



FEUP Universidade do Porto
Faculdade de Engenharia

Análise automática de imagens de culturas
celulares para avaliação da eficiência da
transferência de genes

João Filipe Correia Ribeiro

Julho de 2006

Orientadora: Prof. Ana Maria Mendonça

Co-orientadora: Prof. Ana Paula Pêgo

621.3(047.3)/
LEEC
2006/RIBj

ção





FEUP Universidade do Porto
Faculdade de Engenharia

Análise automática de imagens de culturas
celulares para avaliação da eficiência da
transferência de genes

João Filipe Correia Ribeiro

Julho de 2006

Orientadora: Prof. Ana Maria Mendonça

Co-orientadora: Prof. Ana Paula Pêgo

Ciência.
Inovação
2010



Para a minha mãe

621.3107(3) ILCC 2005/12/10

105237

24 02

10

Agradecimentos

Embora este projecto tenha uma finalidade académica, há contributos de natureza diversa que não podem nem devem deixar de ser realçados. Por essa razão, desejo expressar os meus sinceros agradecimentos:

À Prof. Ana Maria Mendonça, à Prof. Ana Paula Pêgo e ao Dr. Hugo Oliveira pela ajuda, compreensão e disponibilidade.

À FEUP e ao INEB por toda a organização, e condições fornecidas para a realização do projecto.

Ao Programa Operacional Ciência e Inovação - 2010 pela bolsa.

A toda a gente que ao longo deste tempo de elaboração do projecto tiveram de suportar as minhas intensas conversas sobre o projecto e mesmo assim mostraram-se sempre disponíveis para me ajudar.

Aos meus amigos que foram perguntando pelo trabalho.

A toda a gente do INEB pelo excelente ambiente de trabalho e social criado.

Por último gostaria de agradecer também a todos aqueles de forma directa ou indirecta foram ajudando.

A todos, os meus sinceros agradecimentos.

Índice

1	Resumo	1
2	Introdução	3
2.1	Necessidade do projecto	4
2.2	Objectivos do projecto	5
2.3	Imagens tipo	5
2.4	Principais dificuldades	8
3	Detecção dos contornos das células	13
3.1	Introdução	14
3.2	Detecção do fundo	14
3.3	Subdivisão das imagens	17
3.4	Delimitação das células	19
3.5	Análise do método	20
4	Detecção das células	21
4.1	Introdução	22
4.2	Pré-processamento das imagens	22
4.3	Filtragem	23
4.4	Detecção dos centros das células	26
4.5	Eliminação de falsas detecções	28
4.6	Células não detectadas	29
4.7	Análise do método	30

5	Classificação das células	33
5.1	Introdução	34
5.2	Detecção das zonas azuis	34
5.3	Classificação das células	36
5.4	Análise do algoritmo	38
6	Programa	41
6.1	Introdução	42
6.2	Interface gráfica	42
7	Resultados e Conclusões	47
7.1	Resultados	48
7.1.1	Resultados nas imagens de desenvolvimento do algoritmo .	48
7.1.2	Resultados nas imagens de avaliação do algoritmo	49
7.2	Exemplos de resultados	51
7.3	Conclusões	60
	Referências	61

Lista de Figuras

2.1	Imagem de fluorescência	6
2.2	Imagem de campo claro	7
2.3	Célula em metafase	8
2.4	Células sobrepostas	9
2.5	Dificuldade de distinguir se no centro da imagem existe uma célula nova ou parte de outra.	9
2.6	Célula de difícil visualização	10
2.7	Imagem saturada	10
2.8	Interferência do substrato azul	11
2.9	Células sem gene marcadas com a azul por interferência das células vizinhas.	11
2.10	Células com gene marcadas com um azul muito ténue.	11
3.1	Imagem de uma célula	15
3.2	Célula após abertura com elemento estruturante <i>disk</i>	15
3.3	Imagem após a utilização do filtro TopHat	16
3.4	Célula após binarização	17
3.5	Imagem dos contornos obtidos com a detecção do fundo	17
3.6	Imagem dos contornos detectados com o algoritmo watershed	18
3.7	Imagem dos contornos detectados com a detecção do fundo e com o algoritmo watershed	19
3.8	Resultado final do cálculo dos contornos das células	19
4.1	Célula após abertura com estrutura <i>disk</i> com raio 10	23

4.2	Representação 3D do filtro A com dimensões 30x30	25
4.3	Representação 3D do filtro B com dimensões 30x30	25
4.4	Células após filtragem com o filtro A (para melhorar a visualização desta imagem aumentou-se a iluminação e o contraste)	26
4.5	Deteccção dos centros das células	27
4.6	Áreas das zonas detectadas	28
4.7	Células não detectadas usando o algoritmo de deteccção de células mas por estarem localizadas numa zona onde foi detectada com sucesso o seu contorno.	29
4.8	Resultado da deteccção das células	31
5.1	Binarização da imagem de campo claro	35
5.2	Binarização da imagem de campo claro pós-processada	35
5.3	Separação das zonas das células	36
5.4	Resultado da classificação das células	39
5.5	Resultado da classificação das células numa zona com azul muito ténue	39
6.1	Interface gráfica	42
7.1	Resultado da imagem de fluorescência, (colheita nº5 usando parâmetros por defeito)	52
7.2	Resultado da imagem de campo claro, (colheita nº5 usando parâmetros por defeito)	53
7.3	Resultado da imagem de fluorescência, (colheita nº4 usando parâmetros por defeito)	54
7.4	Resultado da imagem de campo claro, (colheita nº4 usando parâmetros por defeito)	55
7.5	Resultado da imagem de fluorescência, (colheita nº6 usando como parâmetros para a contagem do número de células que exprimem o gene: Pontos interiores=94% e Pontos exteriores=99%)	56

7.6	Resultado da imagem de campo claro, (colheita nº6 usando como parâmetros para a contagem do número de células que exprimem o gene: Pontos interiores=94% e Pontos exteriores=99%)	57
7.7	Resultado da imagem de fluorescência, (colheita nº11 usando parâmetros por defeito)	58
7.8	Resultado da imagem de campo claro, (colheita nº11 usando parâmetros por defeito)	59

Lista de Tabelas

4.1	Exemplo de um filtro A com dimensões 5x5	24
7.1	Contagem do número total de células nas imagens de desenvolvi- mento do algoritmo	48
7.2	Contagem do número de células que exprimem o gene nas imagens de desenvolvimento do algoritmo	49
7.3	Contagem do número total de células nas imagens de avaliação do algoritmo	49
7.4	Contagem do número de células que exprimem o gene nas imagens de avaliação do algoritmo	50

Capítulo 1

Resumo

Com o objectivo principal de desenvolvimento de sistemas não-virais seguros e eficazes de terapêutica dos genes para promover a neuro regeneração no sistema nervoso periférico, estão a ser estudados os polímeros baseados em quitosano (polissacarídeo derivado da quitina).

Para este estudo é necessário calcular a eficiência de transfecção que pode ser definida como a percentagem de células que exprimem o gene exógeno em estudo, que pode ser estimada através da contagem das células a partir de duas imagens da mesma colheita obtida através de dois métodos. O número total de células na cultura é estimado com base nas imagens de fluorescência onde se usou um fluorocromo que marca o DNA de cor vermelha. Na imagem de campo claro pretende-se identificar as células marcadas com coloração azul, correspondendo esta cor a um substrato metabolizado apenas pelas células com o gene beta-galactosidase.

Este projecto surge devido ao facto da contagem das células, para além de cansativa e demorada, dar origem a grande dispersão nas contagens devido a situações de dúvida. Assim, com este projecto desenvolveu-se um algoritmo capaz de produzir de forma automática dois valores com o número total de células e com o número de células onde o gene se exprime, bem como criar duas imagens

onde são identificados os contornos e os centros das células detectadas; uma vez que é facilmente perceptível a localização das zonas da imagem e das células onde o algoritmo teve maior dificuldade de contagem, é fácil corrigir os valores resultantes do algoritmo, ou seja, o algoritmo permite com um trabalho mínimo para o utilizador aliar as qualidades da facilidade da contagem automática às qualidades da experiência do utilizador na contagem manual, como pode ser visto nos exemplos das imagens apresentadas nos capítulos seguintes.

Capítulo 2

Introdução

Sumário: *Neste capítulo será referido o ambiente onde se insere este projecto assim como as suas principais dificuldades.*

2.1 Necessidade do projecto

Com o objectivo principal de desenvolvimento de sistemas não-virais seguros e eficazes de terapêutica dos genes para promover a neuro regeneração no sistema nervoso periférico, estão a ser estudados os polímeros baseados em quitosano (polissacarídeo derivado da quitina).

Actualmente a eficiência da transferência de genes dos vectores desenvolvidos é avaliada *in vitro* usando as células 293T (células humanas embrionárias epiteliais do rim) como modelo e o gene beta-galactosidase (uma enzima que cliva a lactose em glicose e galactose).

A eficiência de transfecção pode ser definida como a percentagem de células que exprimem o gene exógeno em estudo, tomando como referência o número total de células que foram sujeitas ao processo de transfecção, e pode ser estimada através da contagem das células a partir de duas imagens da mesma colheita obtida através de dois métodos. O número total de células na cultura é estimado com base nas imagens de fluorescência onde se usou um fluorocromo que marca o DNA de cor vermelha, tal como se pode observar na Figura 2.1. Na imagem de campo claro, visível na imagem da Figura 2.2, pretende-se identificar as células marcadas com coloração azul, correspondendo esta cor a um substrato metabolizado apenas pelas células com o gene beta-galactosidase.

Actualmente estão a ser usadas as seguintes ferramentas de apoio à contagem das células:

- Axiovision LE Rel. 4.1 (da ZEISS) - para adquirir a imagem e para tratá-la nos termos do contraste, do brilho, etc.
- Imagem J - para executar o desconvolução das imagens (para determinar limites das células)

sendo ambos programas de uso geral que não permitem uma sistematização e automação do processo de contagem das células, o que obriga a uma contagem

apoiada, demorada e cansativa.

2.2 Objectivos do projecto

O objectivo deste projecto é o desenvolvimento de procedimentos de análise de imagem que possibilitem a avaliação da eficiência da transferência de genes para células eucarióticas (transfecção), através da contagem do número de células em imagens digitais, obtidas usando microscopia de fluorescência e microscopia de transmissão, de culturas celulares *in vitro*.

Como foi dito anteriormente, a eficiência de transfecção pode ser definida como a percentagem de células que exprimem o gene exógeno em estudo, tomando como referência o número total de células que foram sujeitas ao processo de transfecção. A análise das imagens de microscopia de fluorescência deverá permitir a detecção dos limites das células e respectiva contagem automática. O número de células que exprimem o gene em estudo resultará da detecção das células positivas que, nas imagens coloridas obtidas com microscopia de transmissão, se manifestam com uma coloração distinta.

2.3 Imagens tipo

Seguidamente são apresentadas duas imagens tipo, uma imagem de fluorescência onde é possível contar o número total de células, e uma imagem de campo claro onde é possível ver as zonas onde existem células que exprimem o gene.

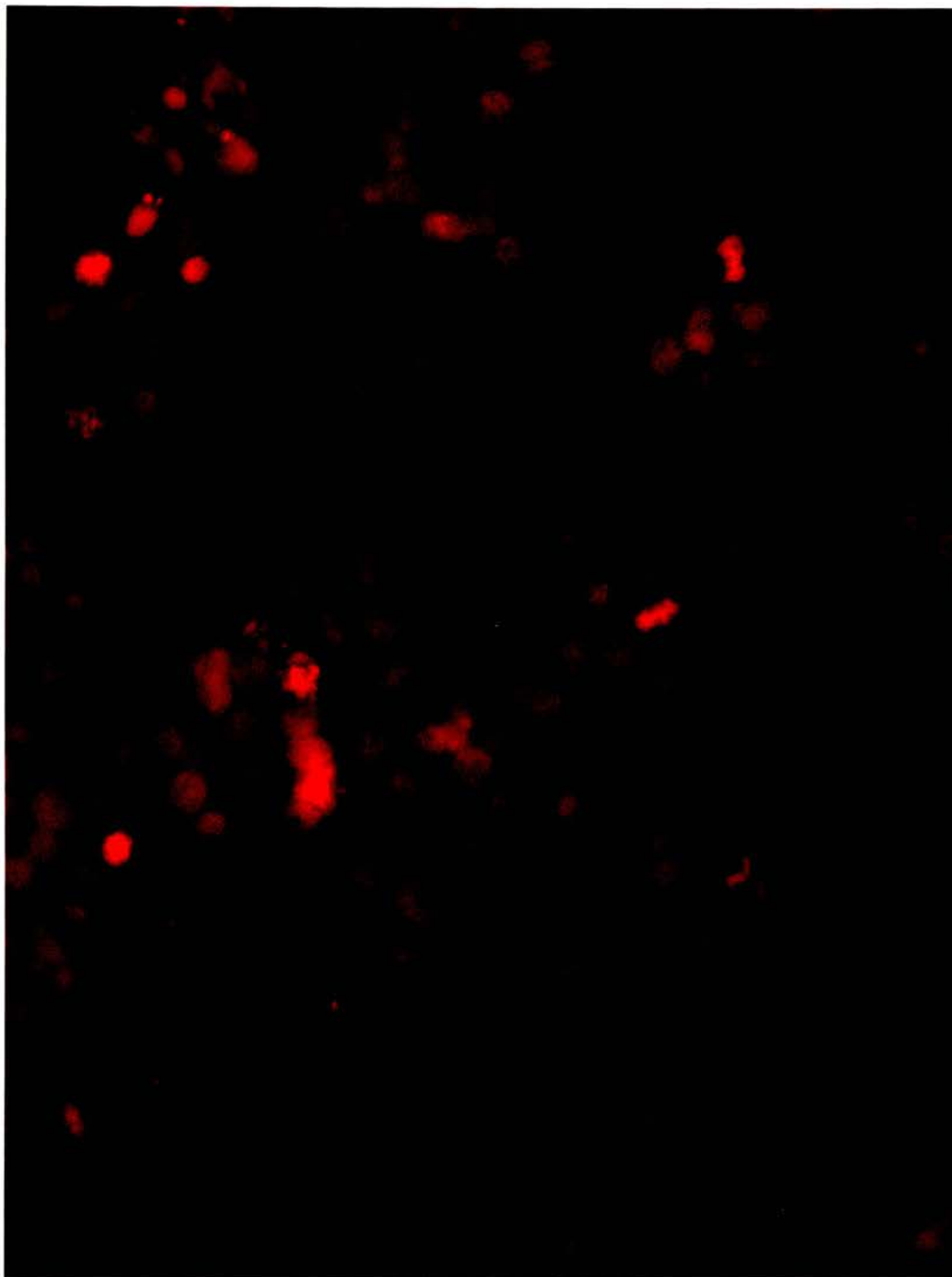


Figura 2.1 — Imagem de fluorescência

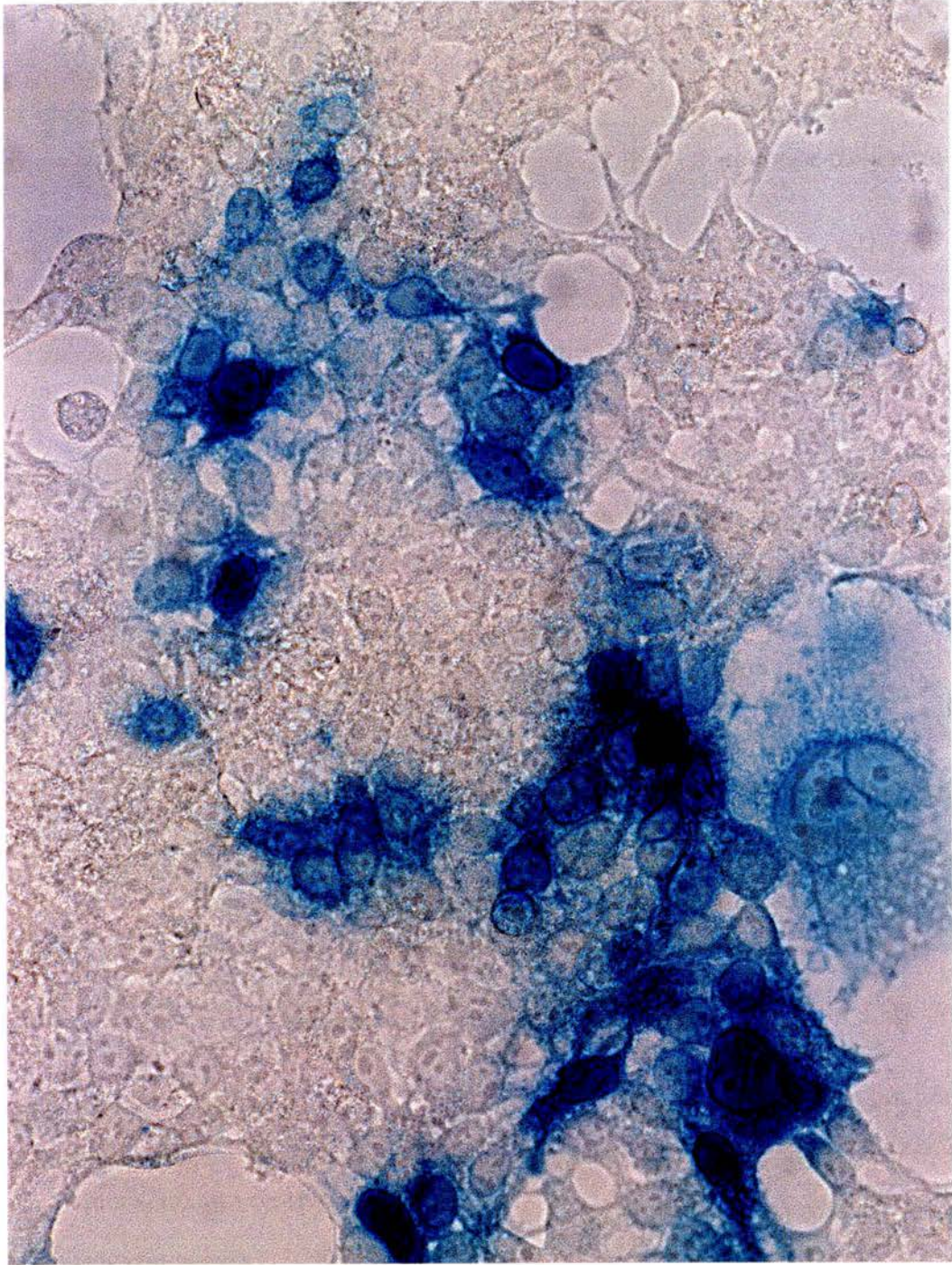


Figura 2.2 — Imagem de campo claro

2.4 Principais dificuldades

Existem vários tipos de dificuldades na contagem das células, sendo as principais seguidamente descritas:

1. Uma das maiores dificuldades deste projecto é a grande variação na qualidade das imagens devido a não existir um sistema de aquisição montado exclusivamente para este projecto, o que provoca com que as imagens tenham aspectos muito diversos.
2. Em alguns casos verificam-se células em metáfase (material genético a ser dividido, Figura 2.3), devendo ser contabilizado apenas um núcleo, já que ainda se trata de apenas uma só célula.

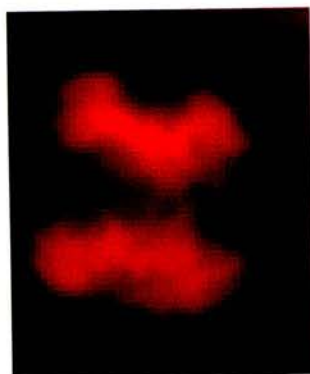


Figura 2.3 — Célula em metáfase

3. Nas imagens de fluorescência podem existir zonas com muitas células juntas e sobrepostas, como o caso da imagem da Figura 2.4, o que dá origem a muitos erros de contagem.

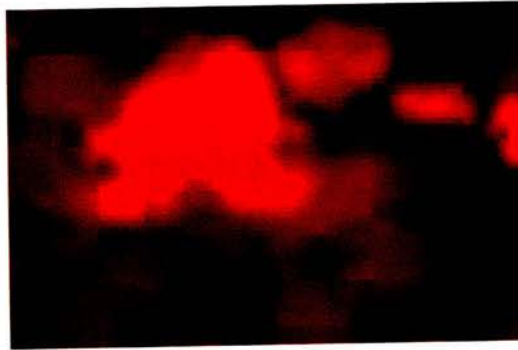


Figura 2.4 — Células sobrepostas

4. Existem algumas zonas na imagem em que não é fácil perceber o número correcto de células existentes; por exemplo na Figura 2.5 é difícil perceber se existem três ou quatro células .



Figura 2.5 — Dificuldade de distinguir se no centro da imagem existe uma célula nova ou parte de outra.

5. Algumas células são de difícil visualização devido à intensidade de luz por ela reflectida ser muito baixa; por exemplo na imagem da Figura 2.6 existe uma célula no centro da imagem.



Figura 2.6 — Célula de difícil visualização

6. Em alguns casos a imagem adquirida apresenta algumas zonas extensas onde a iluminação é demasiado elevada e satura a imagem, tal como se mostra na Figura 2.7.

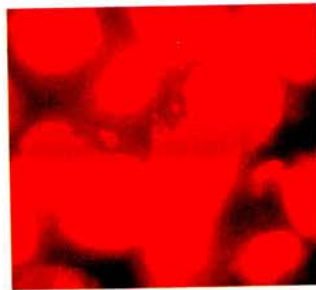


Figura 2.7 — Imagem saturada

7. Algumas células marcadas com o substrato azul (de campo claro) podem apresentar uma fluorescência menos intensa ou mesmo ausente, existindo interferência; por exemplo na imagem de fluorescência correspondente à zona representada na imagem seguinte apenas aparece um espaço vazio, negro, como pode ser visto na imagem da Figura 2.8.

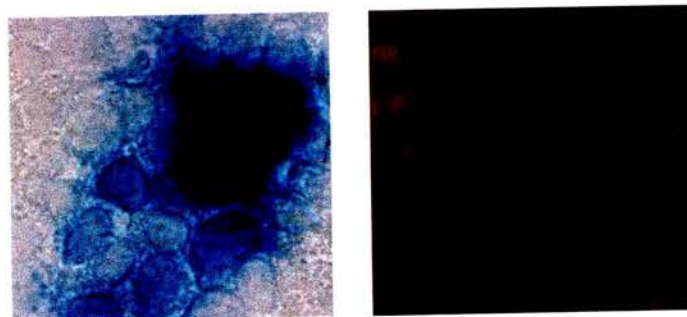


Figura 2.8 — Interferência do substrato azul

8. Algumas células com gene marcadas com corante azul produzem interferência nas células vizinhas fazendo com que elas também apresentem uma tonalidade azul mesmo não exprimindo o gene, Figura 2.9.

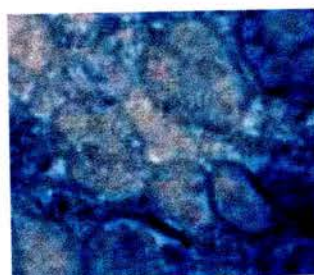


Figura 2.9 — Células sem gene marcadas com a azul por interferência das células vizinhas.

9. Algumas células com o gene são marcadas com um azul muito ténue o que provoca problemas em conseguir distinguir se têm o gene ou se a sua tonalidade provém de alguma interferência, figura 2.10.



Figura 2.10 — Células com gene marcadas com um azul muito ténue.

Capítulo 3

Detecção dos contornos das células

Sumário: *Neste capítulo será descrito o processo de detecção dos contornos das células, que se baseia num algoritmo de detecção do fundo da imagem e da etiquetagem das zonas ocupadas por cada célula.*

3.1 Introdução

A detecção dos contornos das células teria apenas o objectivo de facilitar a identificação das células. No entanto devido às imagens das células serem um pouco esbatidas foi impossível ter uma detecção dos contornos eficaz através de um algoritmo habitual de detecção de orlas como por exemplo o algoritmo de detecção de orlas de *Canny*; portanto foi necessário criar um algoritmo alternativo de detecção dos contornos, que se baseia na conjugação da identificação do fundo da imagem e da divisão da imagem em zonas através de um algoritmo de *watershed*; cada zona idealmente é ocupada apenas por uma célula.

Devido à capacidade deste algoritmo catalogar todas as zonas da imagem, dá-nos a possibilidade de criar métodos de detecção das células com um filtro para zonas onde a detecção das células é relativamente fácil, e a utilização de um filtro diferente para zonas onde a detecção é mais difícil, como será melhor explicado no Capítulo 4.

3.2 Detecção do fundo

O primeiro passo para a detecção do fundo é a realização de uma operação morfológica de abertura das imagens com um elemento estruturante do tipo *disk* com raio=10; este valor para o raio foi determinado experimentalmente. A abertura da imagens é uma operação morfológica muito utilizada para eliminar segmentos pequenos na imagem, ou para separar objectos que estão ligados por elementos pequenos, e consiste em fazer uma erosão da imagem seguida de uma dilatação com o mesmo elemento estruturante que é dado como parâmetro. No caso deste trabalho era um requisito que o elemento estruturante fosse bastante menor do que as células para não as deformar, mas que simultaneamente fosse suficientemente grande para separar as células; pode-se ver nas Figuras 3.1 e 3.2 um exemplo da utilização da abertura das imagens nas células.

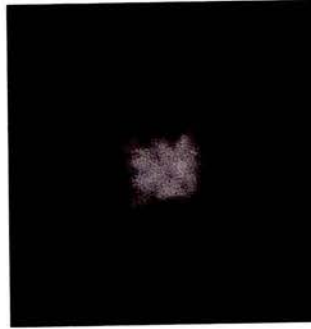


Figura 3.1 — Imagem de uma célula



Figura 3.2 — Célula após abertura com elemento estruturante *disk*

Seguidamente pretende-se separar as células do fundo tendo sido usada uma operação morfológica top-hat com um elemento estruturante do tipo *disk* com raio=40; esta é uma operação morfológica em que o objectivo é, numa escala de cinzentos, separar os objectos do fundo usando a transformada de top-hat [1], dependendo de se tratar de objectos claros em fundo escuro, ou objectos escuros em fundo claro, a transformada de top-hat é definida como:

Objectos claros: $Tophat(A, B) = A - (A \circ B) = A - \max_B(\min_B(A))$

Objectos escuros: $Tophat(A, B) = A - (A \bullet B) = A - \min_B(\max_B(A))$

onde o elemento estruturante **B** é definido de forma a ser maior do que os elementos **A** que se quer detectar, e se possível deve ter uma forma convexa.

Sabendo que o raio das células costuma variar entre os 20 e os 30 pixels e que o raio do elemento estruturante top-hat deveria ser consideravelmente maior do que o raio das células, optimizou-se o raio do elemento estruturante experimentalmente; assim conseguiu-se ter uma imagem em que o fundo é homogéneo e com valores de intensidade relativamente baixos comparados com os valores de intensidade nas células, como pode ser visto na Figura 3.3.

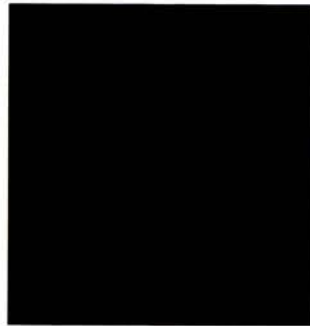


Figura 3.3 — Imagem após a utilização do filtro TopHat

No entanto, quando duas células estão próximas o valor de intensidade no fundo entre as células ainda é superior ao valor de algumas células, isto dificulta a binarização da imagem, pois se usássemos apenas um threshold, dependendo de termos escolhido um limiar alto ou baixo iríamos classificar algumas células como sendo parte do fundo, ou classificar zonas do fundo como sendo uma zona onde existe uma célula; para resolver este problema, e de forma a fazer com que o algoritmo fosse mais flexível e dinâmico, em vez de um limiar usou-se a transformada de mínimo estendido; a transformada de mínimo estendido produz uma imagem binária, que é calculada suprimindo todos os mínimos regionais cuja profundidade, relativamente aos pixels vizinhos, seja inferior a h (h é um escalar), representando com valores 1 os pixels que são os mínimos regionais dessa operação e com zeros todos os outros pixels; neste caso colocou-se o nível de diferença h igual a 6 e conectividade 8; assim consegue-se separar as células do fundo de forma dinâmica sendo possível detectar todas as células, mesmo aquelas com níveis de intensidade muito baixo e considerar como fundo os pontos que apesar de terem

níveis de intensidade elevados estão próximos de zonas com níveis de intensidade ainda mais elevados, como se pode ver no exemplo da Figura 3.4 o resultado da binarização da imagem através do processo anteriormente descrito.

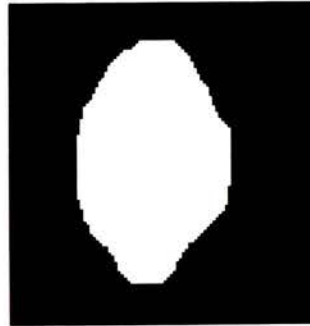


Figura 3.4 — Célula após binarização

A partir destas imagens binarizadas pode-se desenhar facilmente as orlas das zonas detectadas, Figura 3.5.

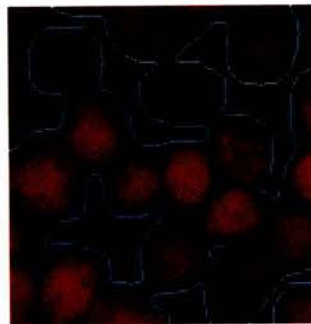


Figura 3.5 — Imagem dos contornos obtidos com a detecção do fundo

3.3 Subdivisão das imagens

Com os passos anteriores conseguiu-se separar eficazmente as células do fundo, no entanto existem células que estão juntas a outra células ou mesmo sobrepostas e os passos anteriores não conseguem estimar a posição onde se dividem as células de forma a ser possível desenhar os respectivos contornos das células, o

que obrigou a criar uma sequência de passos que complementassem as insuficiências do método anterior. Essa sequência foi construída baseada no algoritmo de *watershed*; este algoritmo produz uma matriz de etiquetas que identificam cada região da imagem, baseando-se num princípio que simula uma vertente de água para várias poças; começa-se a encher todas as poças ao mesmo tempo e quando a água das poças começa a passar para a poça vizinha, então está encontrado o limite dessas regiões.

Para se aplicar este algoritmo nas imagens de fluorescência foi necessário, após transformar a imagem em tons de cinzento, inverter a intensidade da imagem para que as zonas das células fossem as poças que o algoritmo *watershed* detecta.

Assim, após a aplicação deste algoritmo, ficamos com uma matriz que delimita a zona ocupada por cada célula, mas, também este algoritmo apresenta algumas falhas para o objectivo de desenhar as orlas das células, uma vez que coloca a divisão das zonas das células sensivelmente equidistantes de todas as células, o que é excelente quando as células estão muito próximas ou mesmo juntas, mas quando estas estão muito afastadas os limites das zonas serão muito afastados dos limites reais que se pretende detectar, Figura 3.6.

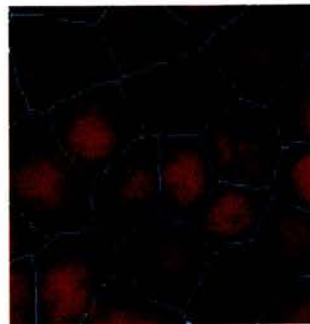


Figura 3.6 — Imagem dos contornos detectados com o algoritmo watershed

3.4 Delimitação das células

Para se conseguir delimitar as células e desenhar as orlas correctamente foi necessário juntar o melhor dos dois processos de detecção das orlas anteriormente descritos. Tecnicamente o que se fez foi fazer a junção através de um *and* lógico dos resultados da detecção do fundo e dos resultados do algoritmo *watershed*, conseguindo-se assim através do algoritmo watershed delimitar as células quando estão muito próximas umas das outras e através da detecção do fundo eliminar as áreas entre as células que não pertencem a nenhuma célula mas sim ao fundo, como pode ser visto nas Figuras 3.7 e 3.8.

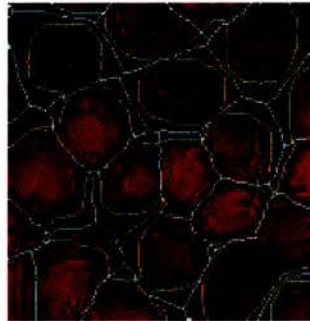


Figura 3.7 — Imagem dos contornos detectados com a detecção do fundo e com o algoritmo watershed

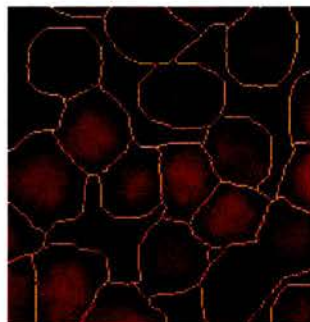


Figura 3.8 — Resultado final do cálculo dos contornos das células

3.5 Análise do método

Este método de cálculo dos contornos das células mostrou ser bastante eficaz e dinâmico fornecendo bons resultados em imagens de qualidade muito distinta, funcionando tanto em imagens com uma iluminação muito fraca onde este método é capaz de detectar e delimitar células quase imperceptíveis aos humanos, como em imagens com iluminação bastante elevada, apenas dando resultados piores em zonas onde existem células muito juntas e sobrepostas.

Este método tem a vantagem de etiquetar cada zona o que irá permitir calcular a área de cada zona e com base nessa área estimar se essa zona é de difícil detecção das células, neste caso é possível usar um filtro mais eficaz para essas zonas, como será melhor explicado no capítulo 4.

Através da etiquetagem das zonas também seria possível estimar a quantidade de células na imagem, no entanto não foi este o método que se usou para esse fim devido a ter sido possível construir um algoritmo mais eficaz de detecção e contagem das células.

4.1 Introdução

A contagem das células é o objectivo mais importante deste projecto por ser a tarefa mais difícil e trabalhosa para um utilizador executar devido à grande quantidade de células que cada colheita contém, e ao facto de existirem zonas na colheita onde é muito difícil efectuar a contagem devido aos vários problemas já descritos nos capítulos anteriores.

Para além de se pretender obter uma boa contagem da quantidade de células existentes na imagem, é também objectivo conhecer a localização de onde a célula foi detectada de forma a permitir ao utilizador detectar possíveis erros na contagem, e ajudar nos casos que ele tem dúvidas acerca do número real de células existentes.

Devido à grande variação da iluminação ao longo da imagem foi necessário que o algoritmo de contagem das células se baseasse em operações dinâmicas que se adaptassem ao tipo de iluminação da zona onde a célula está inserida; isto foi conseguido através de uma operação de abertura das imagens e da utilização de filtros com o objectivo de obrigar a que os pixels dos centros das células tivessem valores de iluminação superiores aos valores dos pixels vizinhos, conseguindo-se assim detectar os centros das células através do cálculo os máximos locais da imagem.

4.2 Pré-processamento das imagens

Tal como no algoritmo de detecção do fundo, também aqui o primeiro passo é a realização de uma operação morfológica de abertura das imagens com um elemento estruturante do tipo *disk* com raio=10, isto permite que seja mais fácil detectar cada célula como uma só pois cria espaços entre células e torna-se mais fácil distinguir as células. Tal como foi dito na introdução, pretende-se que os centros das células tenham valores de intensidade superiores aos extremos da

célula; esta função de abertura também se mostra útil nesse objectivo dado que a realização da dilatação após uma erosão faz com que as zonas mais distantes dos centros das células sofram influência dos pixels pertencentes ao fundo e daí resulte que o valor de intensidade final nesses pixels sejam mais baixos do que os que se localizam no centro das células e que não sofrem esta influência, como pode ser visto na Figura 4.1.

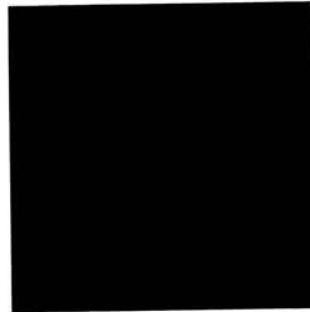


Figura 4.1 — Célula após abertura com estrutura *disk* com raio 10

4.3 Filtragem

Seguidamente, faz-se a filtragem da imagem resultante por dois filtros especialmente construídos para este efeito, e que são complementares. Visto ter sido impossível criar um filtro que conseguisse aliar um baixo número de falsas detecções e fosse capaz de detectar grande parte das células nas zonas onde se encontram muito juntas, criaram-se estes dois filtros: o primeiro filtro *Filtro A*, Figura 4.2, produz resultados com baixo número de falsas detecções, mas quando existem células muito próximas é incapaz de detectar grande parte das células; o *Filtro B*, Figura 4.3, apesar de produzir um maior número de falsas detecções é capaz de detectar células mesmo em zonas muito difíceis como são as zonas onde existem muitas células juntas e sobrepostas. Com estas filtrações pretende-se que o filtros acentuem ainda mais as diferenças de intensidade dos pixels centrais das células para os pixels exteriores.

O *Filtro A* foi construído de forma a ter razoavelmente as mesmas dimensões das células, e a sua forma baseia-se num hexágono obtido pelo mesmo elemento estruturante da dilatação, ou seja, um disco de dimensões 30x30, onde o peso de cada ponto é igual a:

distância à orla mais próxima do hexágono em conectividade 8^6 , nos pontos interiores ao hexágono e zero nos pontos exteriores.

Posteriormente normalizou-se a amplitude dos pontos para que o peso total do filtro fosse 1, o que não afecta o resultado final da detecção mas facilitou visionamento do resultado da filtragem durante o desenvolvimento do filtro.

0	0	0.0098976	0	0
0	0.0098976	0.061057	0.0098976	0
0.0098976	0.061057	0.67659	0.061057	0.0098976
0	0.0098976	0.061057	0.0098976	0
0	0	0.0098976	0	0

Tabela 4.1 — Exemplo de um filtro A com dimensões 5x5

A potência 6 foi calculada de forma experimental, tendo sido verificado que se aumentasse mos esta potência iríamos conseguir detectar células que com potências inferiores não eram detectadas mas, no entanto, também iriam aumentar as falsas detecções, e como foi o critério durante toda a elaboração do algoritmo preferiu-se sempre manter um baixo número de falsas detecções.

Com o *Filtro B* pretende-se aumentar a capacidade de detectar células nas zonas mais difíceis, mesmo que para isso seja necessário aumentar um pouco o número de falsas detecções. Este filtro foi construído com base no filtro A, tendo sido criada uma janela com o mesmo formato que o filtro A, mas com dimensões 70x70 e com potência $\frac{21}{5}$; posteriormente colocou-se um valor de intensidade negativo na orla do hexágono, sendo este filtro capaz de detectar as células nas

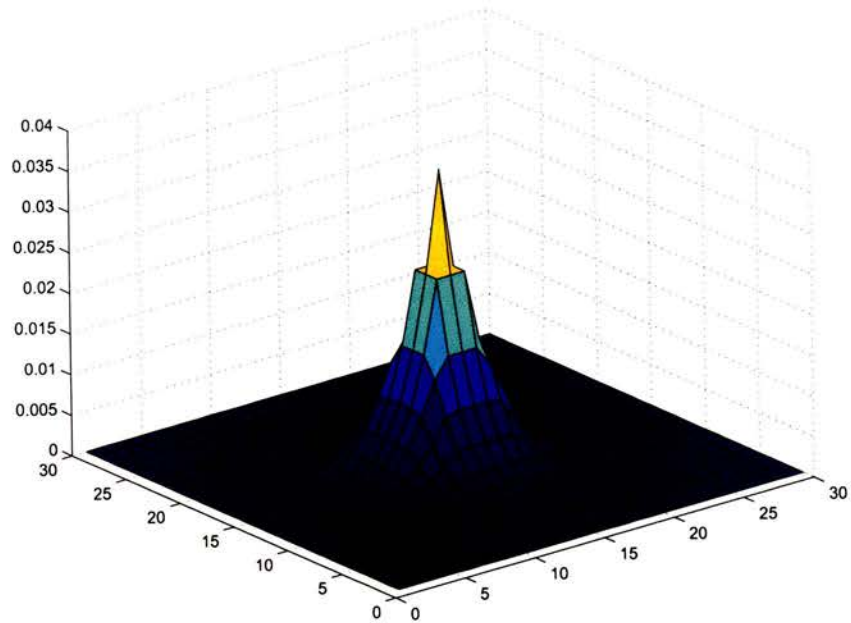


Figura 4.2 — Representação 3D do filtro A com dimensões 30x30

zonas mais difíceis. Porém, como o número de falsas detecções era demasiado elevado, para resolver este problema foi necessário suavizar este filtro, o que foi conseguido através de uma filtragem com o filtro A.

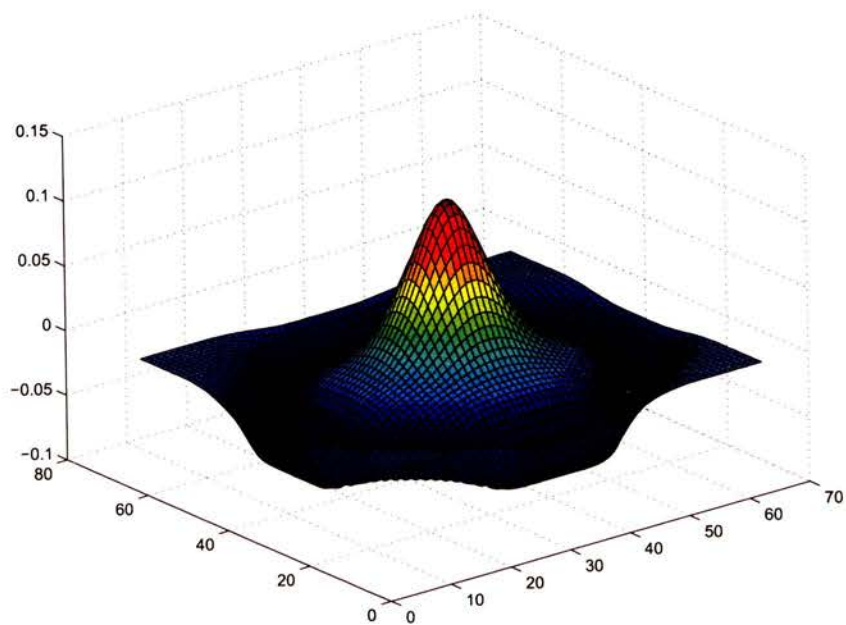


Figura 4.3 — Representação 3D do filtro B com dimensões 30x30

O filtro B tem uma boa capacidade de detecção de células nas zonas mais difíceis mas o número de falsas detecções é mais elevado do que o resultante do filtro A.

4.4 Detecção dos centros das células

A detecção dos centros das células é feita através da aplicação da transformada de máximo estendido às imagens depois de filtradas; esta transformada produz uma imagem binária, que é calculada suprimindo todos os máximos regionais cuja altura seja inferior a h (h é um escalar), e representa com valores 1 os pixels que são máximos regionais e com zeros todos os outros pixels. Neste caso h é zero e foi usada conectividade 8.

O resultado desta operação é uma imagem binária com fundo zero e com alguns pontos com nível 1 localizados nos centros das células.

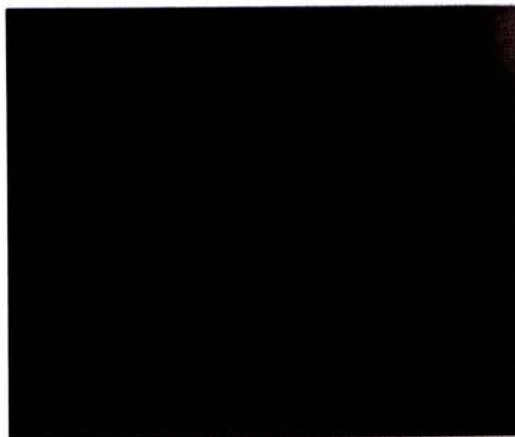


Figura 4.4 — Células após filtragem com o filtro A (para melhorar a visualização desta imagem aumentou-se a iluminação e o contraste)

Como já foi dito anteriormente os dois filtros construídos produzem imagens em que a detecção das células é melhor numas zonas do que noutras; para que a detecção final tenha uma boa qualidade construiu-se um método que dependendo da zona da imagem, ter muitas células juntas ou não, utiliza o filtro

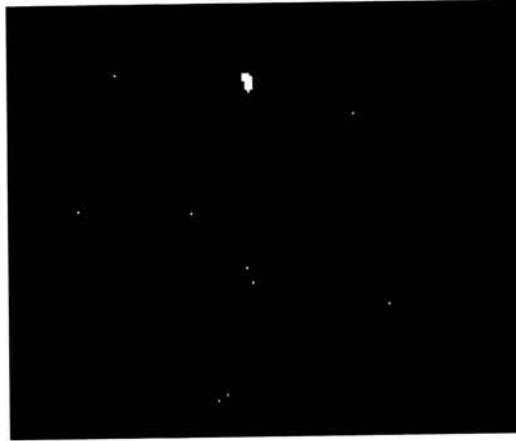


Figura 4.5 — Detecção dos centros das células

que melhor responde a essa situação.

O método baseia-se no estudo das dimensões das zonas delimitadas quando da detecção dos contornos e na classificação dessas zonas como sendo uma zona de difícil detecção, resultante da verificação de que nas zonas onde existem muitas células a detecção das orlas das células também falha produzindo orlas que englobavam mais do que uma célula, logo a área dessa zona vai ser superior às zonas onde a detecção dos contornos foi bem feita e a orla apenas delimita uma célula.

Após verificação em imagens de várias colheitas de células chegou-se à conclusão que seria razoável considerar que as zonas de maior dificuldade são as zonas com área superior ao dobro da área mais frequente; como pode ser visto na Figura 4.6 o pico aparece em aproximadamente 1200 pixels, então neste caso seriam consideradas zonas de maior dificuldade as zonas com área superior a 2400.

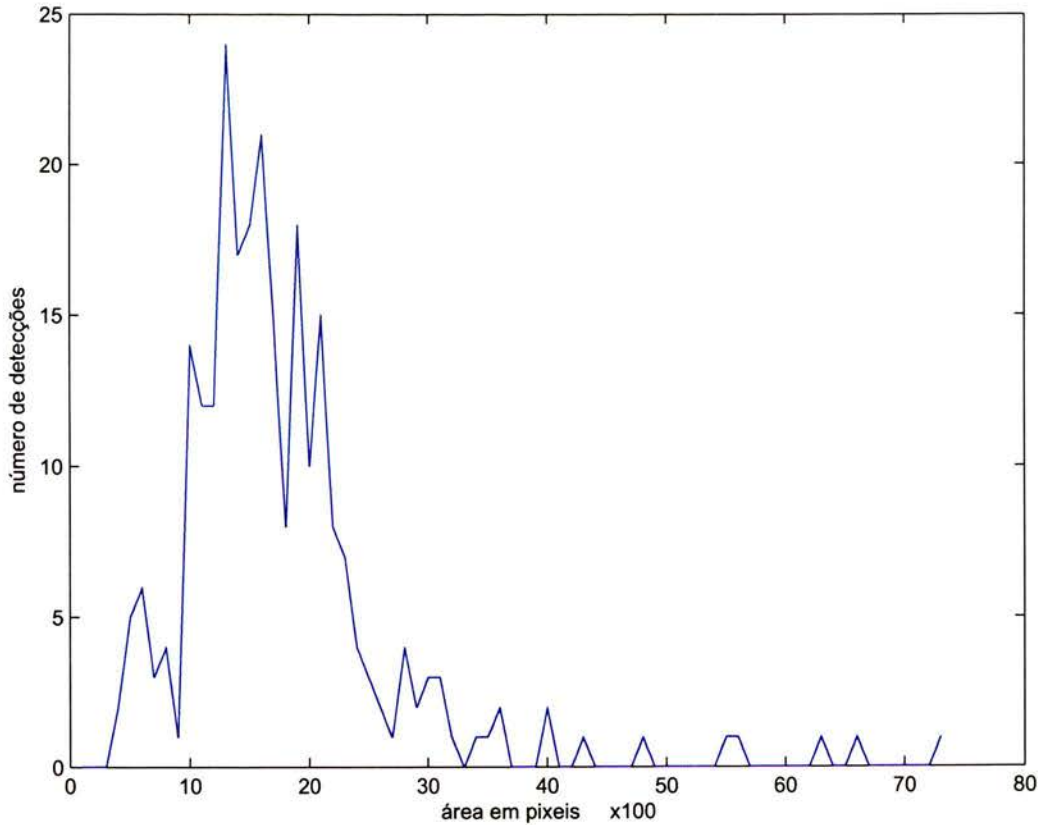


Figura 4.6 — Áreas das zonas detectadas

4.5 Eliminação de falsas detecções

Como pode ser visto na imagem da Figura 4.4 a detecção usando a transformada de máximo estendido produz algumas falsas detecções, que normalmente pertencem a um dos seguintes tipos:

1. Detecção de células onde não existe nenhuma célula.
2. Múltiplas detecções da mesma célula.

O primeiro tipo de falsa detecção é muito fácil de corrigir, pois como anteriormente já detectamos as zonas de fundo, basta eliminar as detecções que se encontram nessas zonas.

O segundo tipo de erro é mais difícil porque as várias detecções da mesma célula podem ocorrer afastadas umas das outras; porém através de uma operação de dilatação consegue-se unir as várias detecções da mesma célula desde que estejam a uma distância inferior ao raio do elemento estruturante de dilatação que no caso foi um círculo de raio=8.

4.6 Células não detectadas

Por outro lado também existem células que não foram detectadas por este algoritmo mas que ainda é possível detectar, como são as células que no processo de cálculo dos contornos foram bem detectadas e se consegue criar uma boa delimitação, mas onde o processo de detecção não determina qualquer célula nessa zona; neste caso podemos estimar que existe pelo menos uma célula nessa zona, e o que foi feito para estimar a quantidade de células existentes, e as suas respectivas localizações foi calcular a erosão das imagens dessas zonas, até restarem apenas pontos isolados, e para eliminar os pontos resultantes dessa operação que estejam duplicados fez-se uma dilatação desses pontos, como se pode ver na imagem da Figura 4.7 o resultado final dessas operações, onde os pontos amarelos foram detectados através dos processos de filtragem e os pontos azuis resultaram deste último tipo de detecção .

O outro tipo de células que apesar de não terem sido detectadas se pode

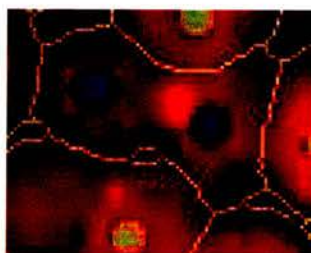


Figura 4.7 — Células não detectadas usando o algoritmo de detecção de células mas por estarem localizadas numa zona onde foi detectada com sucesso o seu contorno.

estimar que elas existem são as células que nas imagens de campo claro estão nas zonas com um azul escuro, como pode ser visto na imagem da Figura 2.8, é frequente que nestas zonas, nas imagens de fluorescência não seja possível ver nenhuma célula, mas como estas só ocorrem em zonas onde existem células com o gene, se detectarmos uma zona deste tipo onde não tenha sido detectada nenhuma célula podemos estimar que existe aí pelo menos uma célula e colocar no centro dessa zona um ponto indicando a presença de uma célula (este processo será melhor explicado no capítulo 5).

4.7 Análise do método

Com este algoritmo consegue-se fazer uma detecção bastante eficaz das células, com um número bastante baixo de falsas detecções, e onde normalmente apenas algumas das células que estão muito próximas umas das outras ou sobrepostas não são detectadas. Pode-se ver na imagem 4.8 parte de uma detecção de células, onde os centros das células são marcados com cores distintas consoante a forma como foram detectadas.

A convenção de cores usada para marcar os centros das células é a seguinte indicada:

1. *Verde - Células detectadas com o filtro A.*
2. *Azul claro - Células detectadas com o filtro B.*
3. *Laranja - Célula não detectada usando o algoritmo de detecção de células, mas cuja localização foi estimada por estar localizada numa zona de azul muito escuro nas imagens de campo claro.*
4. *Azul escuro - Célula não detectada usando o algoritmo de detecção de células, mas cuja localização foi estimada por estar localizada numa zona onde foi detectada com sucesso o seu contorno.*

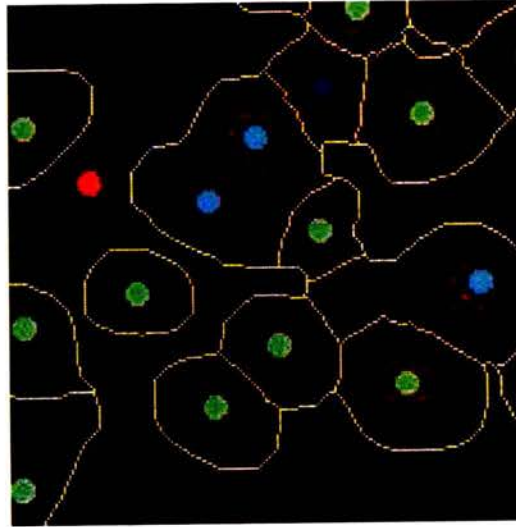


Figura 4.8 — Resultado da detecção das células

Capítulo 5

Classificação das células

Sumário: *Neste capítulo são apresentados os algoritmos de identificação e classificação das células com genes.*

5.1 Introdução

Para identificar as células com o gene foi necessário usar técnicas de análise de imagem colorida para detectar as zonas da imagem onde existe o gene, pois estas são marcadas com uma tonalidade azul nas imagens de campo claro, e essa coloração varia bastante de intensidade e de saturação. A detecção dessas zonas azuis foi feita em dois passos, um para detectar zonas azuis mais claras e mais frequentes que não perturbam muito a imagem de fluorescência, e outro passo para detectar as zonas de azul muito escuro que normalmente não permitem visualizar nenhuma célula nessas zonas nas imagens de fluorescência.

Foi também necessário criar um algoritmo capaz de superar o facto de o corante que marca as células com gene se esbater e colorir células que não têm o gene, e de existirem células em que o corante as marca com uma cor muito tênue.

5.2 Detecção das zonas azuis

Para detectar as zonas com tonalidade azul de forma eficaz e independente da intensidade e da saturação, foi necessário converter as imagens para o formato de representação de imagens coloridas HSV (tonalidade-saturação-intensidade). Este formato representa cada pixel através de três elementos tonalidade, saturação e intensidade todos representados de 0 a 1, tornando possível escolher uma tonalidade sem termos de nos preocupar com a iluminação e a saturação. Deste modo fez-se a binarização da imagem colocando a 1 todos os pixels que tivessem uma tonalidade azul com valores entre 0.47 a 0.57. Devido que toda a imagem tem alguns pixels com tonalidade azul, este processo resulta numa imagem tipo sal e pimenta, Figura 5.1, mas com maior frequência de pontos brancos nas zonas que se queriam detectar; portanto, é necessário eliminar os pontos brancos de todas as outras zonas e os pontos pretos das zonas de interesse.

Para se fazer uma limpeza da imagem e detectar apenas as zonas com

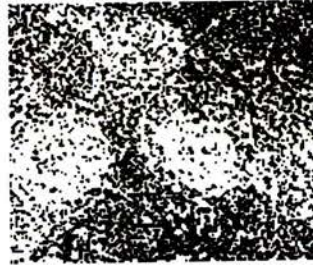


Figura 5.1 — Binarização da imagem de campo claro

níveis e quantidades de azul significativos e chegar a uma imagem do tipo da figura 5.2, seguiram-se os seguintes passos:

1. Abertura da imagem com elemento estruturante disk de raio=3, para eliminar pontos brancos que estejam isolados.
2. Fechamento da imagem com elemento estruturante disk de raio=5 para eliminar espaços entre os pontos brancos.
3. Preenchimentos dos buracos negros totalmente cercados por pontos brancos.
4. Abertura da imagem com elemento estruturante disk de raio=10 para separar as várias zonas detectadas.

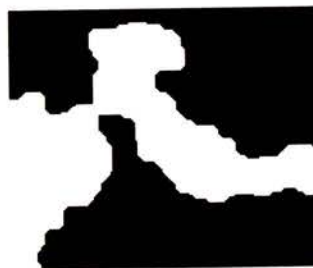


Figura 5.2 — Binarização da imagem de campo claro pós-processada

Para a detecção das zonas com um azul escuro mais intenso foi usado um método semelhante, mas por estas zonas algumas vezes não serem verdadeiramente azuis, mas sim pretas, foi necessário usar também o conhecimento da saturação da imagem; assim, foram consideradas como sendo dessas zonas as

áreas que tivessem valores de tonalidade entre 0.534 a 1 e valores de saturação entre 0.70 a 1. Posteriormente a esta binarização, foi também necessário fazer uma limpeza da imagem seguindo os mesmos passos descritos anteriormente, pode-se ver na figura 5.3 a delimitação das zonas utilizando os métodos anteriormente descritos, onde foi usada a seguinte convenção de cores:

1. *Azul claro - Zona das células com gene.*
2. *Azul escuro - Zona das células com azul muito forte.*
3. *Amarelo - Contornos das células detectados nas imagens de fluorescência.*
4. *Violeta - Zona de influencia do corante.*

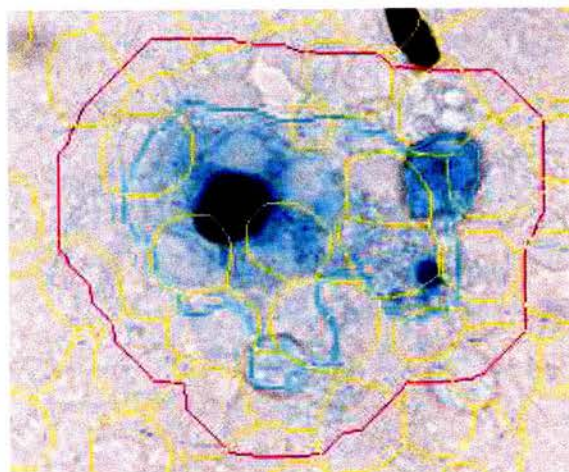


Figura 5.3 — Separação das zonas das células

5.3 Classificação das células

Para classificar as células como exprimindo o gene ou não, é necessário ter em atenção o facto de existirem células com uma tonalidade azul mas que não contêm o gene. Essa situação normalmente ocorre quando a célula está próxima de uma célula com gene que foi intensamente marcada com o corante.

Também existem células marcadas pelo corante com uma tonalidade de azul muito ténue e que contêm o gene. Para que o algoritmo de classificação fosse robusto foi necessário que ele tivesse em atenção a zona em que a célula se encontra e tomasse essa informação em conta; assim foram divididas as imagens em dois tipos de zonas:

1. As zonas das células com influência das células marcadas com um azul forte
2. As zonas sem influência das células marcadas com um azul forte.

As zonas com influência do corante foram encontradas de uma forma muito simples, bastando para isso dilatar com um elemento estruturante disk de raio=35 as zonas marcadas como tendo tonalidade azul. O resultado desta operação pode ser visto na figura 5.3, onde a zona detectada é delimitada com uma linha violeta.

Seguidamente, aproveitando o facto de já possuímos o conhecimento da localização dos centros das células, fez-se a classificação das células avaliando a percentagem de pontos com tonalidade azul em torno do centro da célula, mais especificamente, seguiram-se os seguintes passos:

1. Binarizou-se a imagem de campo claro tal como foi anteriormente descrito, a 1 os pontos azuis e a 0 os outros pontos. A binarização foi realizada usando níveis de tonalidade entre 0.5 a 0.6, o que resulta numa imagem parecida com a da figura 5.1, mas que selecciona um tipo de azul um pouco mais próximo do violeta.
2. Avaliou-se a percentagem de pontos com nível 1 num raio de 12 pixels em redor do centro da célula.
3. Classificou-se cada célula como tendo gene se a percentagem de pontos a 1 for superior a 77% caso a célula esteja dentro da zona azul e 44% se estiver fora dessa zona. Estas percentagens foram calculadas de forma experimental e funcionaram para a quase totalidade das imagens disponíveis

para o desenvolvimento do algoritmo, no entanto devido à grande variação da qualidade das imagens não é possível garantir que estes números servirão para todas as imagens, para contornar este problema colocou-se na interface gráfica a opção de alterar estes valores, ver secção 6.2 na página 44.

Posteriormente, e como foi referido no capítulo 4, as zonas de azul muito escuro das imagens de campo claro quase sempre correspondem a zonas nas imagens de fluorescência onde é impossível detectar as células. Porém como estas manchas só acontecem quando existe pelo menos uma célula nessa zona, incluiu-se no algoritmo uma excepção para criar artificialmente uma célula no centro destas zonas, o mesmo foi feito para as zonas de azul mais claro que não contêm nenhuma célula no seu interior; nestes casos foi necessário ter o cuidado de que quando aparece uma zona de azul sem nenhuma célula mas com uma zona de azul escuro no interior, apenas se incluir uma célula.

5.4 Análise do algoritmo

Este algoritmo produz resultados bastante bons e é capaz de superar as dificuldades de variação da quantidade de azul do corante caso a célula esteja próxima ou distante de outras células com gene e de as células sem gene muito próximas das células com gene também apresentarem uma tonalidade azul, como pode ser visto nos exemplos das Figuras 5.4 e 5.5 onde foi usada a seguinte convenção de cores:

1. *Linhas*

- *Azul claro - Zona das células com gene.*
- *Azul escuro - Zona das células com azul muito forte.*
- *Amarelo - Contornos das células detectados nas imagens de fluorescência.*
- *Violeta - Zona de influencia do corante.*

2. Pontos

- Verde- *Célula sem gene.*
- Laranja claro - *Célula com gene.*
- Laranja escuro- *Célula com gene que se encontra numa zona não visível na imagem de fluorescência.*

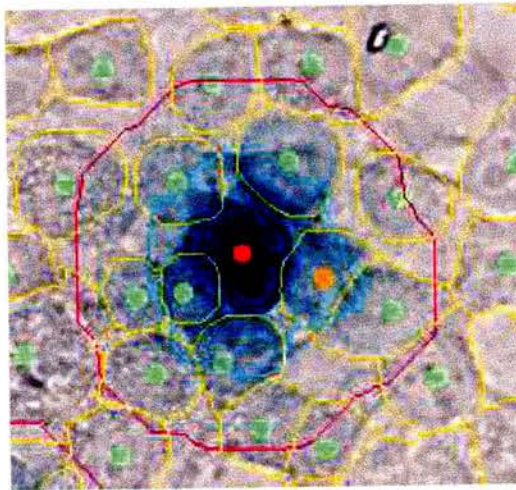


Figura 5.4 — Resultado da classificação das células

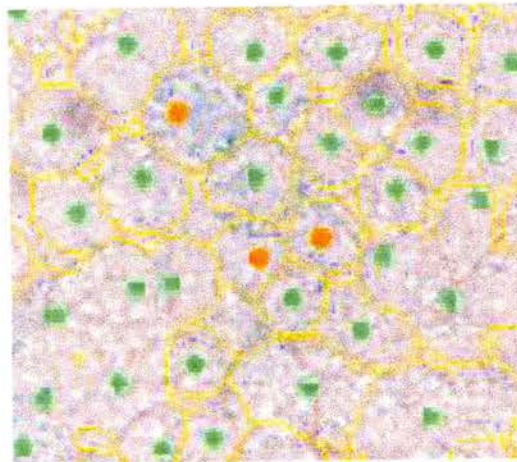


Figura 5.5 — Resultado da classificação das células numa zona com azul muito ténue

Capítulo 6

Programa

Sumário: *Neste capítulo será dada uma visão geral sobre a interface gráfica.*

6.1 Introdução

Para que o software de cálculo do número de células pudesse ser usado de uma forma fácil por pessoas sem um conhecimento aprofundado do algoritmo criou-se uma interface gráfica bastante simples e intuitiva.

O facto do algoritmo ser automático e quase todos os parâmetros se adaptarem naturalmente às imagens, facilitou a implementação de uma interface simples, pois o número de parâmetros a introduzir durante a operação é muito reduzida.

6.2 Interface gráfica

Na interface principal, apresentada na Figura 6.1, existem o botão *Fluoresc.* que permite escolher uma imagem de fluorescência, o botão *Campo claro* que permite escolher uma imagem de campo claro e o botão *Calcular* que faz correr o algoritmo. Existe também a opção *Abrir imagem* que, se estiver activa, permite a abertura de uma janela com essa imagem.

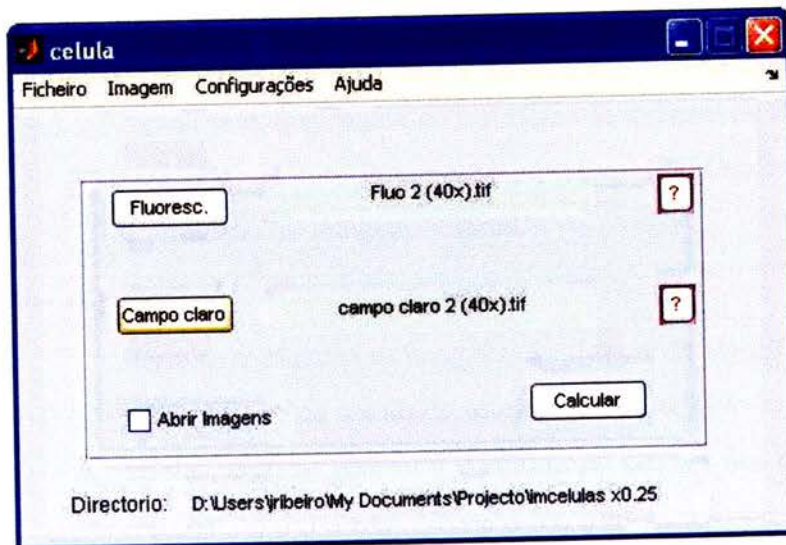


Figura 6.1 — Interface gráfica

Na parte superior da interface existem as opções:

1. *Ficheiro*

(a) *Definir Directório de origem*

Permite definir como directório de origem a pasta que contém as imagens que se querem analisar, o que faz com que seja mais rápida a pesquisa das imagens.

(b) *Fechar tudo*

Fecha todas as janelas de imagens que foram abertas através do programa.

(c) *Sair*

Fecha a janela do programa.

2. *Imagem*

(a) *Iluminação*

Permite escolher o tipo de iluminação do resultado apresentado nas imagens de fluorescência.

Existem três tipos de iluminação, *Original*, que coloca no resultado da imagem de fluorescência a mesma iluminação da imagem original, *Equalizada*, que coloca no resultado da imagem de fluorescência a imagem original com iluminação equalizada, e *Duplicada* que duplica a iluminação da imagem original.

(b) *Opções*

Permite configurar as imagens resultantes do algoritmo, definindo se se pretende que no resultado estejam desenhadas as orlas e os centros das células, caso se pretenda desenhar os centros das células é necessário escolher se devem ser desenhados com várias cores para distinguir o método que foi utilizado para detectar cada célula, ou se se deve utilizar a mesma cor para todas as células, utilizando apenas uma cor diferente

no resultado das imagens de campo claro para marcar as células que foram classificadas como tendo gene.

3. *Configurações*

(a) *Limiares Genes*

Permite alterar os valores das percentagens de pixels azuis na região circular de raio 12 e centro na célula, que são necessários para a classificação da célula como exprimindo o gene.

Como foi referido na secção 5.3, classificou-se cada célula como exprimindo o gene se a percentagem de pontos for superior a 77% caso a célula estivesse dentro da zona azul e 44% se a célula estiver fora dessa zona; porém, como devido à grande variação nas imagens não foi possível garantir que estes números servem para todos os tipos de imagens, colocou-se na interface gráfica a opção de alterar estes valores.

(b) *Configurações originais*

Permite carregar todas as configurações originais.

- Iluminação original.
- Desenha as orlas.
- Desenha os centros das células distinguindo o método de detecção utilizando várias cores.
- Percentagem de pontos azuis necessários para classificar as células como tendo gene 77% no interior e 44% no exterior.
- Directório de origem em C:.

4. *Ajuda*

(a) *Ajuda Célula*

Abre um ficheiro de instruções com ajudas para a utilização do programa.

(b) *Sobre Célula*

Abre uma janela com a ficha técnica do programa.

Capítulo 7

Resultados e Conclusões

Sumário: *Neste capítulo serão comparados os resultados das contagens automáticas e manuais e será feita uma breve conclusão do projecto.*

7.1 Resultados

As tabelas 7.1 e 7.2 mostram as contagens efectuadas manualmente e as contagens automáticas para as imagens disponíveis no desenvolvimento do algoritmo; já as tabelas 7.3 e 7.4 mostram os resultados para as imagens de validação do algoritmo; na tabela 7.3 mostra também os valores de uma contagem automática em que os resultados foram revistos manualmente por uma pessoa especializada.

Seguidamente, serão apresentados as imagens resultantes das colheitas onde se obtiveram os melhores e os piores resultados tanto para as imagens de desenvolvimento como para as imagens de validação.

7.1.1 Resultados nas imagens de desenvolvimento do algoritmo

Número de células na imagem		
	Contagem manual	Automática
1	320	341
2	257	282
3	359	382
4	288	317
5	543	517

Tabela 7.1 — Contagem do número total de células nas imagens de desenvolvimento do algoritmo

Número de células na imagem que exprimem o gene		
	Contagem manual	Automática
1	9	17
2	41	53
3	29	39
4	16	23
5	20	32

Tabela 7.2 — Contagem do número de células que exprimem o gene nas imagens de desenvolvimento do algoritmo

7.1.2 Resultados nas imagens de avaliação do algoritmo

Número de células na imagem			
	Contagem manual	Automática	Automática revista manualmente
6	394	400	406
7	341	351	347
8	394	404	440
9	425	448	450
10	475	457	497
11	479	458	512

Tabela 7.3 — Contagem do número total de células nas imagens de avaliação do algoritmo

Número de células na imagem que exprimem o gene		
	Contagem manual	Automática
6	179	169
7	47	41
8	98	135
9	182	138
10	236	200
11	245	282

Tabela 7.4 — Contagem do número de células que exprimem o gene nas imagens de avaliação do algoritmo

Para as colheitas 6, 8 e 9 foi necessário adaptar os parâmetros de selecção das células com genes devido a apresentarem coloração azul mesmo em zonas sem células ou com células sem o gene beta-galactosidase; em todas as outras colheitas foram usados os parâmetros por defeito.

Como se pode verificar na tabela 7.3 o algoritmo detectou quase sempre um número superior de células do que a contagem manual, no entanto após os resultados terem sido verificados por pessoas especializadas verificou-se que em todas as colheitas, com excepção da colheita número 7, que o número de células assim contabilizado era superior ao das contagens automáticas, o que leva a querer que as contagens automáticas são tão fiáveis como as contagens manuais; sendo que para provar isto seria necessário analisar um número superior de colheitas, o que não foi possível.

7.2 Exemplos de resultados

São aqui apresentadas as imagens em que o algoritmo produziu melhores resultados bem como as imagens de onde se obteve os piores resultados.

As imagens são apresentadas pela seguinte ordem:

- Figura 7.1 e 7.2 - Melhores resultados nas imagens de desenvolvimento.
- Figura 7.3 e 7.4 - Piores resultados nas imagens de desenvolvimento.
- Figura 7.5 e 7.6 - Melhores resultados nas imagens de validação.
- Figura 7.7 e 7.8 - Piores resultados nas imagens de validação.

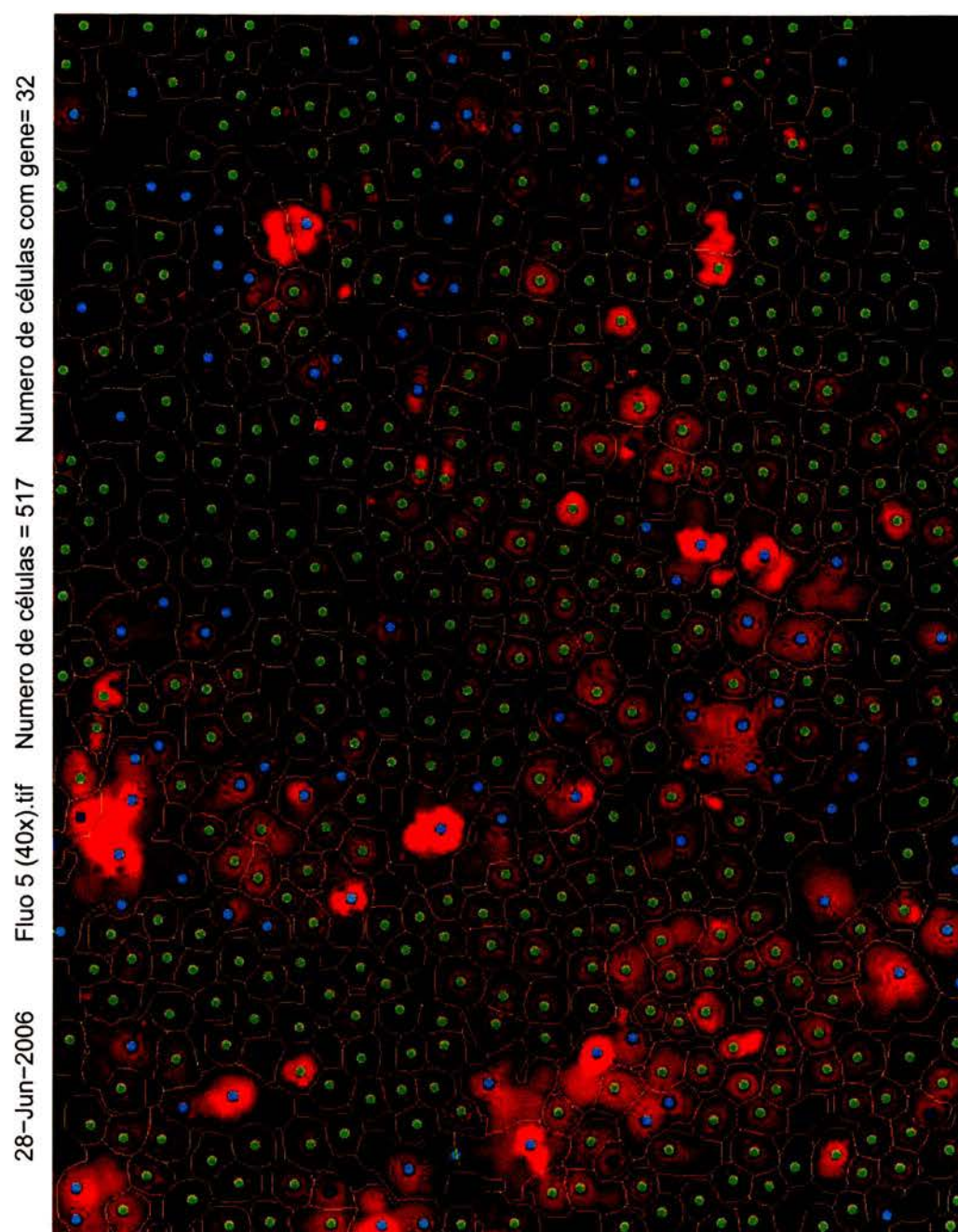


Figura 7.1 — Resultado da imagem de fluorescência, (colheita n^o5 usando parâmetros por defeito)

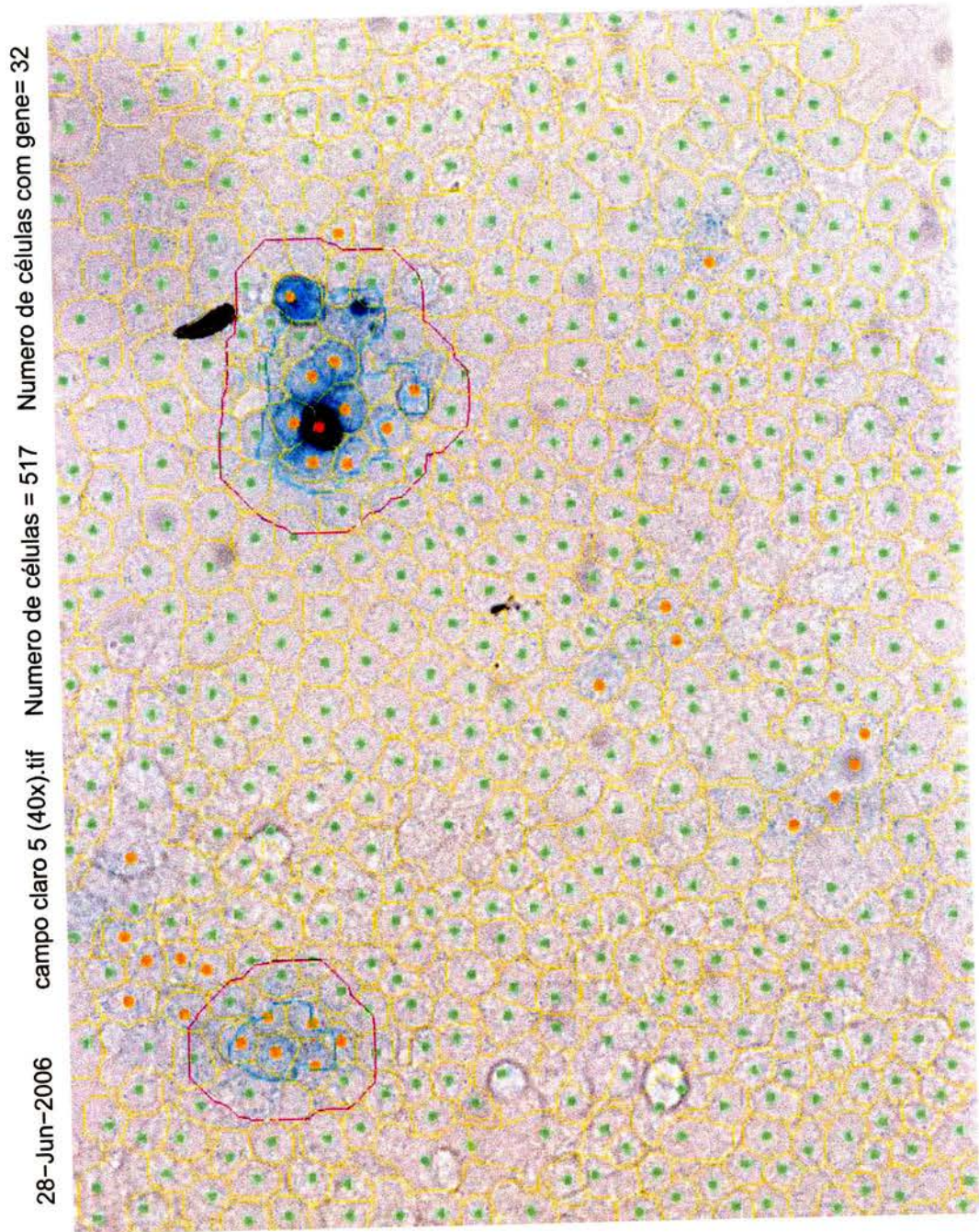


Figura 7.2 — Resultado da imagem de campo claro, (colheita n°5 usando parâmetros por defeito)

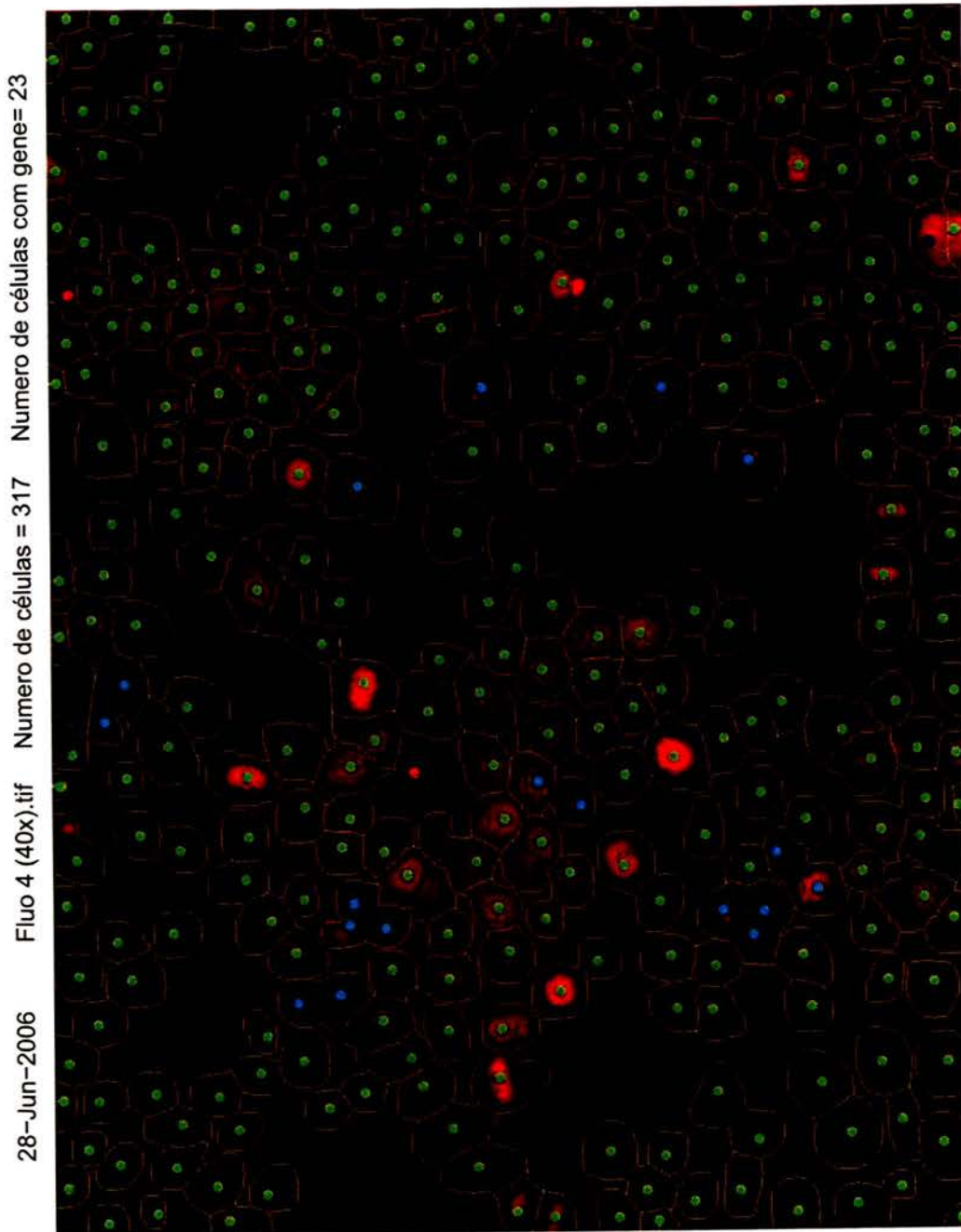


Figura 7.3 — Resultado da imagem de fluorescência, (colheita n^o4 usando parâmetros por defeito)

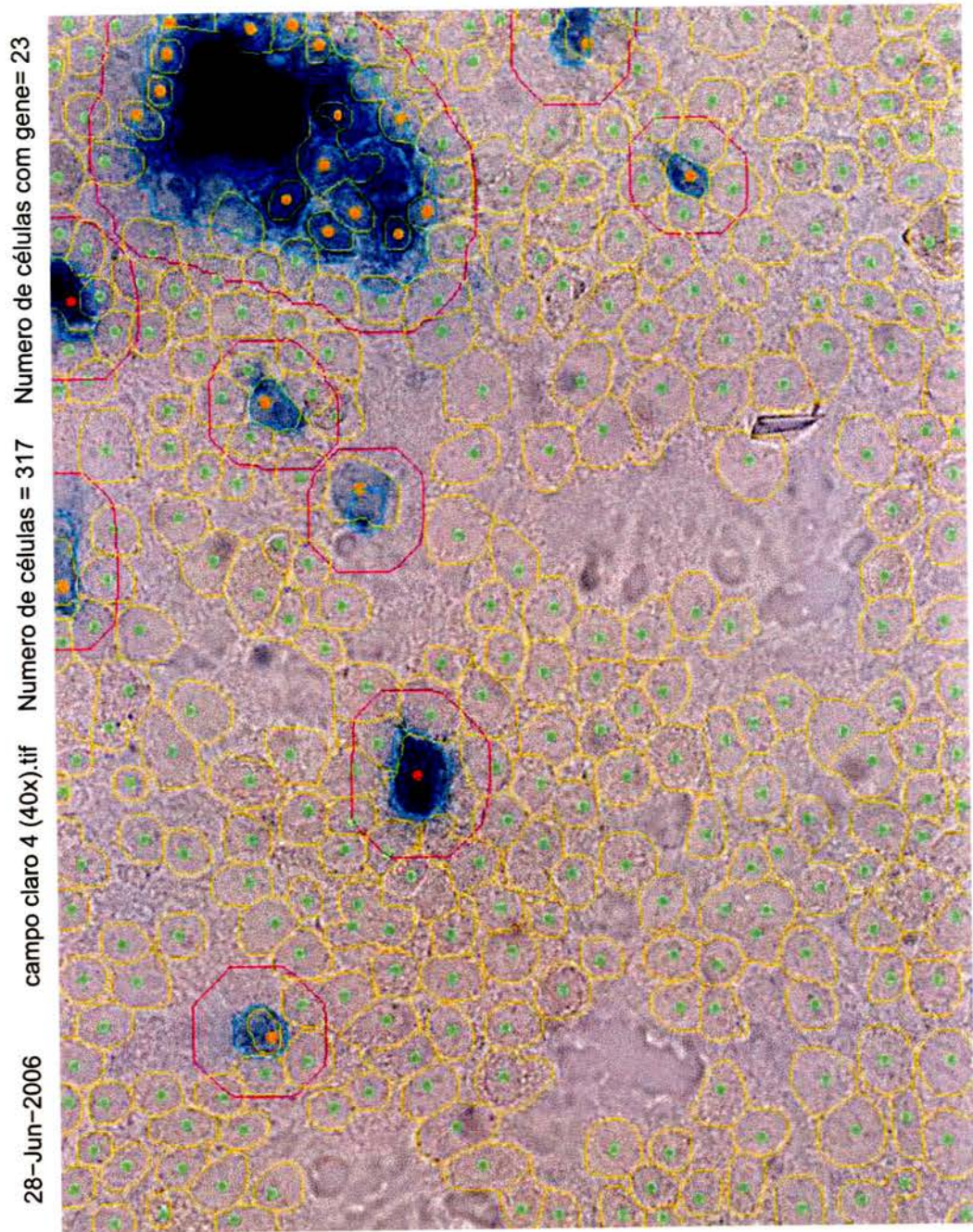


Figura 7.4 — Resultado da imagem de campo claro, (colheita n^o4 usando parâmetros por defeito)

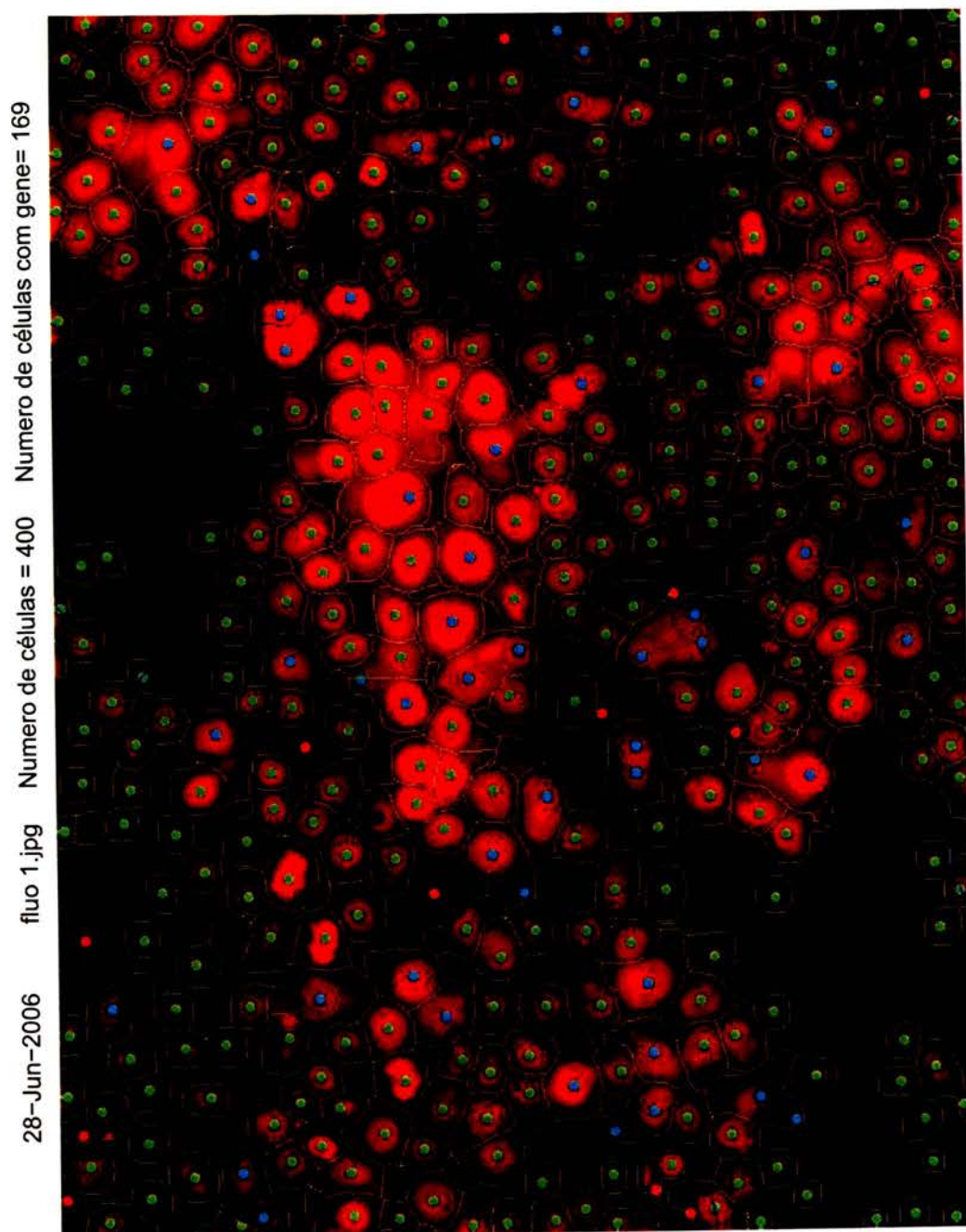


Figura 7.5 — Resultado da imagem de fluorescência, (colheita nº6 usando como parâmetros para a contagem do número de células que exprimem o gene: Pontos interiores=94% e Pontos exteriores=99%)

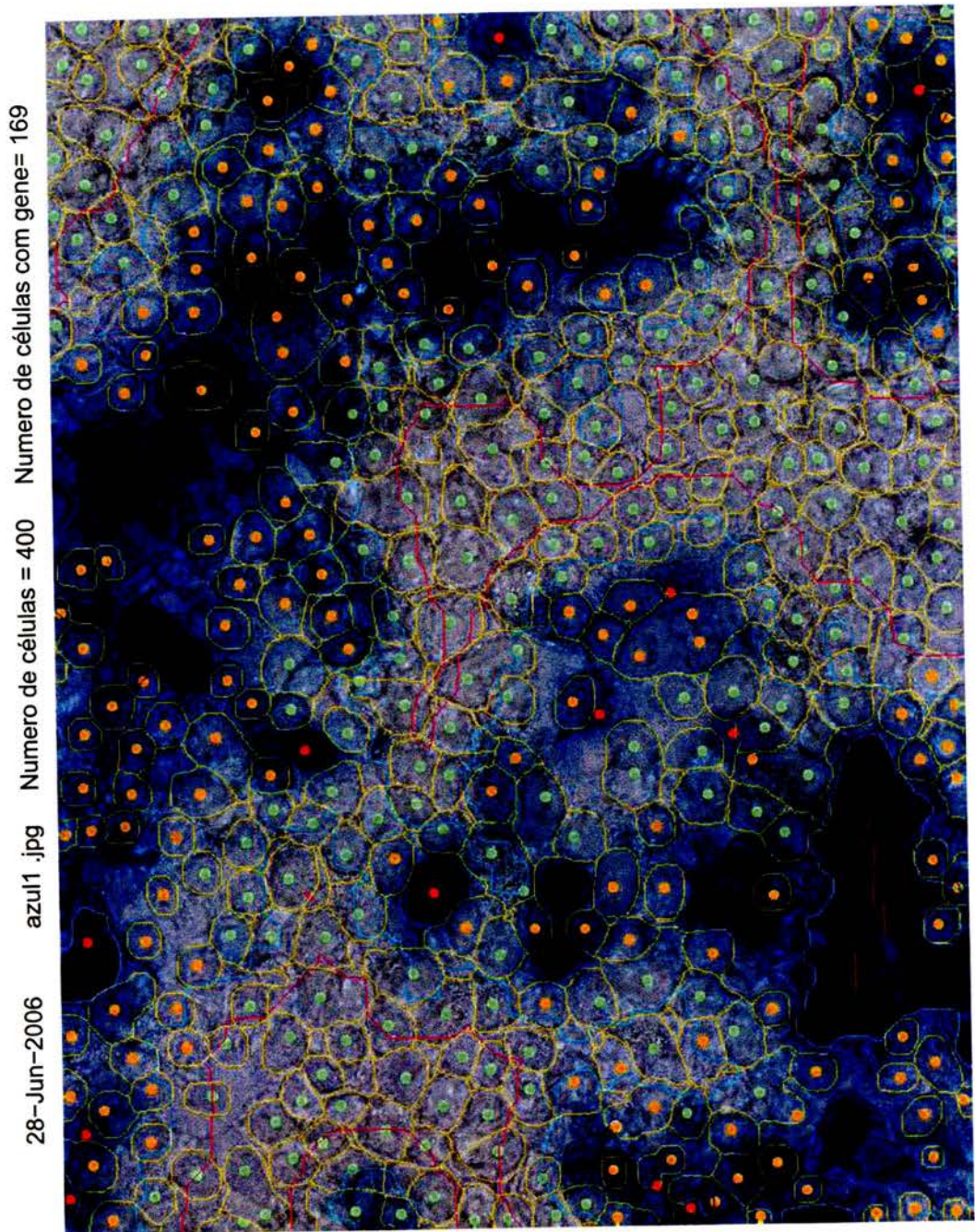


Figura 7.6 — Resultado da imagem de campo claro, (colheita nº6 usando como parâmetros para a contagem do número de células que exprimem o gene: Pontos interiores=94% e Pontos exteriores=99%)

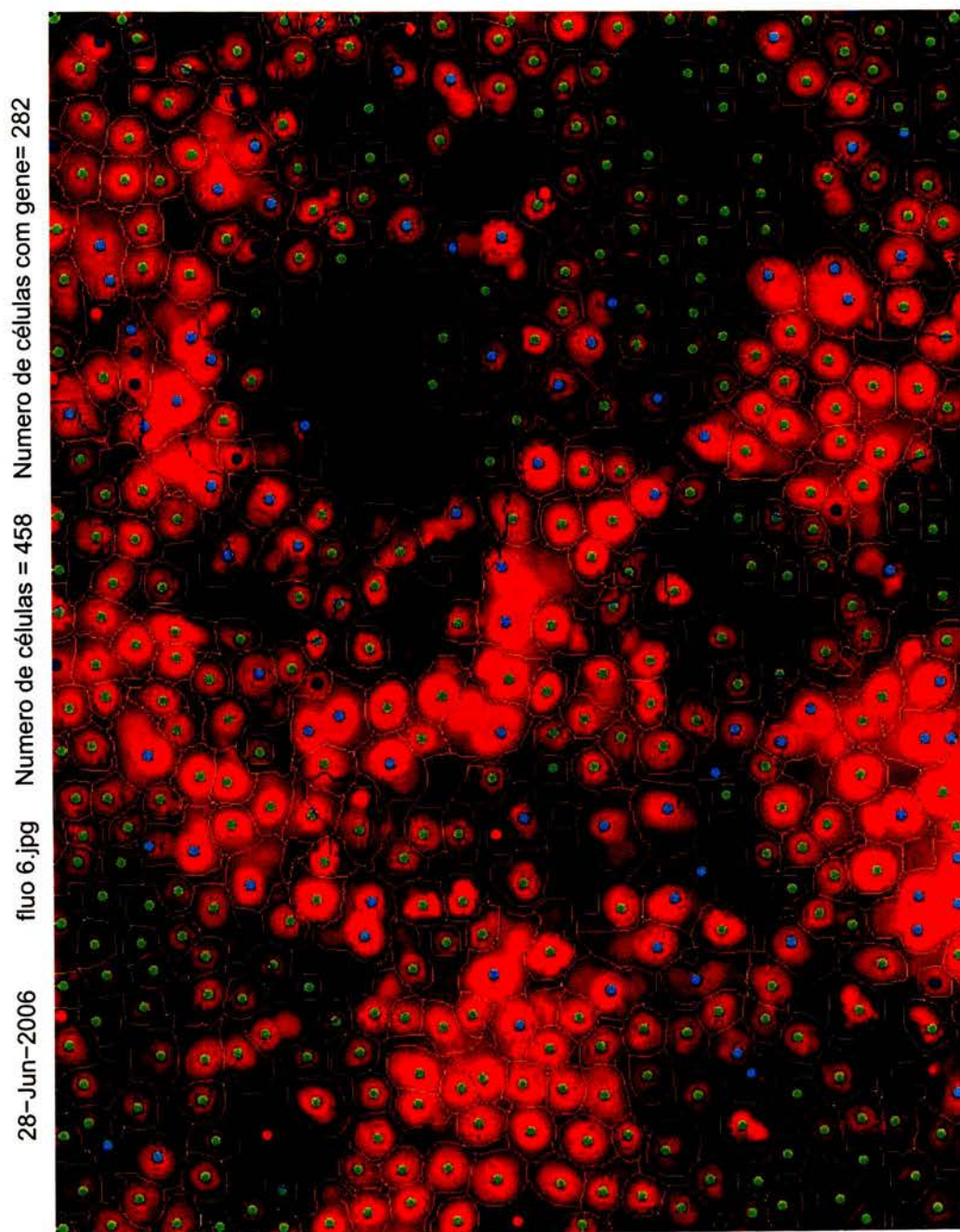


Figura 7.7 — Resultado da imagem de fluorescência, (colheita n^o11 usando parâmetros por defeito)

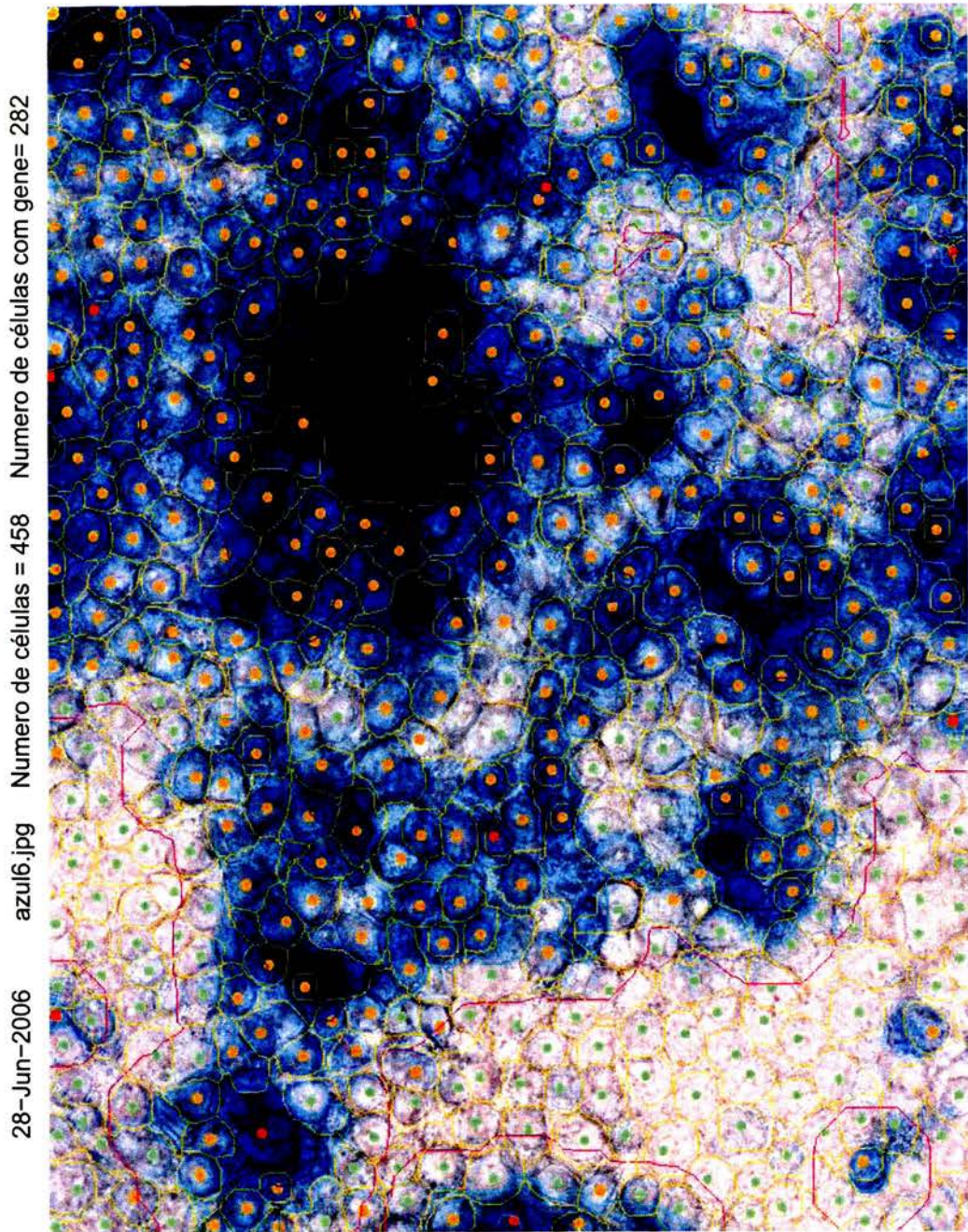


Figura 7.8 — Resultado da imagem de campo claro, (colheita n^o11 usando parâmetros por defeito)

7.3 Conclusões

Desenvolveu-se um algoritmo capaz de calcular de forma totalmente automática o número total de células com resultados muito próximos dos resultados das contagens manuais, como se pode verificar nas tabelas 7.1 e 7.3; sendo que a contagem automática produz valores que normalmente são superiores aos valores das contagens manuais, mas, no entanto as contagens manuais apoiadas com os resultados das contagens automáticas resultam normalmente em valores superiores às contagens automáticas, o que leva a querer que as contagens automáticas tão fiáveis como as contagens manuais.

A contagem do número de células com gene foi mais problemática devido à qualidade muito diversa das imagens de campo claro, tendo sido necessário incorporar na interface gráfica dois parâmetros que podem ser ajustados para permitir ao algoritmo adaptar-se de forma mais eficaz às imagens de menor qualidade, sendo que também aqui se conseguiu obter bons resultados nas contagens, mesmo utilizando os parâmetros por defeito.

Conseguiu-se também que este algoritmo produzisse não apenas dois números, (número total de células e o número de células com gene) mas também duas imagens, onde são identificados os contornos e os centros das células detetadas, sendo facilmente perceptível a localização das zonas e das células onde o algoritmo teve maior dificuldade na contagem. Estes resultados permitem uma correcção muito fácil dos valores de contagem obtidos de forma automática, ou seja permitem com um trabalho mínimo para o utilizador aliar as qualidades da facilidade da contagem automática às qualidades da experiência do humano na contagem manual.

Bibliografia

- [1] I.T. Young, J.J. Gerbrands, L.J. van Vliet, *Image Processing Fundamentals*

Available in HTML version from:

<<http://www.ph.tn.tudelft.nl/Courses/FIP/noframes/fip-Contents.html>>

- [2] Mathworks

Available in HTML version from:

<<http://www.mathworks.com>>

- [3] Digital Image Processing, by Rafael C. Gonzalez and Richard E. Woods;
Prentice Hall © 2002





FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

BIBLIOTECA



0000105237