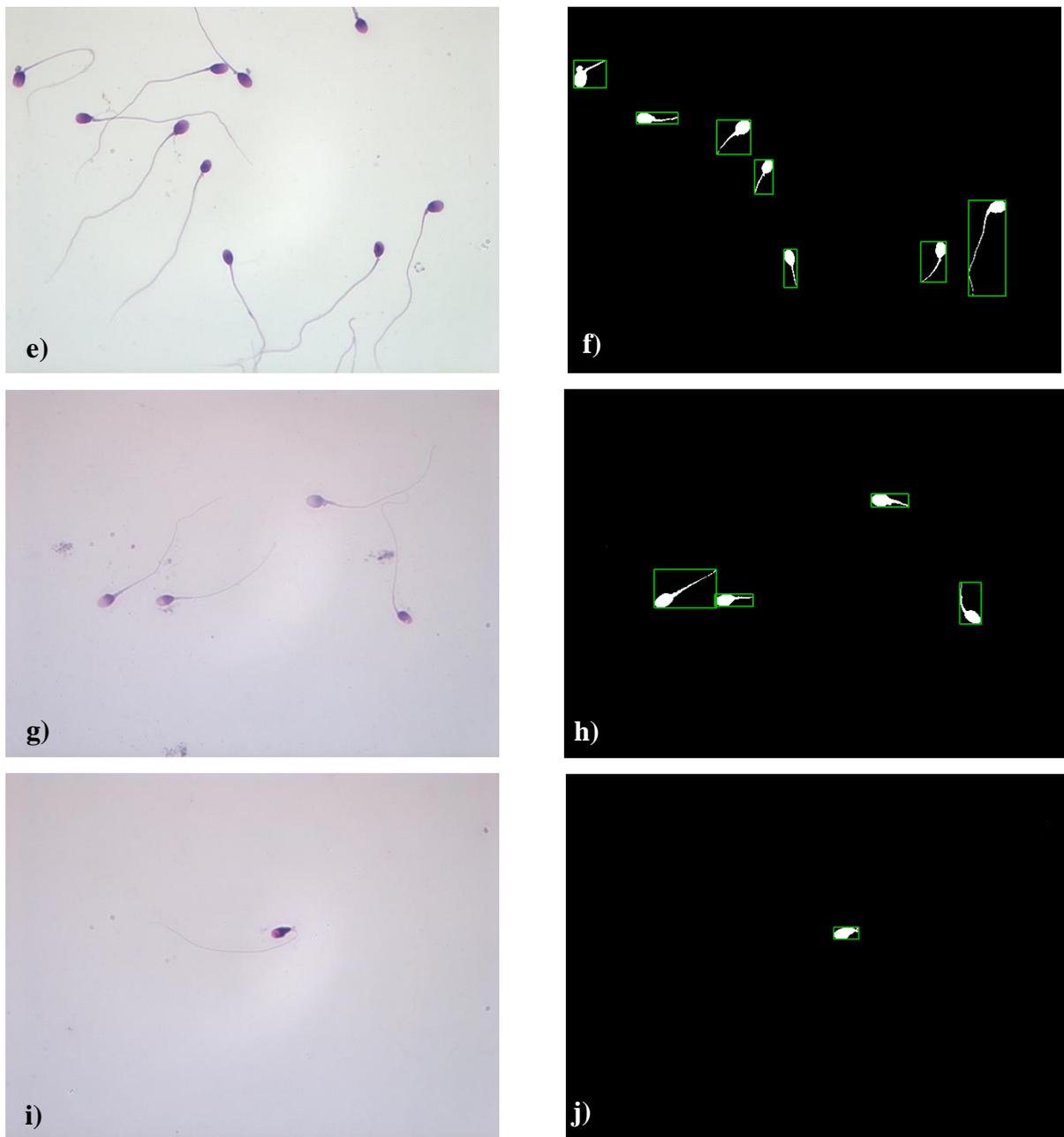


Figura 4.1.1 - Apresentação das Imagens de Análise Originais e o seu aspeto após segmentação: a-b) H3; c-d) H4; e-f) Giemsa_4; g-h) Giemsa_6; i-j) DQ1

Na **Figura 4.1.1** são apresentadas as 5 imagens que passaremos a analisar, bem como a respectiva imagem binarizada contendo as regiões correspondentes à cabeça e secção intermédia dos espermatozóides. É possível observar que de uma forma geral o sistema consegue segmentar eficientemente as várias imagens e isolar os espermatozóides presentes. Porém convém destacar que no caso da imagem **e)** foram removidos 3 espermatozóides aquando da segmentação da imagem. Isto resulta da eliminação de todos os objectos segmentados que se encontrem em contacto com as bordas da imagem.

Na **Tabela 4.1.2** são apresentados os resultados obtidos pelo sistema de classificação SVM, bem como da contagem feita pelo sistema para as 5 imagens apresentadas na **Figura 4.1.1**.

Nome	Nº	Classificação	Perc. Deteções	Falsos Positivos	Perc. Class. Corretas
H3	1	N(83)	100%	0	100%
	2	N(86)			
	3	N(71)			
H4	1	N(97)	100%	0	100%
	2	N(89)			
	3	N(95)			
Giemsa_4	1	N(98)	100% (embora 3 espermatozóides sejam removidos por estar em contacto com a borda da imagem)	0	100%
	2	N(78)			
	3	N(96)			
	4	N(100)			
	5	N(100)			
	6	N(66)			
	7	N(78)			
Giemsa_6	1	N(62)	100%	0	100%

	2	N(90)			
	3	N(97)			
	4	N(73)			
DQ1	1	N(72)	100%	0	100%

Tabela 4.1.2 - Resultados do Sistema de Classificação em SVM para os espermatozoides das imagens ‘H3’, ‘H4’, ‘Giemsa_4’, ‘Giemsa_6’ e DQ1

De uma forma geral pode-se constatar por inspeção visual (já que não foi possível fazer cruzamento com dados clínicos) tanto o processo de segmentação como algoritmo de classificação permitem a obtenção de resultados satisfatórios, sendo capazes de detetar e classificar apropriadamente as cabeças dos espermatozoides em imagens caracterizadas por diferentes colorações.

4.2 - Cálculo da Dimensão do Acrossoma

Nesta breve subsecção, o nosso objetivo será o de registar e comparar os valores obtidos pela análise feita pelo sistema à percentagem do acrossoma face ao núcleo com aquilo que é possível de facto visualizar na imagem. Para esta etapa são consideradas as mesmas imagens que anteriormente (**Figura 4.1.1**).

Nome	Nº	Rácio do Acrossoma	Perc. Class. Corretas
H3	1	46 %	100%
	2	37 %	
	3	16 %	
H4	1	59%	100%
	2	54%	
	3	52%	
Giemsa_4	1	50%	86%
	2	41%	
	3	42%	
	4	41%	
	5	33%	

	6	29%	
	7	42%	
Giemsa_6	1	56%	100%
	2	45%	
	3	54%	
	4	42%	
DQ1	1	42%	100%

Tabela 4.2.1 - Apresentação dos valores obtidos para os valores do rácio do acrossoma face ao núcleo para as imagens em análise com base no método proposto (as classificações erróneas encontram-se destacadas a vermelho)

O sistema mostra-se capaz de detetar de forma aproximada o rácio do acrossoma face ao resto da cabeça na maioria dos casos. A principal exceção é espermatozóide 6 da imagem ‘Giemsa_4’ que apresenta um valor inferior ao esperado por uma rápida inspeção visual, embora esta forma de análise seja subjetiva e mais imprecisa.

4.3 - Análise da Secção Intermédia

Nesta subsecção daremos continuidade ao processo anterior e passar-se-á à análise dos parâmetros da sua secção intermédia dos espermatozoides das diferentes imagens (**Figura 4.1.1**).

Este ponto serve-se exclusivamente da análise das características de alguns dos espermatozoides de referência para permitir fazer uma rápida classificação dos demais. Entre estas características incluem-se o ângulo existente entre o eixo maior da cabeça e o eixo maior do seu acoplamento com a secção intermédia, a distância em pixels existente entre estes no início da secção, a sua largura e uma comparação entre o máximo e o mínimo do seu histograma de colunas. O propósito das primeiras é determinar situações de inserção assimétrica ou pescoço torcido, enquanto que as restantes partem da necessidade de determinar a existência de uma secção intermédia espessa ou a presença de gotas citoplasmáticas.

Nome	Nº	Eixo Menor	Distância	Ângulo	Hist. Colunas	Classificação
Normal	1	16	1	0°	13	Normal
Espessa	1	24	4	3°	9	Espessa
Gota_Citoplasmática	1	32	1	0°	29	Gota Cit.

Inserção_Assimétrica	1	15	5	3°	11	Ins. Ass.
DQ1	1	30	1	0°	7	Espessa
H4	1	10	3	3°	9	Normal
	2	13	3	3°	14	Normal
	3	12	2	0°	6	Normal
H3	1	23	3	3°	14	Espessa
	2	31	5	3°	4	Espessa
	3	29	6	2°	14	Espessa
Giemsa_4	1	17	6	4°	11	Ins. Ass.
	2	19	3	0°	15	Normal
	3	20	3	3°	13	Normal
	4	19	2	0°	15	Normal
	5	16	2	0°	11	Normal
	6	20	6	0°	13	Normal
	7	18	1	3°	12	Normal
Giemsa_6	1	14	3	3°	12	Normal
	2	16	2	0°	10	Normal

	3	32	3	3°	30	Gota Cit.
	4	18	4	7°	15	Normal

Tabela 4.3.1 - Apresentação dos valores para os parâmetros de forma para a secção intermédia dos espermatozóides em análise (os espermatozóide de uma dada imagem que seriam tipicamente eliminados na etapa anterior encontram-se a cinzento)

Existem algumas observações a fazer acerca da **Tabela 4.3.1**. É de notar que tanto no caso típico de secções intermédia espessas como na presença de gotas citoplasmáticas (ou deformidades semelhantes) existe um aumento substancial da dimensão do eixo menor da secção face a uma morfologia considerada normal. Para o caso das gotas citoplasmáticas nota-se também que a diferença entre a média dos mínimos locais do histograma de colunas e do seu máximo absoluto é também notavelmente maior. No que toca às caudas de espessura normal podemos ainda distinguir até certo ponto uma maior distância entre os eixos maiores da cabeça isolada e da região originalmente segmentada no ponto de extracção ligeiramente superior ao valor esperado.

Como é óbvio este método de classificação encontra-se extremamente restringido por um conjunto de factores que vão desde a qualidade da segmentação original à forma como a secção intermédia é extraída, até mesmo à presença de manchas de coloração acopladas à secção. Mas é sobretudo a falta de exemplares a considerar que limita a abordagem a uma combinação da análise visual com os poucos dados disponíveis. Assim sendo, e assumindo a correta segmentação da região, considerou-se que qualquer secção intermédia com eixo menor de dimensão superior ou igual a 23 pixéis deveria ser classificada como espessa, verificando-se de seguida a diferença correspondente ao seu histograma de colunas, considerando-se um limiar inferior de 20 pixéis para considerar a existência de gota citoplasmática. Caso a secção intermédia não se insira em qualquer um dos casos anterior, considera-se que uma distância dos eixos no colo espermático igual ou superior a 5 pixéis para a detecção de uma inserção assimétrica. O ângulo existente entre os dois eixos maiores não se apresentou para as amostras testadas como um fator suficientemente discriminante para qualquer tipo de classificação.

Convém realçar novamente que esta abordagem é bastante grosseira e que idealmente, considerando uma segmentação correta e devida extracção da secção intermédia, deveria ser considerado um método de aprendizagem com um número muito mais elevado de amostras de forma a tornar a classificação mais robusta. Apesar das limitações e de ser difícil precisar a morfologia de todas as secções intermédias presentes por mera inspecção visual, de uma forma geral as classificações feitas aparentam ser consistentes com o que é observável.

4.4 - Detecção dos pontos da Cauda

Neste ponto é colocado à prova o algoritmo desenvolvido para deteção das caudas por forma a detetar os pontos correspondentes às dos espermatozóides considerados viáveis até esta etapa (o que exclui as imagens ‘H3’ e ‘DQ1’). Com base no que foi referido anteriormente, esta etapa tem finalidades que se aproximam mais de mera recolha de dados estatísticos do que propriamente numa classificação eficiente das caudas tendo em conta não só as limitações do algoritmo, como também o facto de ser difícil encontrar uma base de comparação viável para as diferentes imagens. Na tabela seguinte são apresentados um conjunto de dados referentes às caudas das imagens analisadas e da cauda do espermatozóide de morfologia normal utilizado como referência:

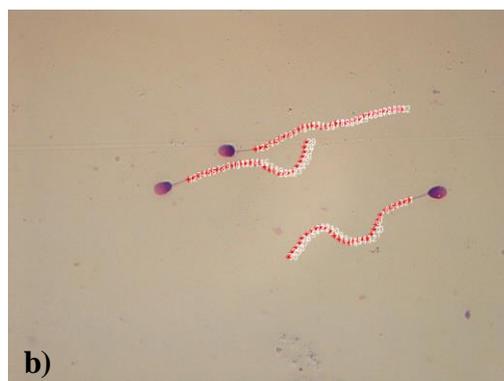
Nome	Nº	Nº Pontos	Distância Total
Normal	1	35	841
H3	1	29	625
	2	32	762
	3	30	670
Giemsa_4	1	33	321
	2	1	-
	3	35	883
	4	32	828
	5	20	462
	6	26	626
	7	18	424
Giemsa_6	1	29	696
	2	21	516
	3	26	579
	4	20	460

Tabela 4.4.1 - Apresentação dos resultados obtidos para o algoritmo de detecção da cauda proposta para os espermatozóides das imagens em análise (destacam-se a vermelho os casos em que o algoritmo não permite fazer um seguimento completo da cauda e a cinzento os espermatozóides eliminados em etapas anteriores)

Existe um conjunto de observações a fazer acerca dos dados apresentados na **Tabela 4.4.1**. Tanto no caso da imagem ‘Giemsa_4’ como em ‘Giemsa_6’ verifica-se um caso de uma cauda que o algoritmo não conseguiu detetar por completo: no primeiro caso isto deve-se ao facto do ponto inicial do algoritmo se encontrar sobre a intersecção de duas caudas o que leva à

segmentação de um conjunto de pontos que interpolados resultam num ponto bastante próximo do inicial; no segundo caso verifica-se um problema de proximidade com uma outra cauda, o que resulta também numa situação de cariz semelhante. Aquando da utilização do algoritmo sobre o espermatozóide 6 da imagem ‘Giemsa_4’ leva a que a partir do cruzamento de cauda passem a ser detetados pontos de uma outra cauda.

Apesar do algoritmo se mostrar capaz de fazer uma detecção eficaz dos pontos das caudas, o valor classificativo dos dados adquiridos é bastante limitado. O algoritmo mostrou-se incapaz de fazer o seguimento das imagens de referência ‘Cauda_Dobrada’ e ‘Cauda_Enrolada’ devido à fraca qualidade das imagens e ao grau do defeito presente em ambas as imagens, o que dificulta substancialmente o funcionamento esperado do algoritmo. Torna-se assim difícil quantificar possíveis defeitos desta natureza que poderiam doutra forma ser determinados pela variação do gradiente direccional dos pontos adquiridos, ou mesmo da distância reduzida que os pontos inicial e final apresentariam para uma cauda dobrada. Para além disto, é possível ainda constatar que devido à falta de amostras com morfologia conhecida *a priori* e ao desconhecimento dos parâmetros sob os quais foi realizada a captura inicial foi realizada, é impossível classificar com precisão as caudas constituídas pelos pontos detetados. Apesar de existir alguma consistência geral relativamente ao número de pontos adquiridos e à distância euclidiana entre os pontos iniciais e finais (com exceção de caudas incompletas por intersecção com os limites da imagem ou, no caso, da cauda do espermatozóide 1 da imagem ‘Giemsa_4’ devido ao facto desta se encontrar ligeiramente dobrada), é difícil garantir de forma minimamente precisa em que intervalos estes e outros parâmetros se deveriam encontrar para garantir a correta classificação morfológica da cauda. Na **Figura 4.4.1** apresentam-se os resultados da aplicação do algoritmo sobre as imagens ainda consideradas nesta etapa, representando a vermelho os pontos detetados sobre as caudas presentes nas imagens originais consideradas nesta breve análise.



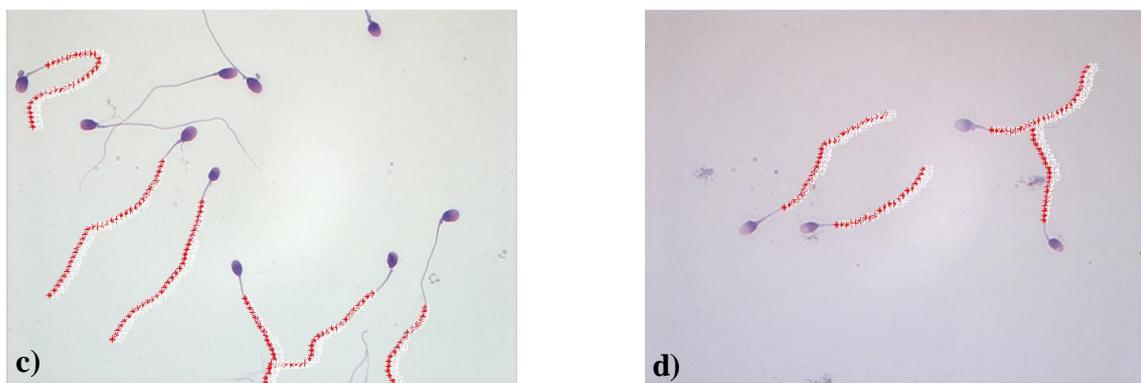


Figura 4.4.1 - Resultados do algoritmo de detecção de caudas proposto para as imagem de referência e imagens de análise não eliminadas nas etapas anteriores: a) Normal; b) H3; c) Giemsa_4; d) Giemsa_6

4.5 – Resultados Gerais

Antes de terminar este capítulo será conveniente fazer a apresentação dos resultados gerais obtidos para as 89 imagens, para as quais se procedeu à contagem do número de espermatozoides presentes, o número de falsos positivos e ainda a percentagem de classificações corretas para a análise dos diversos elementos constituintes do espermatozoide.

Pasta	Nº Deteções	Falsos Positivos	Perc. Deteções	Perc. Class. Corretas (SVM)	Perc. Class. Corretas (Sec. Int.)	Perc. Class. Corretas (Acro.)
Morfologia	10	0	100%	100%	100%	100%
Spz Diff- Quik	33	0	100%	100%	100%	100%
Hemacolor	23	2	95,8%	100%	86,9%	91,3%

Giemsa	43	0	83,7%	93,0%	88,3%	90,7%
Fotos Hemacolor 15-5-07	56	3	98,2%	94,6%	92,7%	92,9%
Diff Quik	24	0	100%	95,8%	87,5%	91,7%
Total	198	5	95,6%	97,4%	92,9%	93,9%

Tabela 4.5.1 - Apresentação dos resultados gerais obtidos para as diversas componentes do sistema

Pela **Tabela 4.5.1** verifica-se que apesar do número limitado de amostras, os resultados são geralmente promissores. Foram contados um conjunto de 198 espermatozoides de um total de 207, para os quais se obteve uma percentagem de 97,4% classificações corretas para o sistema de classificação com base em SVM, 93,9% para a verificação das dimensões do acrossoma e 92,9% para a classificação da Secção Intermédia. É de recordar mais uma vez que não houve cruzamento com quaisquer dados clínicos, por isso estes resultados provêm de uma análise visual das imagens disponíveis com todas a subjectividade que daí advém.

Capítulo 5 – Conclusão e Trabalho Futuro

Este último capítulo é dedicado à revisão sumária do trabalho realizado com o propósito de criar um sistema de avaliação e classificação de espermatozóides de acordo com a sua morfologia.

Apesar da natureza largamente exploratória da abordagem feita, estipulou-se inicialmente que o sistema teria como base a análise morfológicas das diversas componentes que constituem um espermatozóide: cabeça, acrossoma, secção intermédia e cauda. Após levantamento da literatura relevante foi feita a decisão de seguir uma abordagem semelhante à dos sistemas SAC utilizados comercialmente com base na análise de um conjunto de parâmetros utilizados para classificar a morfologia da cabeça e secção intermédia do espermatozóide. Devido ao desconhecimento das condições de captura das diversas imagens distribuídas entre múltiplas colorações, e apesar do número extremamente limitado de imagens contendo espermatozóides de morfologia previamente conhecida, decidiu-se proceder a uma classificação com base num modelo de aprendizagem. Pretendia-se também desenvolver um sistema para detecção dos pontos da cauda que posteriormente poderiam ser utilizados para a sua avaliação.

O sistema proposto partiu da conjugação de um conjunto de métodos propostos na literatura, procurando no entanto sugerir um conjunto de alternativas e simplificações de forma a adequar-se aos recursos disponíveis. Incluem-se neste conjunto de soluções alternativas os métodos de segmentação da secção intermédia e acrossoma, bem como o método de interpolação local para determinar os pontos da cauda. Como foi possível observar através da **Tabela 4.5.1** os resultados obtidos são de uma forma geral promissores, embora ainda restem um conjunto de melhorias a serem desenvolvidas para otimizar o desempenho do sistema.

Ao terminar é possível afirmar que o sistema cumpre minimamente os objectivos propostos, mas, acima de tudo, propõe soluções que poderão ser refinadas para obtenção de um sistema mais eficiente e capaz de uma avaliação mais robusta dos diversos elementos constituintes de um dado espermatozóide.

De seguida são sugeridas um conjunto de possíveis melhorias a implementar com base nas conclusões retiradas, e que não foram consideradas para o âmbito deste projeto devido ao tempo limitado disponível para o seu desenvolvimento:

- Seria impreterível procurar obter mais imagens digitais de amostras para referência obtidas segundo parâmetros consistentes de preparação e captura de forma a facilitar o processo de classificação;
- Melhorar o desempenho computacional do algoritmo utilizado para tratamento da imagem;

- Procurar um método para lidar com regiões da segmentação de múltiplas cabeças em contacto utilizando, por exemplo, o método de segmentação baseado em Watershed sugerido em [24];
- Implementação de métodos mais robustos para a segmentação das regiões que delimitam o núcleo, o acrossoma e a secção intermédia;
- Testar alternativas para a detecção dos pontos da cauda mais consistentes com a capacidade de lidar com a sobreposição de artefactos sobre a cauda e com critérios de paragem automática mais gerais.
- Desenvolvimento de um sistema eficiente de classificação para os pontos da cauda.

Bibliografia

- [1] Sá, R., Sousa, M., Barros, A. “Manual do Casal Infértil”, Biologia na Noite, Amadeu Soares (Editor), Editora Afrontamento, Porto, 2007.
- [2] Tavares, R., Morais, S., “Qualidade Espermática”, Departamento de Ciências da Vida, Faculdade de Ciências e Tecnologia da Universidade de Coimbra, Coimbra, 2009/2010.
- [3] World Health Organization (1999) WHO laboratory manual for examination of human semen and sperm cervical mucus interaction. 4th Edition, Cambridge University Press, Cambridge.
- [4] K. M. Silverberg, T. Turner, “4 – Evaluation of Sperm”, em *Textbook of Assisted Reproductive Technologies: Laboratory and Clinical Perspectives*, 39-52. Eds Gardner DK, Weissman A, Howles CM, Shoham Z, Third edition. Informa Healthcare, London.
- [5] Reproductive Care Center, “Sperm Analysis” e outros, <http://www.fertilitydr.com/semen-analysis.html>, Sandy, Utah
- [6] The Fertility Institutes, “Sperm Evaluation” e outros, http://www.fertility-docs.com/sperm_eval.html.
- [7] Wang, Y., Jia, Y., Yuchi, M., Ding, M. (2011) “The Computer-assisted Sperm Analysis (CASA) Technique for sperm Morphology Evaluation”, *International Conference on Intelligent Computation and Bio-Medical Instrumentation*, 2011, p. 279-282
- [8] Maree, L., du Plessis, SS., Menkveld, R. e van der Horst, G. (2010) “Morphometric dimensions of the human sperm head depend on the staining method used”, *Human Reproduction*, Vol.25, nº .6, p.1369-1382
- [9] Carrillo, H., Villarreal, J., Sotaquira, M., Goelkel, M. e Gutierrez, R. (2005) “Spermatozoon segmentation towards an objective analysis of human sperm morphology.” *Proceedings of the 5th International Symposium on image and Signal Processing and Analysis*, Zagreb, 15-17 Setembro de 2005, p. 522-527.
- [10] Carrillo, H., Villarreal, J., Sotaquira, M., Goelkel, M. and Gutierrez, R. (2007) A computer aided tool for the assessment of human sperm morphology. *Proceedings of the 7th IEEE International Conference on Bioinformatics and Bioengineering (BIBE)*, Boston, 14-17 Outubro de 2007, p. 1152-1157.
- [11] Nowshiravan Rahatabad, F., Moradi, M.H. e Nafisi, V.R. (2005) “A multi steps algorithm for sperm segmentation in microscopic image.” *Proceedings of the World Academy of Science, Engineering and Technology*, Vol. 12, p. 43-45.
- [12] Nafisi, V.R., Moradi, M.H. and Nasr-Esfahani, M.H. (2005) “Sperm identification using elliptic model and tail detection.” *Proceedings of the World Academy of Science, Engineering and Technology*, Vol. 6, p. 205-208.

- [13] Otsu, N. (1979) “A threshold selection method from gray- level histograms.” *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 9, pp. 62-66.
- [14] Abbiramy, V.S. e Shanthi, V. (2010) “Spermatozoa segmentation and morphological parameter analysis based detection of teratozoospermia.” *International Journal of Computer Applications*, Vol. 3, pp. 19-23.
- [15] RC., Johnston, GN., Clarke, DY. Liu, BHW., Gordon, (1995) “Assessment of the sperm quality analyzer,” *Fertil Steril*, Vol. 63, p. 1071-1076, Maio de 1995
- [16] S., Komori, Y., Hamada, A., Hasegawa, H., Tsubamoto *et al.* (2003) “A digital method of sperm immobilization test: comparison to the conventional method.” *American Journal of Reproductive Immunology*, Vol. 50, n°. 6, p.481-484, Dezembro de 2003
- [17] Agarwal, Ashok e Sharma, Rakesh K., “Automation is the key to standardized semen analysis using the automated SQA-V sperm quality analyzer,” *Fertility and Sterility*, Vol. 87, n°.1, Janeiro de 2007
- [18] Soler, C., Monserrat., JJ. de, Gutiérrez, R., *et al* (2003) “Use of the Sperm-Class Analysis for objective assessment of human sperm morphology,” *International Journal of Andrology*, Vol. 26, p. 262-270, 2003
- [19] Blanchard, M., Haguenoer, K., Apert, A., Poret, H., Barthelemy, C., Royere, D. e Guerif, F. (2010) “Sperm morphology assessment using David’s classification: time to switch to strict criteria? Prospective comparative analysis in a selected IVF population”, *International Journal of Andrology*, Vol. 34, p. 145-152, 2010
- [20] Leung, C. e Lu, Z. “Detection and Tracking of Low Contrast Human Sperm Tail”, *Automation Science and Engineering CASE 2010 IEEE Conference*, 2010, p. 263-268
- [21] Abbiramy, V.S., Shanthi, V. e Allidurai, C. (2010) ” Spermatozoa detection, counting and tracking in video streams to detect Asthenozoospermia.” *Signal and Image Processing (ICSIP)*, Chennai, p.265-270, 2010.
- [22] Ravanfar, M. e Moradi, M. (2011) “Low Contrast Sperm Detection and Tracking by Watershed and Particle Filter”, *18th Iranian Conference of Biomedical Engineering ICBME*, p. 260-263, 2011
- [23] Luo, G., Huang, W. e Li, S. (2010) “2-D Maximum Entropy Spermatozoa Image Segmentation Based on Canny Operator”, *Intelligent Computing and Integrated Systems ICISS 2010 International Conference*, 2010, p. 0-3
- [24] Wang, D. e Valloton (2010) P. “Improved Marker Controlled Watershed Segmentation with local Boundary Priors”, *25th Conference of Image and Vision Computing New Zealand (IVCNZ)*, 2010, p. 1-6

- [25] National Programme on Technology Enhanced Learning, “Power Transform” e outros, http://www.nptel.iitm.ac.in/courses/Webcourse-contents/IIT-KANPUR/Digi_Img_Pro/chapter_8/8_14.html
- [26] Abidi, M. A. e. Gonzalez, R. C., “Data Fusion in Robotics and Machine Intelligence”, Academic Press, Nova Iorque, 1992
- [27] Adams, R., Bischof, L. (1994) “Seeded Region Growing”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 16, n°. 6, p. 641-647, Jun. 1994
- [28] Bijar, A., Mikaeli, M., “Sperm’s tail identification and discrimination in microscopic images of stained human semen smear”. *Proceeding of the 7th International Symposium on Image and Signal Processing and Analysis (ISPA)*, Croatia, 4-6 Setembro de 2011, p.709-714.
- [29] Bijar, A., Mikaeli, M., Benavent, A. P. e Khayati, R. (2012) “Fully Automated Identification and Discrimination of Sperm’s Parts in Microscopic Images of Stained Human Smear”, *Journal of Biomedical Science and Engineering*, Vol. 5, p. 384-395, 2012
- [30] Rajan, J., Kanaan, K. e Kaimal, M. R. (2008) “An improved hybrid model for molecular image denoising.” *Journal of Mathematical Imaging and Vision*, Vol. 31, p. 73-79, 2008.
- [31] Pirzadeh, H. “Computational geometry with the rotating calipers.” Tese de Mestrado, *School of Computer Science, McGill University*, Montreal, 1999
- [32] Toussaint, G. T. (1983) “Solving geometric problems with the rotating calipers.” *Proceedings of IEEE MELECON83*, Atenas, Maio de 1983, p. 24-26
- [33] Benavent, A.P., Ruiz, F.E. e Sáez, J.M. (1999) “Learning gaussian mixture models with entropy-based criteria” *IEEE Transactions on Neural Networks*, Vol. 20, p. 1756-1771, 1999.
- [34] Bijar, A., Khanloo, M.M., Benavent, A.P. e Khayati, R. (2011) “Segmentation of MS lesions using entropy-based EM algorithm and Markov random fields.” *Journal of Biomedical Science and Engineering (JBISE)*, Vol. 4, p. 552-556, 2011.
- [35] Wang, Z., Bovik, A.C., Sheikh, H.R. e Simoncelli, E.P. (2004) “Image quality assessment: From error measurement to structural similarity.” *IEEE Transaction on Image Processing*, Vol. 13, p. 1-14.
- [36] Rényi, A. (1961) “On measures of entropy and information.” *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, 20 Junho-30 de Julho 1960, p. 547-561
- [37] Alegre, E., Rodriguez, R.A., Barreiro, J. e Ruiz, J. (2009) “Use of contour signatures and classification methods to optimize tool life in metal machining”, *Estonian Journal of Engineering*, Vol. 15, n° 1, p. 3-12, Janeiro de 2009
- [38] MathWorks Documentation Center, “Support Vector Machines (SVM)”, <http://www.mathworks.com/help/stats/support-vector-machines-svm.html>

- [39] You, Y.L. and Kaveh, M. (2000) “Fourth-order partial differential equations for noise removal.” *IEEE Transactions on Image Processing*, Vol. 9, p. 1723-1730, 2000.
- [40] Perona, P. and Malik, J. (1988) “Scale-space and edge detection using anisotropic diffusion.” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 12, p. 629-639, 1988
- [41] Hamza, A.B., Escamilla, P.L., Aroza, J.M. and Roldan, R. (1999) “Removing noise and preserving details with relaxed median filters.” *Journal of Mathematical Imaging and Vision*, Vol. 11, p. 161-177, 1999
- [42] Hamza, A.B. and Krim, H. (2001) “Image denoising: A nonlinear robust statistical approach.” *IEEE Transactions on Signal Processing*, Vol. 49, p. 3045-3054, 2001
- [43] Andrew, A. (1997) “Another efficient algorithm for convex hulls in two dimensions.” *Information Processing Letters*, Vol. 9, p. 216-219, 1997
- [44] Fitzgibbon, A., Pilu, M. e Fisher, R.B. (1999) “Direct Least Square Fitting of Ellipses”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 21, n°. 5, Maio de 1999
- [45] Fitzgibbon, A., Pilu, M. e Fisher, R. B., Código Matlab de “Direct Least Square Fitting of Ellipses”, http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/PILU1/, Department of Artificial Intelligence, University of Edinburgh, Edimburgo, Escócia, Julho 1996
- [46] Peixoto, P., Gonçalves, J. e Araújo, H. (2002) “Real-Time Gesture System Based on Contour Signatures”, *Proceedings of the 16th International Conference on Pattern Recognition (ICPR'02)*, 2002, Vol. 1, p. 447-450
- [47] Schmiddsberger, F. e Stolzenburg, F. (2011) “Semantic Object Recognition in Digital Images”, *Frieder Stolzenburg and Florian Ruh*, p. 67-72, 2011
- [48] Lorena, A. C., Carvalho, A. C. (2007), “Introdução às Support Vector Machines”, *Revista de Informática Teórica e Aplicada (RITA)*, Vol. 14, n° 2,43-67, 2007
- [49] Vapnik, V. “The Nature of Statistical Learning Theory” (2005), *Data mining and knowledge discovery*, Springer, 2005
- [50] Haykin, S. (1994) ”Neural Networks: A Comprehensive Foundation”, 1ª Edição, *Prentice Hall PTR Upper Saddle River*, EUA, 1994
- [51] Vapnik, V. (1998) “Statistical Learning Theory”, Wiley-Blackwell, 12 Outubro de 1998
- [52] Shannon, C. (1948) “A Mathematical Theory of Communication”, *Bell System Technical Journal*, Vol. 27, p. 379-423, 1948.
- [53] Cover, M. e Thomas, J.A. (1991)“Elements of Information Theory.”, Wiley and Sons, Nova Iorque, EUA, 1991
- [54] Leonenko, N., Pronzato, L. e Savani, V. (2005) “A Class of Renyi Information Estimators for Multidimensional Densities”, *The Annals of Statistics*, Vol. 36, n° 5, p. 2153-2182, 2005.

- [55] Kozachenko, L., Leonenko, N. (1987) “On statistical estimation of entropy of a random vector”. *Problems Inform. Transmission*, Vol. 23, p. 95-101, 1987.
- [56] Sampaio, R. “Automated evaluation of gamete and embryo quality for assisted reproduction”, Tese de Mestrado, Departamento de Engenharia Eletrotécnica, Faculdade de Ciências e Tecnologia - Universidade de Coimbra, Setembro de 2012
- [57] Homepage of Kalle Rytanen, “TIM 1.2.0 : A C++ library for efficient estimation of information-theoretic measures from time-series in arbitrary dimensions”, <http://www.cs.tut.fi/~timhome/tim/tim.htm>, 2011

Anexo I - Resultados da classificação SVM

Imagens Spz Diff-Quik

Nome	Nº	Classificação	Perc. Deteções	Falsos Positivos	Perc. Class. Corretas
1	1	N(68)	100%	0	100%
2	1	A(74)	100%	0	100%
3	1	N(92)	100%	0	100%
4	1	N(97)	100%	0	100%
5	1	N(93)	100%	0	100%
6	1	N(64)	100%	0	100%
7	1	N(81)	100%	0	100%
8	1	N(68)	100%	0	100%
9	1	A(74)	100%	0	100%
10	1	N(66)	100%	0	100%
11	1	N(81)	100%	0	100%
12	1	N(76)	100%	0	100%
13	1	N(98)	100%	0	100%

14	1	N(68)	100%	0	100%
15	1	N(74)	100%	0	100%
16	1	N(78)	100%	0	100%
17	1	N(80)	100%	0	100%
18	1	N(76)	100%	0	100%
19	1	N(75)	100%	0	100%
20	1	A(59)	100%	0	100%
21	1	A(62)	100%	0	100%
22	1	A(70)	100%	0	100%
23	1	A(61)	100%	0	100%
24	1	N(60)	100%	0	100%
25	1	N(91)	100%	0	100%
26	1	N(81)	100%	0	100%
27	1	N(71)	100%	0	100%
28	1	N(100)	100%	0	100%
29	1	N(100)	100%	0	100%
30	1	N(59)	100%	0	100%

31	1	N(83)	100%	0	100%
32	1	N(96)	100%	0	100%
33	1	N(76)	100%	0	100%

Tabela IV.1 - Resultados do Sistema de Classificação em SVM para os espermatozóides da pasta 'Spz Diff-Quik'

Hemacolor

Nome	Nº	Classificação	Perc. Deteções	Falsos Positivos	Perc. Class. Corretas
H1	1	N(81)	100%	1	100%
	2	N(84)			
	3	A(56)			
	4	N(97)			
H2	1	A(67)	100%	0	100%
	2	N(88)			
H3	1	N(83)	100%	0	100%
	2	N(86)			
	3	N(71)			
H4	1	N(97)	100%	0	100%
	2	N(89)			

	3	N(95)			
H5	1	N(72)	89%	0	100%
	2	N(89)			
	3	N(82)			
	4	A(59)			
	5	N(71)			
	6	A(51)			
	7	N(82)			
	8	A(84)			
H6	1	N(76)	100%	0	100%
	2	N(85)			
H7	1	N(82)	100%	1	100%

Tabela IV.2 - Resultados do Sistema de Classificação em SVM para os espermatozóides da pasta ‘Hemacolor’

Fotos Hemacolor 15-5-07

Nome	Nº	Classificação	Perc. Deteções	Falsos Positivos	Perc. Class. Corretas
hemacolor1	1	A(78)	100%	0	100%
	2	N(76)			
	3	N(82)			
	4	N(80)			
	5	N(67)			
	6	N(71)			
	7	N(84)			
	8	N(78)			
	9	N(89)			
hemacolor2	1	A(57)	100%	0	100%
	2	N(97)			
	3	N(64)			
hemacolor3_cabeca_amorfa	1	A(60)	100%	0	100%
hemacolor3_insercao_ass	1	N(81)	100%	0	100%

hemacolor3_ peca_intermedia_espada	1	N(67)	100%	0	60%
	2	A(59)			
	3	A(63)			
	4	A(54)			
	5	N(89)			
hemacolor4_ escuros	1	A(71)	100%	1	86%
	2	N(67)			
	3	A(79)			
	4	N(82)			
	5	N(92)			
	6	N(75)			
	7	N(80)			
hemacolor5_ claro	1	N(90)	100%	1	100%
	2	N(92)			
hemacolor6_ claro	1	A(64)	100%	0	100%
	2	A(69)			
	3	A(73)			

hemacolor7_ normal	1	N(100)	100%	0	100%
hemacolor8_ claros_e_esc uros	1	N(65)	100%	1	100%
	2	N(82)			
	3	N(76)			
	4	N(72)			
	5	N(86)			
	6	N(91)			
	7	N(68)			
Hemacolor9_ cabeca_amor fa	1	A(61)	100%	0	100%
	2	A(81)			
	3	A(66)			
	4	N(70)			
	5	N(72)			
hemacolor10 _cauda_parti da	1	A(66)	75%	0	100%
	2	N(84)			
	3	A(73)			
Hemacolor11 _cabeca_amo	1	A(53)	100%	0	100%

rfa	2	A(88)			
hemacolor12 _cabeca_alon gada	1	N(70)	100%	0	100%
	2	A(86)			
	3	A(59)			
	4	N(72)			
hemacolor13 _cauda_enro lada	1	A(61)	100%	0	100%
	2	N(81)			
	3	N(86)			

Tabela IV.3 - Resultados do Sistema de Classificação em SVM para os espermatozoides da pasta 'Fotos Hemacolor 15-5-

07'

Diff-Quik

Nome	Nº	Classificação	Perc. Deteções	Falsos Postivos	Perc. Class. Corretas
DQ1	1	N(67)	100%	0	100%
DQ2	1	N(91)	100%	0	100%
	2	N(74)			
	3	N(68)			
	4	A(68)			

	5	A(70)			
	6	N(98)			
DQ3	1	N(58)	100%	0	100%
	2	A(70)			
	3	N(64)			
	4	N(69)			
	5	A(65)			
	6	A(56)			
DQ4	1	N(90)	100%	0	100%
	2	N(62)			
	3	N(78)			
	4	N(70)			
DQ5	1	A(82)	100%	0	100%
	2	N(79)			
	3	N(82)			
	4	N(90)			
	5	N(71)			

	6	N(77)			
DQ6	1	N(59)	100%	0	100%

Tabela IV.4 - Resultados do Sistema de Classificação em SVM para os espermatozoides da pasta 'Diff Quik'