

Faculdade de Engenharia da Universidade do Porto



FEUP

**Detecção de núcleos de células em sequências de
imagens de microscopia confocal**

Tiago Guichard

VERSÃO FINAL

Dissertação realizada no âmbito do
Mestrado Integrado em Engenharia Electrotécnica e de Computadores
Major Telecomunicações

Orientador: Professora Doutora Ana Maria Mendonça

Julho de 2008

© Tiago Guichard, 2008

Resumo

Este trabalho nasce da necessidade de encontrar um método para a detecção e segmentação automática de núcleos de células em imagens de microscopia confocal. Esta necessidade provém do facto da segmentação destes objectos ser um trabalho moroso, gerando normalmente resultados imprecisos e muito subjectivos, enquanto que a segmentação com recurso a métodos de processamento e análise de imagem é muito mais cómoda e rápida, mantendo-se sempre objectiva. As imagens em estudo representam secções da raiz de uma planta, a *Arabidopsis Thaliana*, que é usada com grande frequência em estudos de biologia de desenvolvimento.

Com este objectivo em mente, e atendendo às características dos núcleos, designadamente a sua forma aproximadamente circular, começa-se por usar um filtro de realce para facilitar a localização das áreas da imagem que possam ser associadas com os objectos de interesse. O filtro usado tem uma resposta tanto mais alta quanto mais perto do centro de uma região circular está do ponto em análise. Desta forma, na imagem realçada, as regiões dos núcleos estão associadas a valores elevados comparativamente ao resto da imagem. Tendo em conta este facto, determinam-se os máximos locais da imagem que são os potenciais centros dos núcleos.

Posteriormente é feita a análise dos pontos determinados pelo processo antes descrito, e com base na sua posição e características da região envolvente, são removidos os pontos que não satisfazem determinadas condições. De seguida são reajustadas as áreas onde os núcleos das células podem ser encontrados de forma a permitir a separação de alguns núcleos que, por pertencerem a camadas celulares diferentes, resultam sobrepostos.

Abstract

This work results from the need to find a method to perform automatic detection and segmentation of cell nuclei on images of confocal microscopy. The main reason for this project results from the fact that manual segmentation of these objects is a slow task, normally ending up with imprecise and subjective results, while segmentation based on image processing and analysis methods is faster and more convenient since the results always remain objective. The images to study represent root sections of a plant, *Arabidopsis Thaliana*, which is frequently used in development biology studies.

With this purpose in mind, and since the cell nuclei show rounded shape, the first step of this method starts by the use of an enhancement filter to ease the detection of the localization of the areas that may correspond to nuclei. The filter response is higher in the points that are closer to the center of a rounded region. So, on the enhanced image, the nucleus regions are associated with high intensity values compared to the rest of the image. As a result, the image local maxima are potential centers of the nuclei.

Afterwards, an analysis of the points found after applying the enhancement filter is performed, and based on their position and surrounding area properties, some points are removed. Finally, some areas are adjusted so it is possible to allow the splitting of some nuclei that are overlapped because they belong to different cell layers.

Agradecimentos

Em primeiro lugar, agradeço à Professora Ana Maria Mendonça toda a ajuda e disponibilidade demonstrada durante todo o trabalho. Ao Professor Aurélio Campilho por me ter despertado o interesse nesta área da Engenharia Electrotécnica. Ao Pedro e à Mónica pelos conselhos e ideias oferecidos ao longo do trabalho. E finalmente aos meus pais por me terem apoiado todos estes anos que frequentei a Universidade.

Índice

Resumo	iii
Abstract	v
Agradecimentos	vii
Índice	ix
Lista de figuras	xi
Lista de equações.....	xiii
Abreviaturas e Símbolos	xv
Capítulo 1	1
Introdução	1
1.1. Motivação	1
1.2. Estrutura da Dissertação	2
Capítulo 2	3
Revisão Bibliográfica	3
2.1. Introdução	3
2.2. Realce de Imagem	3
2.3. Segmentação de núcleos	7
Capítulo 3	11
Metodologia	11
3.1. Introdução	11
3.2. Realce da Imagem.....	12
3.3. Determinação dos máximos locais	14
3.4. Análise e selecção dos núcleos	18
3.5. Conclusão	24
Capítulo 4	27
Resultados	27
4.1. Introdução	27
4.2. Realce da Imagem.....	28
4.3. Identificação dos núcleos	31
4.4. Conclusão	34

Capítulo 5	35
Conclusões	35
5.1. Conclusão.....	35
Referências	37

Lista de figuras

2.1 - Filtro de Moeda.....	5
2.2 - Filtro de Íris.....	5
2.3 - Filtro de Anel Adaptativo.....	6
3.1 - Esquema do método desenvolvido.....	11
3.2 - Exemplo de uma imagem original.....	12
3.3 - Filtro SBF	13
3.4 - A: Imagem original, B: Resultado do realce da imagem original, B: <i>Zoom</i> da imagem original da zona marcada em B, C: <i>Zoom</i> da imagem realçada da zona marcada em B	14
3.5 - Máximos locais da imagem realçada.....	15
3.6 - Máximos locais da imagem realçada marcados sobre: A- na imagem original, B- na imagem realçada.....	15
3.7 - Exemplos de alguns máximos localizados nas paredes	16
3.8 - A: Imagem das paredes, B: Após filtro de simetria de fase	17
3.9 - Antes (A) e depois (B) da utilização da informação da informação das paredes	17
3.10 - Antes (A) e depois (B) de introduzir a informação das paredes	18

3.11 - Características da Elipse	19
3.12 - A: Raios máximos, B: Elipses equivalentes	19
3.13 - Segmentação por crescimento de regiões	20
3.14 - A: Regiões marcadas utilizando as elipses calculadas a partir dos dados do Filtro SBF, B: Zoom da Região seleccionada.....	20
3.15 - A: Antes da tentativa de separação, B: Depois da tentativa de separação...	21
3.16 - Esquema do algoritmo de selecção	22
3.17 - Antes (A) e depois (B) de aplicar o algoritmo de selecção	23
3.18 - A: Imagem original. B: Imagem realçada. C: Imagem das paredes. D: Máximos locais da imagem realçada. E: Máximos locais tendo em conta a informação das paredes. F: Simetria de fase da imagem das paredes. G: Zonas dos máximos marcadas na imagem realçada. H: Zonas dos máximos que têm intersecções marcadas na imagem da média geométrica entre a imagem original e a imagem realçada. I: Imagem Original com as zonas seleccionadas marcadas.....	24
3.19 - Resultado Final da aplicação da metodologia.....	25
4.1 - Imagens originais.....	27
4.2 - A: Imagem original, B: Filtro de moeda. C: Filtro de íris, D: Filtro de anel adaptativo, E: Filtro SBF	29
4.3 - Máximos locais de: A: Filtro de moeda, B: Filtro de íris, C: Filtro de anel adaptativo, D: Filtro SBF	30
4.4 - Máximos Locais	31
4.5 - Elipses equivalentes	32
4.6 - Imagens Finais	33

Lista de equações

2.1 - Orientação do vector de gradiente	4
2.2 - Definição genérica de Coeficiente de Convergência	4
2.3 - Definição de coeficiente de convergência para o Filtro de Moeda.....	5
2.4 - Definição de coeficiente de convergência para o Filtro de Íris	5
2.5 - Definição de coeficiente de convergência para o Filtro ARF	6
3.1 - Definição de coeficiente de convergência para o Filtro SBF	13
3.2 - Equação da Elipse	18

Abreviaturas e Símbolos

Lista de abreviaturas (ordenadas por ordem alfabética)

DEEC	Departamento de Engenharia Electrotécnica e de Computadores
FEUP	Faculdade de Engenharia da Universidade do Porto
CF	Coin Filter
IF	Iris Filter
ARF	Adaptative Ring Filter
SBF	Sliding Band Filter

Símbolos:

θ, \varnothing - ângulos

Capítulo 1

Introdução

1.1 Motivação

Este documento descreve um método desenvolvido para seleccionar, em imagens de microscopia confocal, áreas onde se podem encontrar os núcleos das células da raiz de uma planta, a *Arabidopsis Thaliana*. Esta planta cumpre uma função muito semelhante à da Mosca da fruta (*Drosophila Melanogaster*), mas para a biologia vegetal.

As principais razões para o uso da *Arabidopsis Thaliana* para estudos de biologia de desenvolvimento são:

- possui um ciclo de vida de 6 semanas (desde a germinação até à fase adulta), o que reduz o tempo de inactividade e facilita a análise genética de várias gerações ao fim de pouco tempo;
- é muito pequena (aproximadamente 20 cm), o que possibilita uma mais fácil captação de imagens durante o desenvolvimento;
- tem um genoma pequeno, o que facilita a identificação de mutações genéticas devido a estímulos externos;
- pode-se auto-fertilizar e polinizar;
- produz milhares de sementes por planta;
- pode-se transformar facilmente através de mutagénicos comuns;
- tem uma enorme variedade de espécies mutantes.

Como tal, esta planta é para os investigadores uma ferramenta muito valiosa. No entanto, a segmentação manual dos núcleos das células é, devido principalmente à quantidade de imagens captadas durante o crescimento e também ao número de células por imagem, lenta, imprecisa e subjectiva. Desta forma, este trabalho surge da necessidade de desenvolver um processo para tornar a análise deste tipo de

imagens mais rápida, mais cómoda e mais objectiva. O recurso a métodos de processamento de imagem é pois natural neste tipo de aplicação.

Algumas características dos núcleos, como por exemplo o tamanho, a forma e a posição relativa entre núcleos, podem conter informações muito interessantes para um investigador desta área. Estas características podem ajudar a descobrir, por exemplo, quais as células em divisão, em que fase de desenvolvimento se encontram, ou quais as zonas de maior actividade ou crescimento.

Desta forma, para conseguir atingir os objectivos propostos, o método que vai ser descrito em pormenor no Capítulo 3 deste documento pode ser resumido nos três passos principais a seguir mencionados:

- realce das imagens para salientar as zonas convexas (núcleos);
- localização dos potenciais núcleos partindo da imagem realçada;
- delimitação final das regiões onde os núcleos podem ser encontrados.

1.2 Estrutura da Dissertação

Esta dissertação contém 5 capítulos, incluindo esta introdução. No capítulo 2 é apresentada a revisão bibliográfica feita. No capítulo 3 descreve-se o método proposto. No capítulo 4 mostram-se os resultados da aplicação do método a algumas imagens de teste. No capítulo 5 apresentam-se as principais conclusões do trabalho realizado.

Capítulo 2

Revisão Bibliográfica

2.1 Introdução

Grande parte dos trabalhos realizados tendo como objectivo a segmentação automática de núcleos de células têm sido efectuados apenas em células animais, ao contrário do que foi desenvolvido neste trabalho, que tem como objectivo a segmentação de núcleos de células de uma planta, a *Arabidopsis Thaliana*. O facto das células animais serem mais estudadas, deve-se às grandes possibilidades que podem advir daí, por exemplo, para a medicina. No entanto, os métodos desenvolvidos para as células animais também podem ser aplicados às células vegetais uma vez que ambos os tipos de células possuem algumas características semelhantes.

Nas secções a seguir, para além da apresentação dos métodos pesquisados sobre segmentação de núcleos de células, serão também referidos métodos para o realce de objectos convexos em imagens, dado que os objectos de interesse deste trabalho possuem esta característica, e o realce prévio das imagens demonstrou ser um passo essencial em todo o processo de análise desenvolvido.

2.2 Realce da Imagem

O realce de uma imagem tem como objectivo salientar as características dos objectos que vão ser alvo de estudo. Esta necessidade resulta, maior parte das vezes, do facto dos objectos terem sido degradados pelo ruído ou terem sofrido distorções durante a fase de aquisição. No entanto, também podem ser usados

métodos de realce para evidenciar apenas os objectos da imagem que possuem determinado tipo de características.

Atendendo a que os objectos que se pretende detectar neste trabalho são predominantemente circulares, serão apresentados apenas métodos para o realce de formas convexas.

Na literatura analisada foram encontrados essencialmente três filtros [1]: de moeda (*coin filter*), de anel adaptativo (*adaptative Ring filter*) e de íris (*iris filter*). Estes filtros pertencem a uma família genericamente denominada de filtros de convergência, uma vez que todos se baseiam na mesma ideia, a maximização do coeficiente de convergência. Este coeficiente é calculado a partir do ângulo do vector gradiente em cada ponto de coordenadas (x,y) da imagem, cuja fórmula é

$$\varnothing(x, y) = \tan^{-1} \frac{G_C(x,y)}{G_R(x,y)}, \quad (2.1)$$

onde $G_C(x,y)$ e $G_R(x,y)$ são os valores das derivadas calculadas segundo as colunas e as linhas da imagem, respectivamente, e $\varnothing(x,y)$ é o ângulo que o vector gradiente forma com o eixo correspondente às linhas.

A partir do ângulo calculado usando a expressão (2.1), define-se o coeficiente de convergência, $C(x,y)$, como sendo o valor médio dos cosenos dos ângulos do gradiente calculados em todos os pontos de uma região R , designada por Região de Suporte do filtro, tal como está descrito na equação (2.2)

$$C(x, y) = \frac{1}{M} \sum_{(k,l) \in R} \cos \theta_i(k, l), \quad (2.2)$$

onde M é o número de pontos da Região R e θ_i é o ângulo do vector gradiente faz com o segmento de recta que liga o ponto de coordenadas (x,y) ao ponto de coordenadas (k,l) .

De forma a facilitar o cálculo, a região de suporte destes filtros corresponde a N semi-rectas que irradiam do ponto (x,y) , onde se pretende calcular o valor do resultado da filtragem.

O que distingue os diversos filtros desta família é essencialmente a definição da região de suporte do filtro.

O filtro de moeda possui uma região de suporte formada por todos pontos pertencentes às semi-rectas de dimensão r , limitadas por uma circunferência de raio máximo R_{max} , tal como se pode verificar na equação 2.3 e na a Figura 2.1. O coeficiente de convergência é tal que

$$CF(x, y) = \max_{0 \leq r \leq R_{max}} \frac{1}{N} \sum_{i=0}^{N-1} \left(\frac{1}{r} \sum_{m=1}^r \cos \theta_{im} \right), \quad (2.3)$$

onde θ_i é o ângulo formado pelo vector gradiente do ponto m e a direcção da semi-recta e r o raio do círculo da região de suporte.

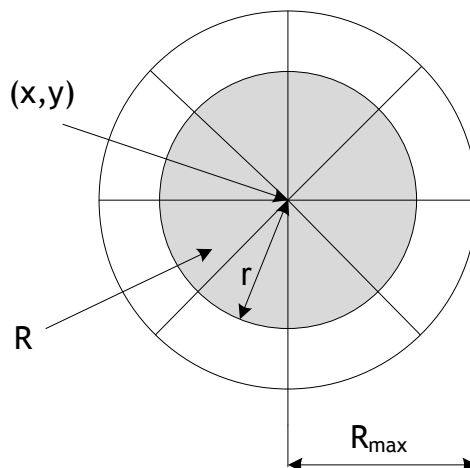


Figura 2.1 - Filtro de Moeda (N=8)

O filtro de Íris adapta a dimensão da região de convergência em cada uma das direcções N de forma a maximizar a média do longo das N semi-rectas, como se pode verificar na equação (2.4) e na Figura 2.2. A equação do filtro é

$$IF(x, y) = \frac{1}{N} \sum_{i=0}^{N-1} \left(\max_{0 \leq n \leq R_{max}} \frac{1}{n} \sum_{m=1}^n \cos \theta_{im} \right). \quad (2.4)$$

onde n corresponde ao número de pixéis ao longo da semi-recta de direcção i usados para o cálculo do filtro, como se mostra na figura seguinte (Figura 2.2).

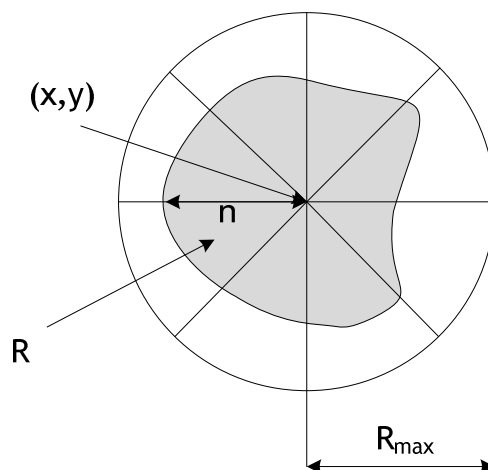


Figura 2.2 - Filtro de Íris (N=8)

Por fim, tem-se o filtro de anel adaptativo (ARF) que usa uma região de suporte em forma de anel de largura fixa, que é um dos parâmetros do filtro. O raio interno vai variando até ser encontrada a região que maximiza o coeficiente de convergência. Esse coeficiente é dado pela equação (2.5),

$$ARF(x, y) = \max_{0 \leq r \leq r_0} \left(\frac{1}{N} \sum_{i=0}^{N-1} \left(\frac{1}{d} \sum_{m=r+1}^{r+d} \cos \theta_{im} \right) \right), \quad (2.5)$$

onde os parâmetros r , r_0 e d são, respectivamente, o valor do raio do círculo interior, o raio máximo do círculo exterior e a largura do anel, como também se pode ver na Figura 2.3.

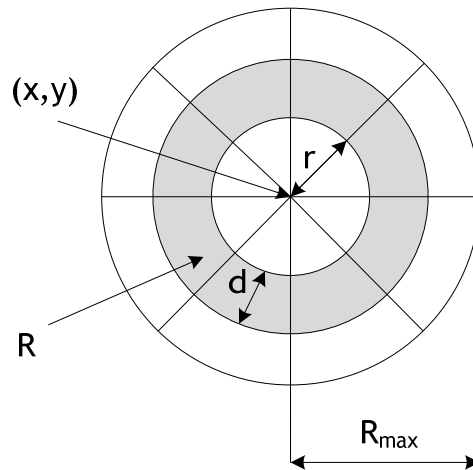


Figura 2.3 - Filtro de Anel Adaptativo (N=8)

O filtro que vai ser usado neste trabalho é também um filtro da classe dos filtros de convergência. É denominado filtro de banda deslizante (*Sliding Band Filter* - SBF) [2] e vai ser designado simplesmente por SBF ao longo deste documento. Apesar de ser um filtro de convergência, tem características um pouco diferentes dos que foram apresentadas antes, dado que combina os conceitos do filtro de íris e do filtro de anel aproveitando as melhores características de cada um destes filtros. A descrição deste filtro será efectuada no Capítulo 3.

2.3 Segmentação de núcleos

Num artigo publicado em 1983 [3] é descrito um método que parte do pressuposto que todos os núcleos têm uma forma circular ou elíptica, e cuja aplicação está condicionada a imagens que contêm um único núcleo completo. Para tal, a imagem inicial com vários núcleos é pré-segmentada em sub-imagens cada uma contendo um núcleo (podendo conter algumas partes de núcleos adjacentes). Posteriormente, para cada sub-imagem, faz-se uma conversão de coordenadas cartesianas para polares. Uma vez que o núcleo tem uma forma aproximadamente circular ou elíptica, na imagem resultante o contorno da célula será uma linha que divide a imagem em duas partes. Partindo da imagem convertida calcula-se o seu gradiente. Depois determina-se a linha que melhor divide a imagem em duas partes. No final, essa linha, quando colocada novamente em coordenadas cartesianas, corresponde ao contorno da célula.

Em 1997 [4] é publicado um artigo que descreve o processo a seguir referido para a segmentação de um aglomerado de núcleos. Os autores começam por efectuar um pré-processamento a todas as imagens de forma a reduzir o ruído introduzido durante a captura. Para tal, cada imagem é multiplicada por uma matriz de controlo de fundo. Esta matriz é calculada a partir da média do desvio padrão de 25 imagens seleccionadas manualmente, que apenas contêm ruído. Posteriormente, num primeiro passo faz-se uma binarização da imagem de forma a separar os núcleos isolados dos aglomerados. O segundo passo vai tratar de preparar a segmentação dos núcleos que estão muito juntos, determinando “sementes” para serem usadas como base para o algoritmo de segmentação baseado em *watershed*. Para tal, é calculada a distância de cada ponto ao fundo (*background*) da imagem, de forma a determinar a localização de cada núcleo. A partir desse resultado marca-se cada núcleo, e depois utilizam-se essas marcações para o algoritmo de segmentação, baseado em *watershed*, determinar os contornos dos núcleos.

Posteriormente, em 1998 [5], é proposta a segmentação dos núcleos através de uma modelação estatística dos dados. Este método começa por calcular as possíveis localizações dos núcleos utilizando os máximos da transformada de Hough simplificada, que é denominada pelos autores de *Compact Hough Transform* (CHT), e que é usada apenas detectar zonas circulares. Com base nos resultados da transformada, e para cada ponto localizado anteriormente, determinam-se alguns pontos do contorno fazendo uma pesquisa radial em 32 direcções diferentes, usando como base a informação fornecida pelo gradiente da imagem e o máximo valor esperado para o tamanho do núcleo. Depois, utiliza-se em cada uma das direcções uma função que maximiza a verosimilhança de forma a melhorar os resultados, e

para finalizar, otimiza-se o contorno, incluindo, ainda, uma condição que garante que a distância entre cada ponto do contorno tem de estar correlacionada com os pixels vizinhos pertencentes a esse contorno, de modo a obter um contorno mais regular.

Em 2001 [6], é publicado um trabalho que segmenta os núcleos usando uma metodologia denominada *kernel-based dynamic clustering*. Este método supõe que os contornos dos núcleos podem ser descritos por um modelo elipsoidal. Em primeiro lugar determina-se o gradiente da imagem, e com base nesta informação obtêm-se pontos com forte probabilidade de pertencerem aos contornos dos núcleos. Após esse passo, usa-se um algoritmo genético adaptado para extrair características relativas aos núcleos e determinam-se os parâmetros do modelo elíptico que melhor se ajusta aos limites do contorno dos núcleos.

Mais tarde, em 2004 [7], é publicado um trabalho que utiliza uma combinação de três tipos de informação para segmentar os núcleos das células. Para tal, são utilizadas informações sobre a intensidade, a forma e os contornos dos objectos das imagens. Este método começa por identificar e separar a imagem em duas partes: o fundo (*background*) e a frente (*foreground*). A parte frontal da imagem é determinada através de uma binarização da imagem original, e a parte do fundo é determinada primeiramente pela binarização da magnitude do gradiente da imagem original seguida da remoção de pequenos objectos dessa imagem. Depois, calculam-se “sementes” para cada uma das imagens anteriores, para posteriormente se fazer uma segmentação por *watershed* usando essas “sementes” como base. Após a segmentação, os objectos muito pequenos são removidos e os objectos com contornos fracos são juntos. Depois calcula-se a distância dos objectos segmentados ao fundo da imagem e aplica-se novamente uma segmentação por *watershed* a esse resultado. No final, tomando por princípio que os núcleos têm uma forma convexa e devem ser independentes uns dos outros, juntam-se as regiões que satisfazem as condições anteriores de forma a obter a segmentação final pretendida.

Um método publicado em 2006 [8] usa o facto dos núcleos terem sido marcados de forma a ficarem fluorescentes quando expostos a certas condições. Esse trabalho divide-se em 3 partes. Em primeiro lugar faz-se uma simples binarização da imagem para segmentar grande parte dos núcleos bem identificados na imagem original. Depois, para os núcleos que estão próximos ou sobrepostos uma vez que pertencem a diferentes camadas, usa-se uma segmentação por *watershed*. Depois da segmentação são extraídas 12 características (máximo, mínimo, média e desvio padrão dos níveis de cinzento do núcleo; área, perímetro, eixo maior, eixo menor, alongamento, ...) e a partir dessas características é determinada a fase em que se encontra a célula para

posteriormente identificar quais aquelas que, potencialmente, se podem vir a dividir, ou já se encontram em fase de divisão.

Outro método recente [9] começa por fazer uma pré-marcação dos núcleos antes de aplicar o algoritmo de *watershed*. Esta pré-marcação é feita através de duas erosões sucessivas sobre a imagem. Primeiro definem-se dois limites para os tamanhos dos objectos que vão sofrer erosão, que são determinados através de uma binarização da imagem original. Depois aplica-se iterativamente uma erosão grosseira, até que o tamanho dos objectos atinja o primeiro limite pré-determinado; de seguida aplica-se, também iterativamente, uma erosão fina até se atingir o segundo limite. A partir da imagem obtida pela sequência de erosões determinam-se as “sementes” que vão ser utilizadas com o algoritmo baseado em *watershed* para concluir a segmentação dos núcleos.

Capítulo 3

Metodologia

3.1 Introdução

Neste capítulo vai ser descrito o método, totalmente automático, para localizar os núcleos das células da raiz da *Arabidopsis Thaliana*, que foi desenvolvido no âmbito deste trabalho. Este método é constituído por 3 partes principais, nomeadamente o realce da imagem de forma a salientar os objectos de estudo, a identificação dos potenciais núcleos, e finalmente, a delimitação das áreas onde se podem encontrar núcleos, tal como se mostra no esquema da Figura 3.1.

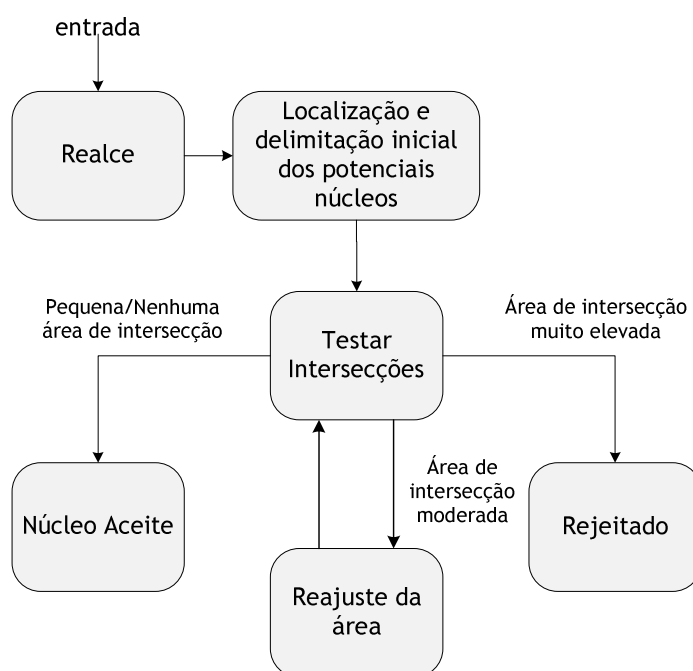


Figura 3.1 - Esquema do método desenvolvido

As imagens disponíveis para estudo são imagens de microscopia confocal, capturadas utilizando uma marcação com uma proteína que quando introduzida na célula se liga ao núcleo. Quando as células são iluminadas por um laser, os núcleos tornam-se fluorescentes. Esta fluorescência permite obter de uma forma mais simples uma imagem com os núcleos relativamente realçados e contrastados. As imagens têm dimensão 512x512 e estão representadas numa escala de 256 níveis de cinzentos. Cada uma contém uma secção da raiz da planta a analisar, como se pode observar na Figura 3.2.

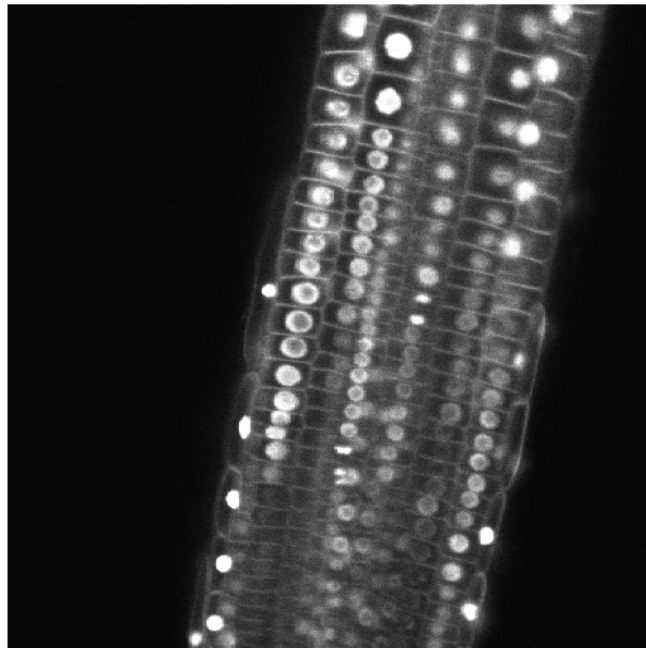


Figura 3.2 - Exemplo de uma imagem original

3.2 Realce da Imagem

Atendendo às características que visualmente surgem destacadas nas imagens pode-se pensar que uma forma simples de detectar a localização dos núcleos pode ser baseada na utilização dos máximos locais da imagem. No entanto, a observação mais cuidadosa da Figura 3.2 revela que muitos dos núcleos das imagens disponibilizadas para o trabalho são menos intensos na sua parte interior do que no exterior, o que impossibilita o uso directo deste método sobre as imagens originais.

Devido a esse facto, é necessário realizar um pré-processamento antes de determinar as possíveis localizações dos núcleos. Dos métodos existentes para realçar a imagem já referidos no Capítulo 2, foi proposto o uso do filtro de Banda

Deslizante (*Sliding Band Filter* - SBF [2]). O principal motivo para o uso deste filtro deve-se ao facto de que é mais flexível do que os outros filtros de convergência quando o interior das zonas a realçar têm um nível de convergência mais aleatório do que as zonas exteriores. A comparação entre os resultados de realce obtidos com os diversos filtros será efectuada no Capítulo 4.

O filtro SBF, tais como os de moeda, de íris e de anel adaptativo, baseia-se na maximização do coeficiente de convergência. A equação do filtro SBF é apresentada na equação 3.1. Tal como acontece com os restantes filtros, também a região de suporte é definida sobre um conjunto de N semi-rectas com origem no ponto (x,y) onde o resultado do filtro está a ser calculado.

$$SBF(x, y) = \frac{1}{N} \sum_{i=0}^{N-1} \left(\max_{R_{min} \leq n \leq R_{max}} \left(\frac{1}{d} \sum_{m=n}^{n+d} \cos \theta_{im} \right) \right) \quad (3.1)$$

Na equação 3.1, d é um parâmetro do filtro e corresponde ao comprimento da banda usada no cálculo, e θ_{im} é o ângulo formado pelo vector gradiente e a semi-recta que descreve a direcção que está a ser analisada.

Para cada uma das N semi-rectas é determinada a posição da banda, especificada através do valor de n na equação (3.1), que maximiza o índice de convergência em cada direcção. Este valor é portanto determinado numa banda fixa de pontos que se vai movimentando desde um valor mínimo do raio (R_{min}) até um raio máximo (R_{max}), tal como pode ser observado na Figura 3.3.

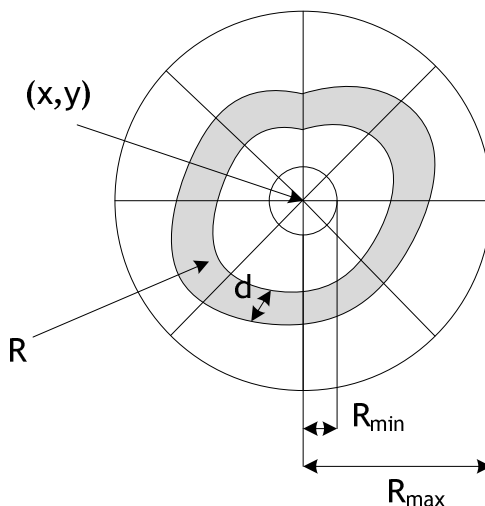


Figura 3.3 - Filtro SBF (N=8)

Tal como o filtro de anel adaptativo, em cada direcção é usado o mesmo número de pontos d para o cálculo do índice de convergência; porém, à semelhança

do filtro de íris, o seu extremo ($n+d$) na semi-recta vai variar entre um valor mínimo e um valor máximo para o raio, de forma a maximizar o coeficiente de convergência.

O resultado da filtragem da imagem original da Figura 3.4-A pode ser observado na Figura 3.4-B, juntamente com um *zoom* da secção assinalada na imagem original (Figura 3.4-C) e na imagem realçada (Figura 3.4-D).

Como se pode observar na Figura 3.4, o filtro resolve em parte o problema das partes menos intensas no interior dos núcleos e também torna a intensidade dos diferentes núcleos mais uniforme. Desta forma será muito mais fácil localizar os núcleos a partir da determinação dos máximos locais.

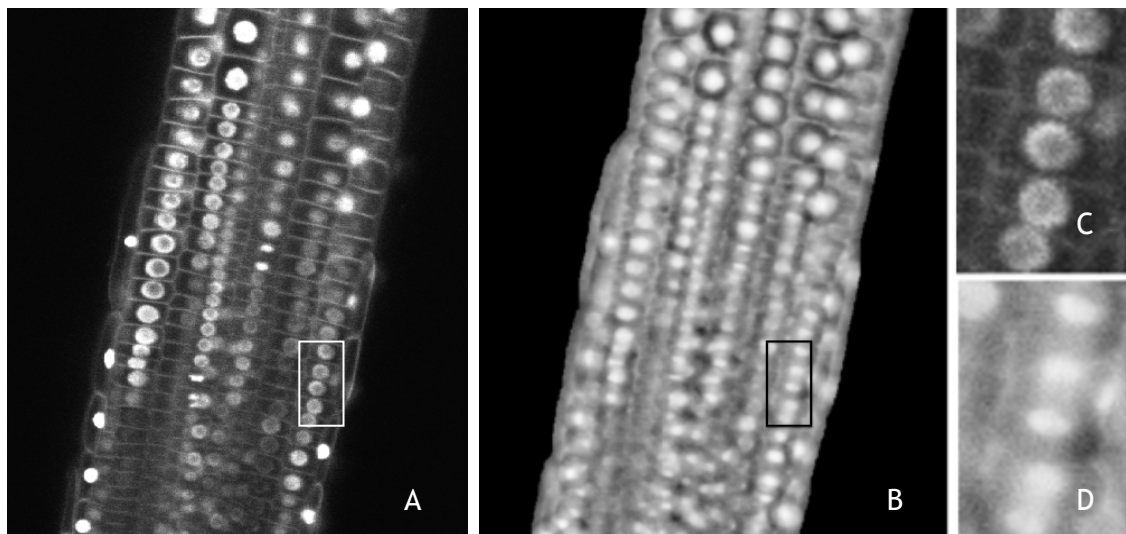


Figura 3.4 - A: Imagem Original, B: Resultado do realce da imagem original, C: Zoom da imagem original da zona marcada em B, D: Zoom da imagem realçada da zona marcada em B.

3.3 Determinação dos máximos locais

Atendendo a que a operação de filtragem torna a intensidade da imagem realçada mais uniforme, e que o seu resultado nos pontos de maior convergência tende para valores próximos de 1, os máximos locais da imagem realçada vão corresponder a possíveis centros para potenciais núcleos. No entanto, o número de máximos locais é muito elevado tal como se pode verificar na Figura 3.5.

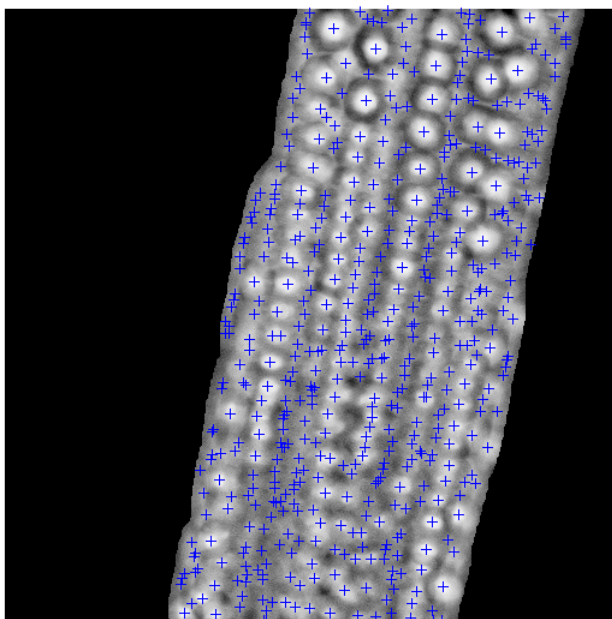


Figura 3.5 - Máximos locais da imagem realçada

De forma a reduzir o número de máximos locais, concluiu-se experimentalmente após testar várias imagens que a média desses máximos é um bom limiar a selecção dos máximos locais a reter. No entanto, para evitar que alguns dos pontos mais intensos da imagem original sejam perdidos, impõe-se que os máximos locais da imagem realçada cuja intensidade seja superior a 0.85 na imagem original não sejam descartados embora na imagem realçada a sua intensidade seja inferior à média.

Nas imagens da Figura 3.6 os máximos locais detectados na imagem realçada são assinalados quer na imagem original (A) quer na imagem realçada (B).

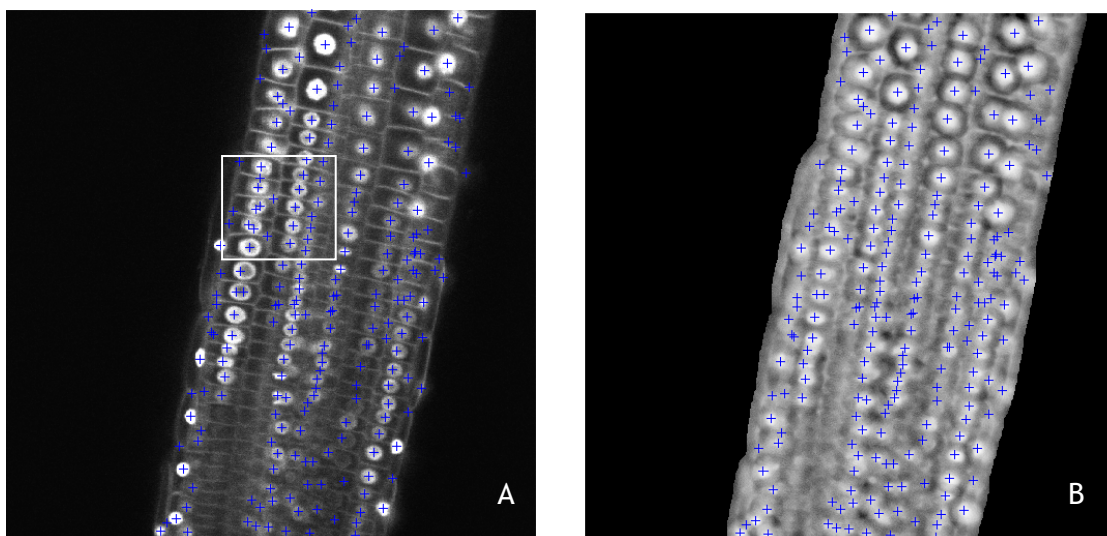


Figura 3.6 - Máximos locais da imagem realçada marcados sobre: A - na imagem original, B - na imagem realçada

Como se pode observar na Figura 3.6 (A e B), existem muitos máximos localizados nas paredes das células, principalmente nas zonas que são comuns a 3-4 células. Este problema deve-se ao facto das paredes estarem muito bem definidas, mesmo naquelas imagens originais onde os núcleos estão mais nítidos, como se pode observar por uma ampliação da região assinalada na Figura 3.6-A que se mostra na Figura 3.7.

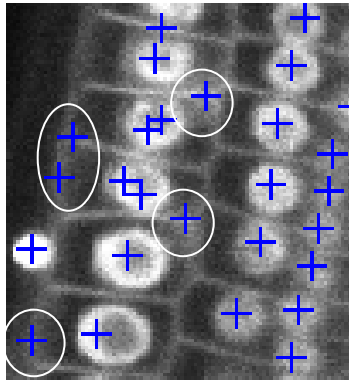


Figura 3.7 - Exemplos de alguns máximos localizados nas paredes

Para resolver este problema, e atendendo a que todas as imagens disponíveis têm associada uma segunda imagem na qual apenas as paredes estão realçadas, como se pode observar na Figura 3.8-A. Vai-se então usar a informação dessa imagem para tentar reduzir o número de máximos localizados nas paredes das células. Uma vez que a imagem das paredes tem muito ruído, e nalguns casos ainda inclui alguns núcleos, vai-se usar um filtro de simetria de fase [10]. O filtro de simetria que é usado neste trabalho utiliza a Transformada *Wavelet* para determinar a informação local de frequência da imagem. Com base nessa informação determinam-se quais os locais onde a simetria de fase é mais elevada, que no caso desta imagem correspondem às paredes. O resultado do filtro pode ser observado na Figura 3.8-B.

Partindo da informação que pode ser observada na Figura 3.8-B, faz-se a sua binarização usando como valor limiar a média das intensidades não nulas. Depois são removidos os máximos da imagem realçada que tomam valor 1 na imagem binarizada das paredes e desta forma consegue-se garantir para a grande maioria dos casos a eliminação dos máximos que se encontram sobre as paredes.

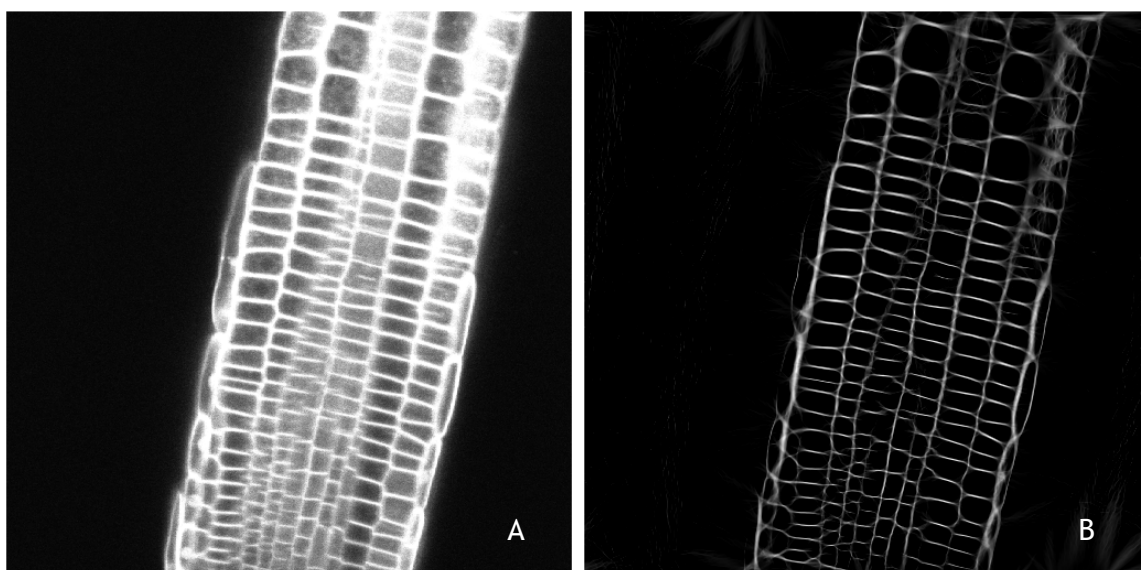


Figura 3.8 - A: Imagem das paredes, B: Após filtro de simetria de fase

Comparando A com B na Figura 3.9, o número de máximos nas paredes é reduzido através da informação das paredes determinada anteriormente.

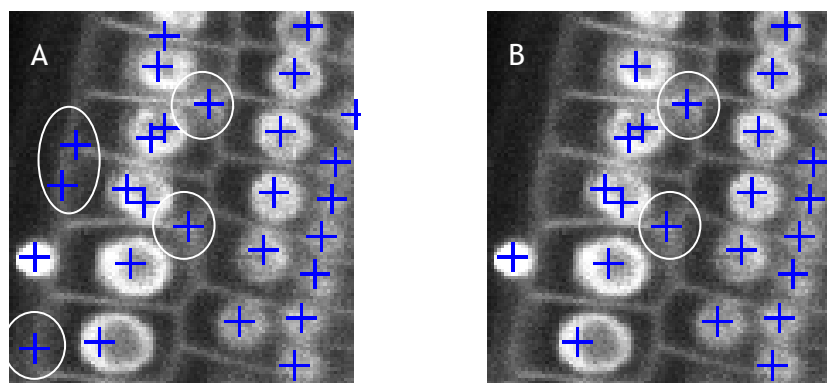


Figura 3.9 - Antes (A) e depois (B) da utilização da informação das paredes

As diferenças anteriores são ainda mais significativas se forem comparadas as imagens completas antes e depois de usar a informação das paredes que se mostram nas Figuras 3.10 A e B.

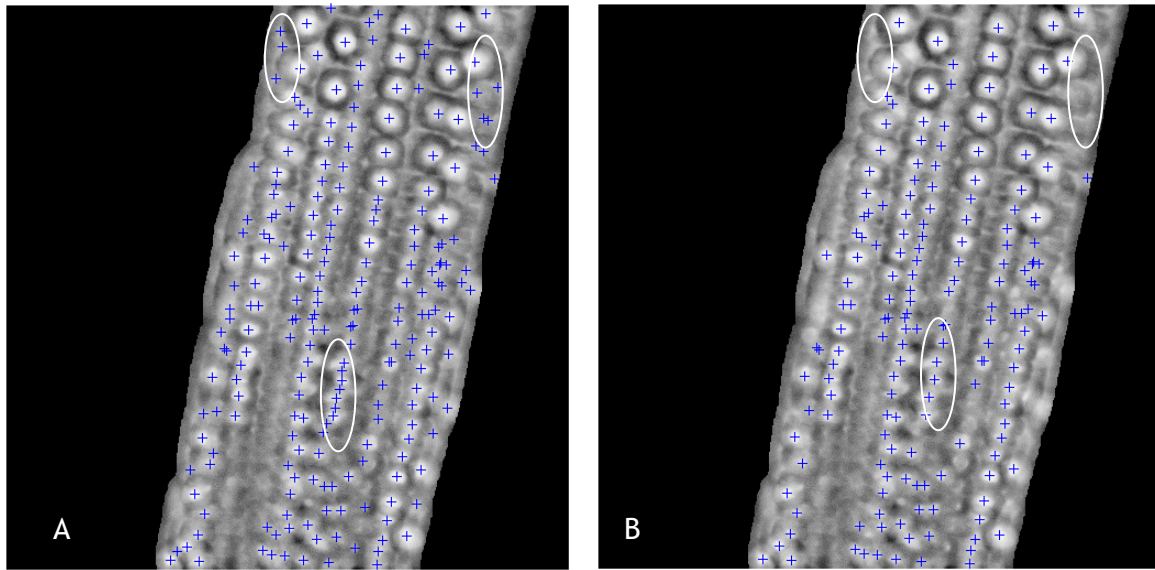


Figura 3.10 - Antes (A) e Depois (B) de introduzir a informação das paredes

Como se pode verificar nas áreas que foram assinaladas nas imagens da Figura 3.10, o número de máximos locais sobre paredes diminuiu muito, face aos resultados inicialmente obtidos.

3.4 Análise e selecção dos núcleos

Usando os máximos identificados no passo anterior, vai-se então tentar seleccionar os núcleos das células. Durante o cálculo da resposta do filtro SBF, para cada ponto é determinado qual o raio máximo utilizado em cada uma das N direcções. Os N pontos detectados por este processo são usados para estimar a equação da elipse que melhor se adapta às respectivas coordenadas [11], de forma a obter uma região que serve para delimitar uma possível localização de um núcleo. Para tal determinam-se os seguintes parâmetros da elipse:

c_x, c_y - coordenadas do ponto P correspondente ao centróide da elipse;

r_1 - eixo maior da elipse;

r_2 - eixo menor da elipse;

θ - ângulo formado pelo eixo maior e pelo eixo dos xx .

Com estes dados usa-se a equação 3.2 para definir a elipse segundo os eixos x e y .

$$\left(\frac{x-c_x}{r_1}\right)^2 + \left(\frac{y-c_y}{r_2}\right)^2 = 1, \quad (3.2)$$

Posteriormente a elipse é rodada de θ para ficar com a orientação correcta, tal como está esquematizado na Figura 3.11.

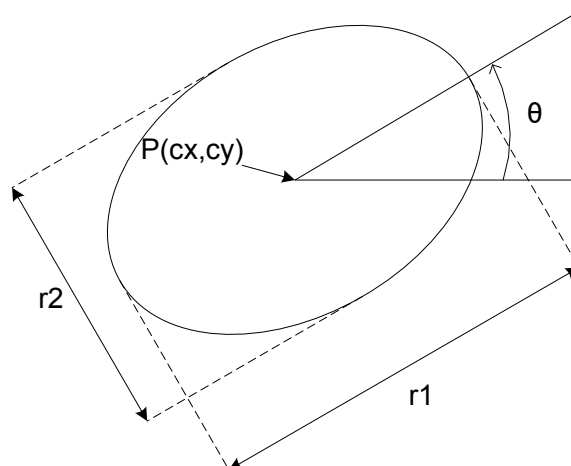


Figura 3.11 - Características da Elipse

Na Figura 3.12-A mostram-se os pontos do raio máximo detectados em cada uma das N direcções ($N=32$), como resultado da aplicação do filtro SBF. Na Figura 3.12-B apresentam-se as elipses que resultaram do processo de adaptação antes mencionado.

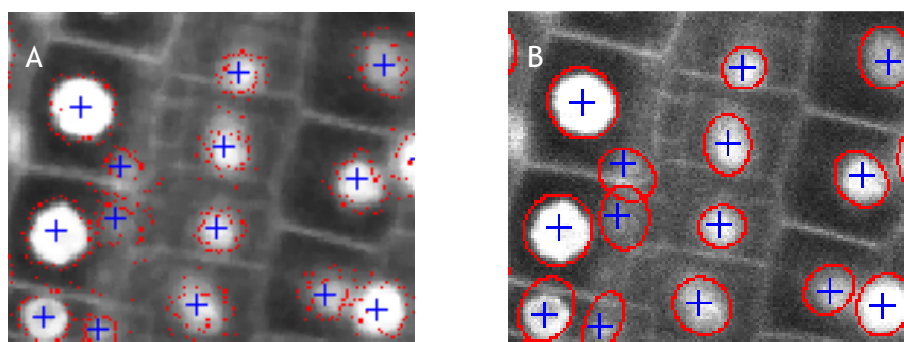


Figura 3.12 - A: Raios máximos, B: Elipses equivalentes

Como se pode constatar nestas figuras, de um modo geral verificou-se que as elipses delimitavam bastante bem as regiões onde existiam núcleos.

Inicialmente estas regiões elípticas foram usadas como máscaras para uma segmentação baseada no crescimento de regiões. No entanto, usando esse algoritmo verificou-se que os limites obtidos para os núcleos eram pouco naturais e frequentemente muito irregulares, tal como se pode verificar nalguns núcleos marcados para uma secção da imagem exemplo na Figura 3.13. Assim foi decidido efectuar a delimitação final dos núcleos usando a informação de posição a partir do filtro SBF, seguindo o procedimento de adaptação de elipses antes referido.

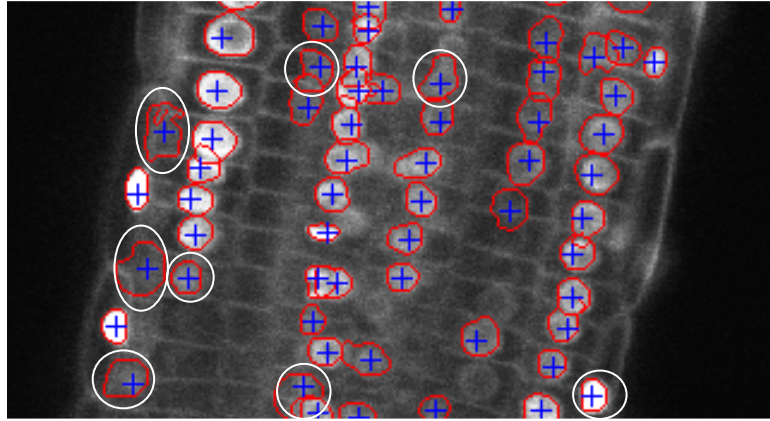


Figura 3.13 - Segmentação por crescimento de regiões

Como se pode observar na imagem da Figura 3.14 existem zonas com uma grande quantidade de potenciais núcleos, particularmente na região vascular da raiz. A elevada concentração de pontos candidatos a núcleos deve-se ao facto de que os núcleos podem existir em camadas diferentes, o que tem como consequência a sobreposição das respectivas áreas, tal como são obtidas através do processo de adaptação de elipses.

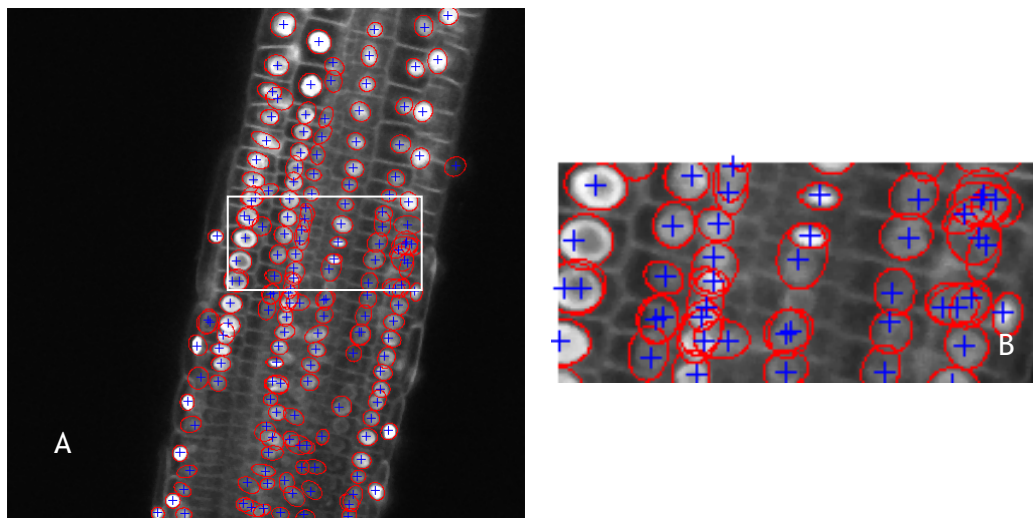


Figura 3.14 - A: Regiões marcadas utilizando as elipses calculadas a partir dos dados do Filtro SBF, B: Zoom da Região seleccionada

De forma a minimizar o problema da sobreposição das áreas dos núcleos, removem-se primeiro todos os máximos cujas regiões não intersectam outras regiões delimitadas pelo processo antes descrito. Em princípio, estas áreas correspondem a núcleos bem definidos na imagem original, e como tal não levantam problemas após a identificação da região que lhes corresponde.

Quando as regiões que se intersectam correspondem apenas a dois máximos, verifica-se que em grande parte das situações correspondem apenas a dois núcleos; a

intersecção das regiões resulta, nestes casos, do facto de algumas delas terem crescido excessivamente numa das direcções, na fase de realce. Partindo do pressuposto que para grande parte dos núcleos o máximo local detectado corresponde ao centróide do núcleo, e que os núcleos têm uma forma aproximadamente circular, a elipse é recentrada na posição do máximo e substituída por um círculo cujo diâmetro é igual ao eixo menor da elipse. Depois recalcula-se a área de intersecção, e se esta área passar a ser menor do que 50% da área do menor círculo então os dois núcleos são mantidos e a região de cada um é guardada com os novos valores; no caso contrário, isto é, se a área de intersecção for superior a 50% da área do menor círculo, é mantido o núcleo correspondente ao máximo local de maior valor e o outro é simplesmente rejeitado. Na figura 3.15 encontram-se exemplos das duas situações.

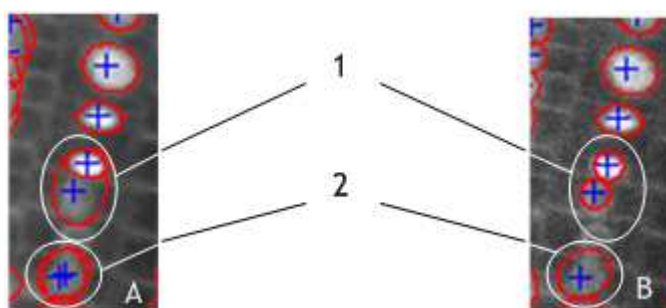


Figura 3.15 - A: Antes da tentativa de separação, B: Depois da tentativa de separação

Estes pequenos fragmentos da imagem na Figura 3.15 (A e B) contêm as duas situações descritas anteriormente: no primeiro caso (1) a separação é bem conseguida e os dois núcleos mantêm-se; no segundo caso (2), como não é possível manter os dois núcleos, é escolhido o que tem maior resposta na imagem realçada.

No caso das regiões com mais do que dois núcleos, devido geralmente à grande quantidade de máximos nessas regiões, opta-se por utilizar um algoritmo diferente e que se detalha no esquema da Figura 3.16. Neste caso, e para garantir que os núcleos bem identificados na imagem original se mantêm, é calculada a média geométrica entre o valor realçado e o valor original. Posteriormente, ordenam-se os núcleos por ordem decrescente de valor da média geométrica, e a partir do núcleo de maior média para o de menor média eliminam-se os máximos locais cujas áreas associadas intersectam a área do máximo local actualmente em análise (máximo de referência) em mais que 50%.

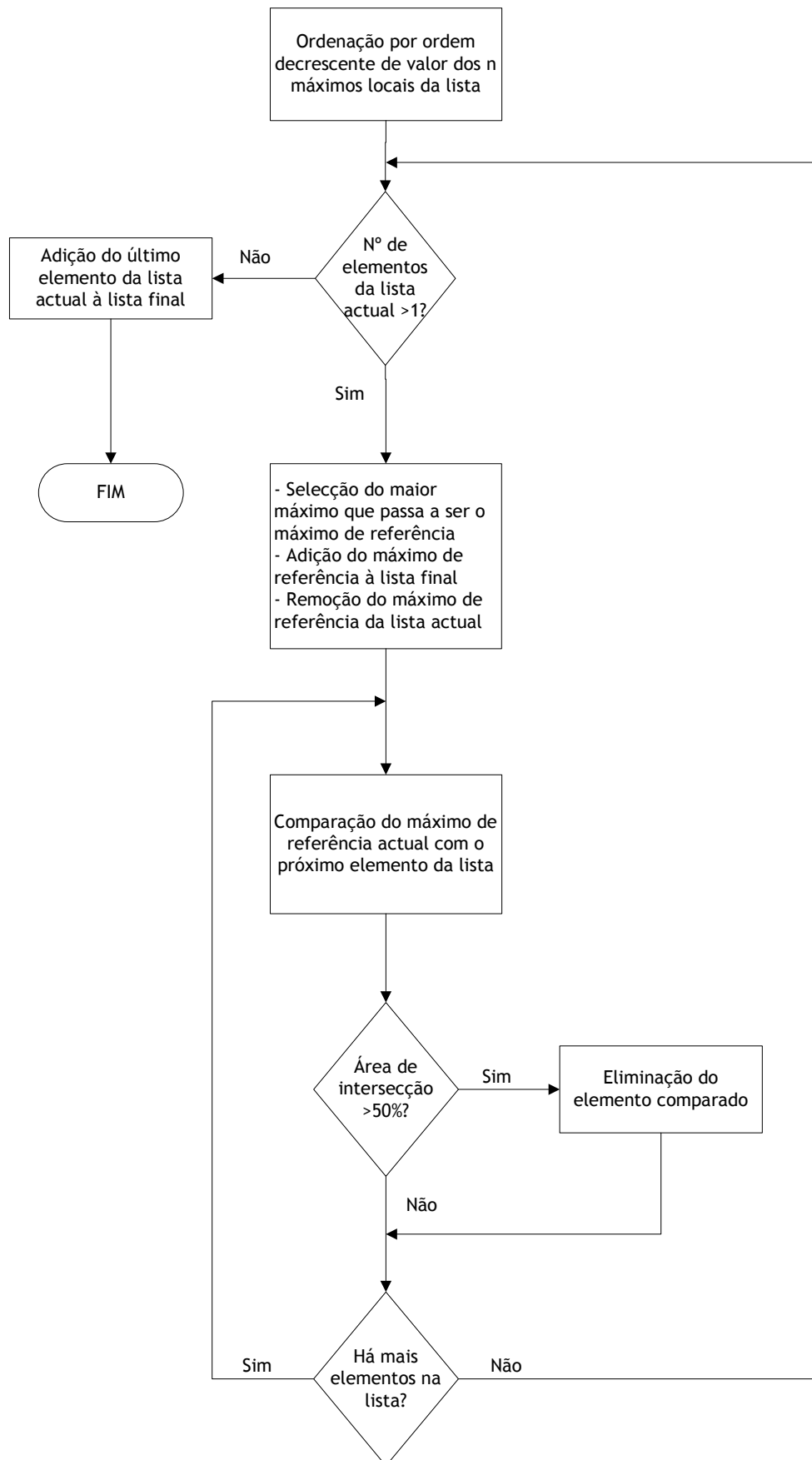


Figura 3.16 - Esquema do algoritmo de selecção

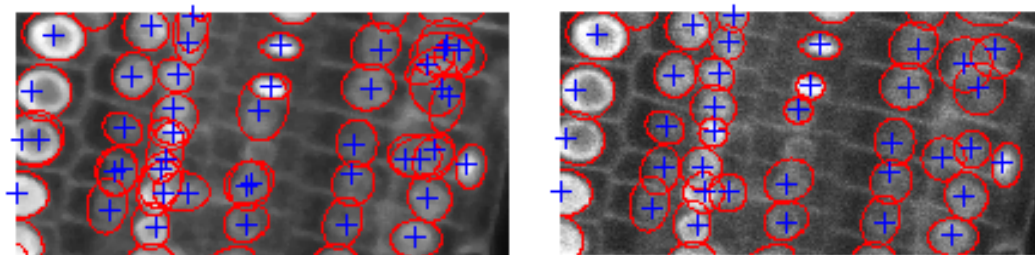


Figura 3.17 - Antes (A) e depois (B) de aplicar o algoritmo de selecção

Como se pode verificar na Figura 3.17, após a aplicação do algoritmo de selecção dos núcleos esquematizado na Figura 3.16, já não existem zonas de grande concentração e sobreposição, e as áreas seleccionadas correspondem em grande parte aos locais onde efectivamente se encontram os núcleos.

Os resultados dos diversos passos da metodologia, aplicados a uma pequena secção da imagem, são ilustrados na Figura 3.18.

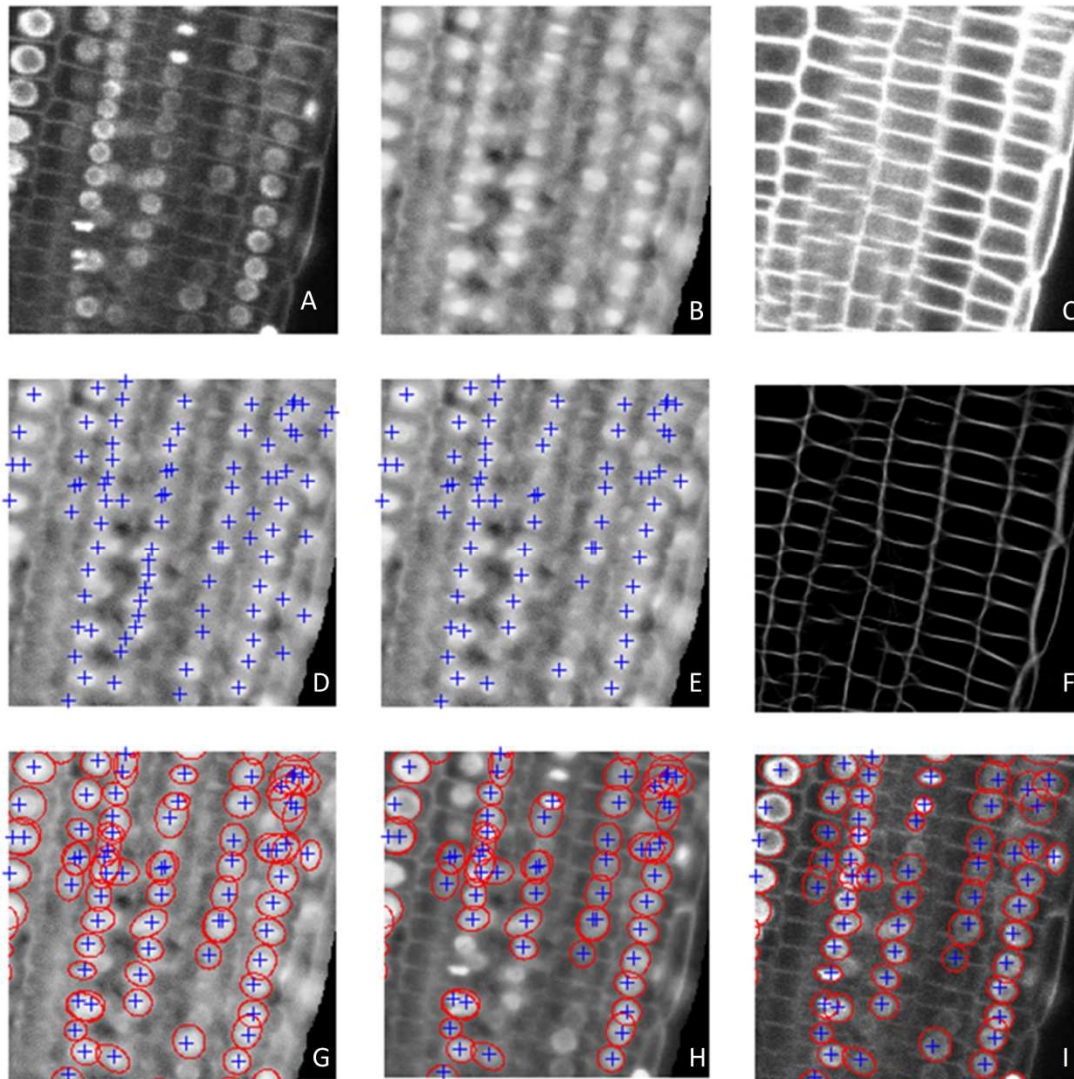


Figura 3.18 - A: Imagem original. B: Imagem realçada. C: Imagem das paredes. D: Máximos locais da imagem realçada. E: Máximos locais tendo em conta a informação das paredes. F: Simetria de fase da imagem das paredes. G: Zonas dos máximos marcadas na imagem realçada. H: Zonas dos máximos que têm intersecções marcadas na imagem da média geométrica entre a imagem original e a imagem realçada. I: Imagem Original com as zonas seleccionadas marcadas.

O resultado final obtido pela aplicação de toda a metodologia descrita neste capítulo para a imagem completa da Figura 3.2 é apresentado na figura 3.19.

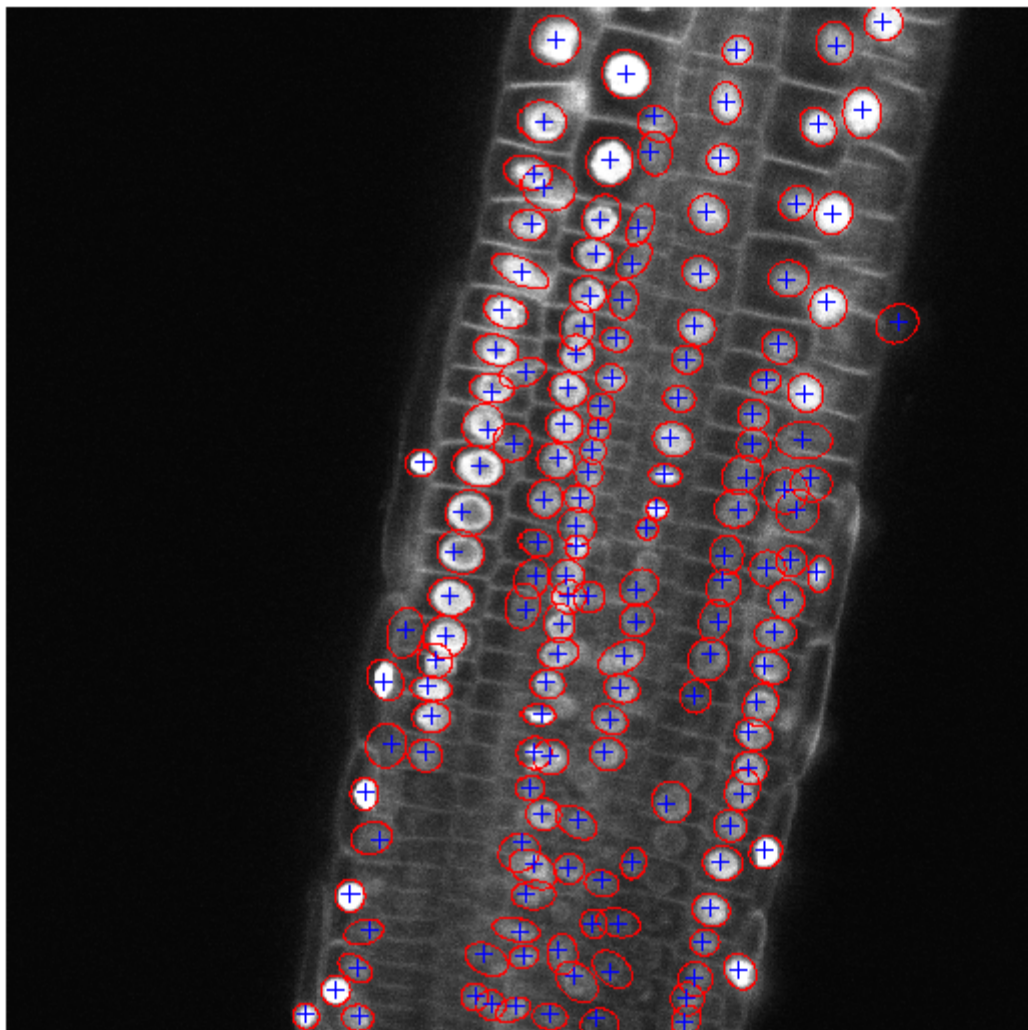


Figura 3.19 - Resultado Final da aplicação da metodologia

3.5 Conclusão

Com esta metodologia pretendeu-se encontrar um método relativamente simples para detectar os núcleos das células em imagens de microscopia confocal da raiz da *Arabidopsis Thaliana*.

Começou por filtrar a imagem original onde os núcleos são visíveis de modo a realçar as regiões que tinham uma forma arredondada. Seguidamente, através da determinação dos máximos locais da imagem realçada, foram identificadas as possíveis localizações dos núcleos.

A partir do resultado do filtro de realce, designadamente da informação da respectiva região de suporte que se adapta naturalmente às características de intensidade da imagem, obtiveram-se elipses para descrever as zonas onde potencialmente os núcleos poderiam existir.

Posteriormente, nas zonas de grande sobreposição de áreas, ordenaram-se os máximos locais dos potenciais núcleos segundo a média geométrica entre o valor da sua intensidade da imagem original e o seu valor na imagem realçada. Depois, partindo do ponto com maior média para o de menor média, eliminaram-se as zonas que intersectavam a área do máximo local em estudo em mais do que 50%.

O procedimento proposto fica concluído após a avaliação e validação de todos os máximos locais.

Capítulo 4

Resultados

4.1 Introdução

Neste capítulo vão ser apresentados os resultados da metodologia descrita no capítulo anterior. Em primeiro lugar vai ser efectuada uma comparação entre os diferentes métodos de realce testados; depois vão ser descritos e mostrados os resultados da aplicação da metodologia a várias imagens diferentes daquelas que foram usadas durante a fase de desenvolvimento. Para testar o método descrito no Capítulo 3 foram usadas 15 imagens diferentes das que foram utilizadas durante o desenvolvimento.

Para ilustrar os resultados vão ser usadas 4 imagens das 15 de teste, que se mostram na Figura 4.1.

Todas as imagens têm dimensão 512x512 pixéis, e estão representadas numa escala de 256 níveis de cinzento. Durante a aplicação desta metodologia vão ser usados, para todas as imagens, os seguintes parâmetros para o filtro de realce:

- $R_{\max}=16$;
- $R_{\min}=2$;
- $N=32$ (Número de direcções usadas na filtragem);
- $d=4$ (largura da coroa).

O limiar usado para garantir que os máximos locais da imagem realçada estão bem definidos na imagem original é de 0.85.

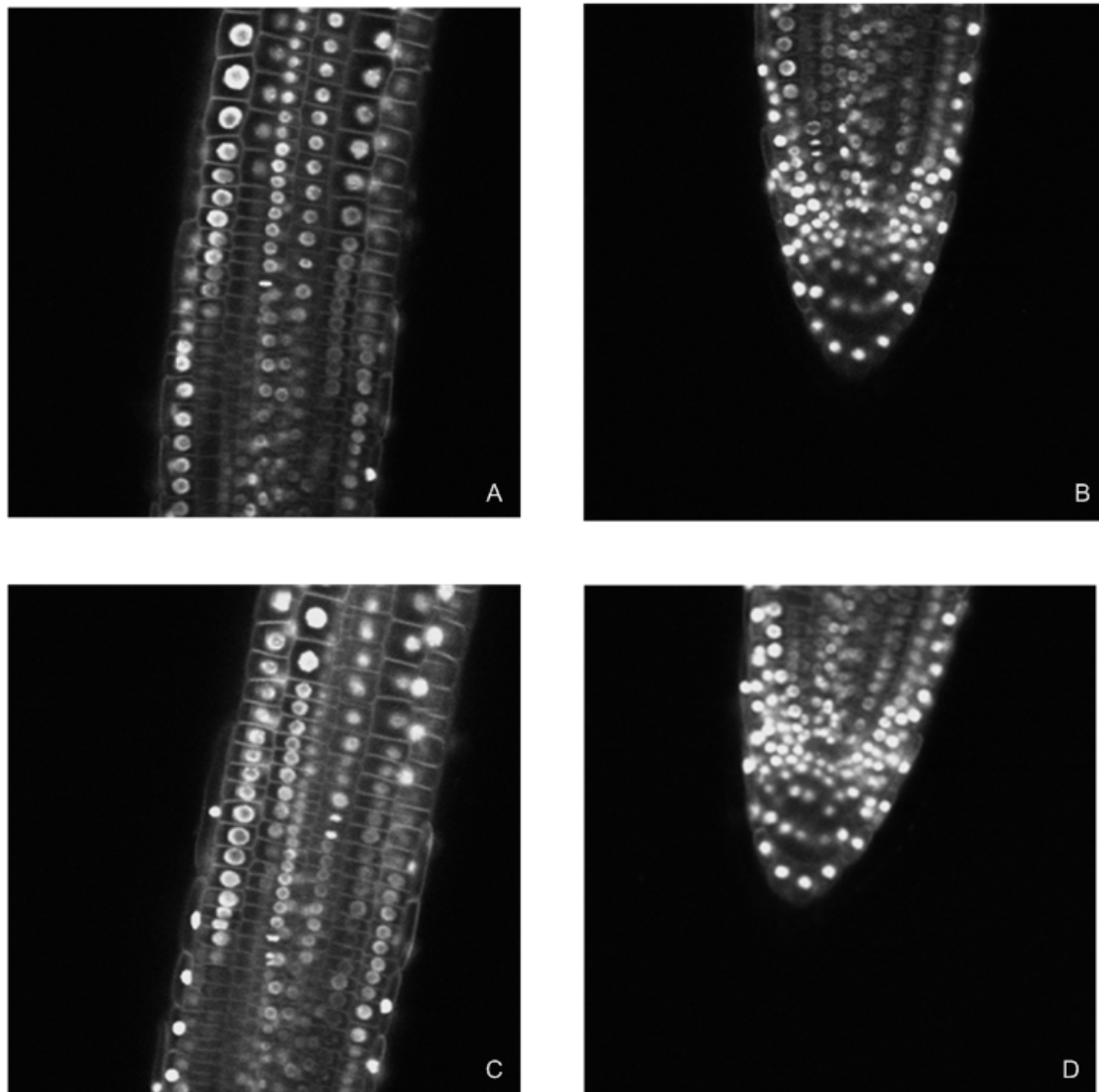


Figura 4.1 - Imagens Originais

Durante a apresentação dos resultados irão ser mostrados os extractos de cada imagem que melhor ajuda a demonstrar a eficácia do método proposto. Os resultados correspondentes a cada imagem estarão representados com a letra que identifica a imagem na Figura 4.1.

4.2 Realce da Imagem

O primeiro passo da metodologia descrita neste trabalho é o realce da imagem. Uma vez que os objectos a serem salientados têm uma forma arredondada, decidiu-se usar um filtro de convergência. Foram testados o filtro de moeda, o filtro de anel adaptativo, o filtro de íris e o filtro de banda deslizante. Os filtros foram aplicados à imagem exemplo do Capítulo 3. Os resultados obtidos para uma secção da imagem exemplo de cada um dos filtros são apresentados na Figura 4.2.

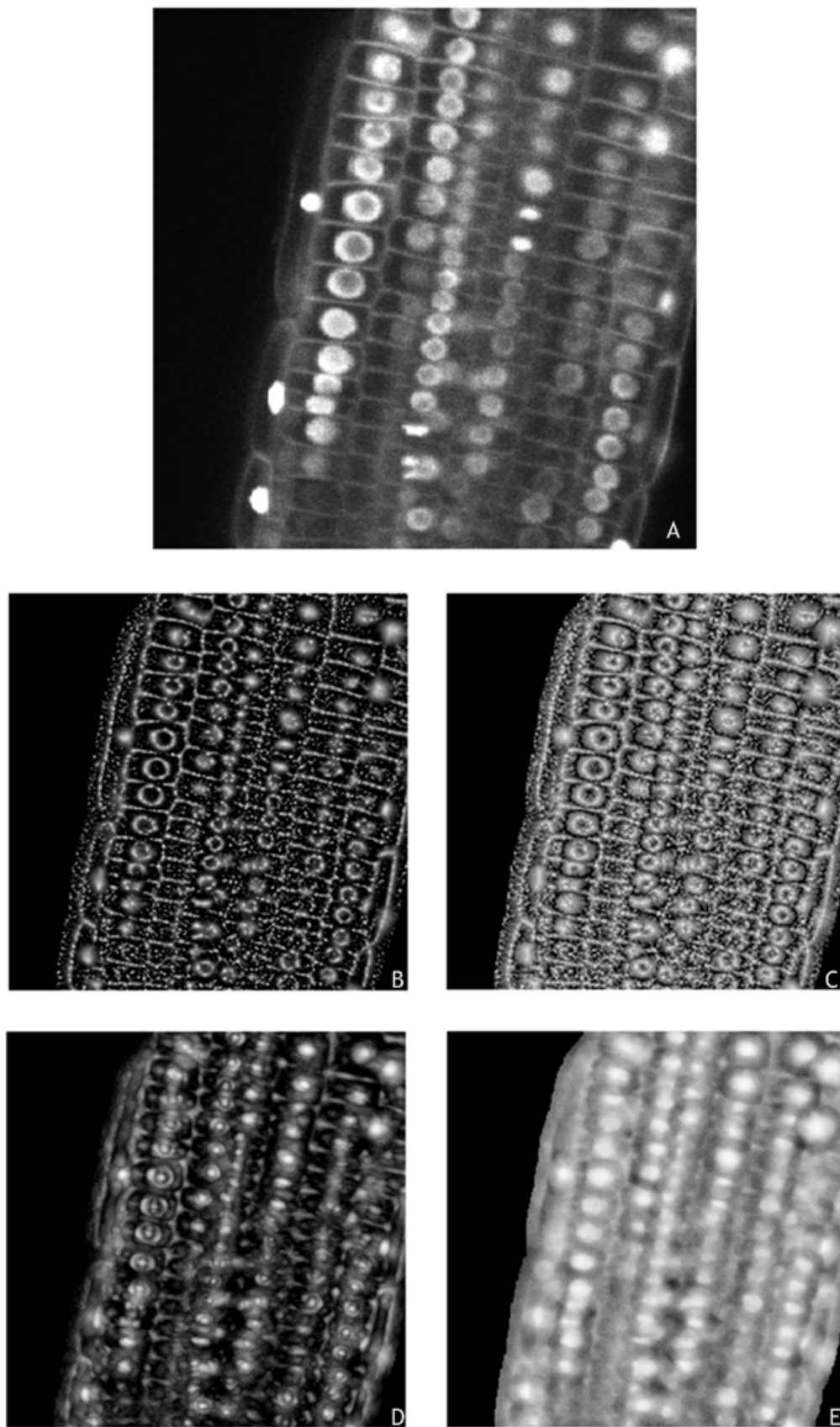


Figura 4.2 - A: Imagem original, B: Filtro de Moeda, C: Filtro de Íris, D: Filtro de anel adaptativo, E: Filtro SBF

Na Figura 4.2, no caso B e C, verifica-se que as zonas mais escuras no interior de cada um dos núcleos continua a existir, e para além disso, as paredes das células também são realçadas. No caso D (Filtro de Anel Adaptativo), nos núcleos com forma aproximadamente circular a resposta ao filtro é muito boa, incluindo no interior; no entanto, quando os núcleos têm uma forma mais elíptica os resultados já não são tão bons. Com o filtro SBF (E), pode-se observar uma melhor homogeneidade da resposta em toda a área do núcleo, incluindo as regiões internas originalmente menos intensas, que na imagem realçada possuem valores semelhantes aos que são obtidos na periferia.

Uma vez que a forma de detecção das possíveis localizações dos núcleos será realizada através dos máximos locais da imagem, na Figura 4.3 assinalam-se os máximos locais para cada uma das imagens realçadas pelos filtros anteriores.

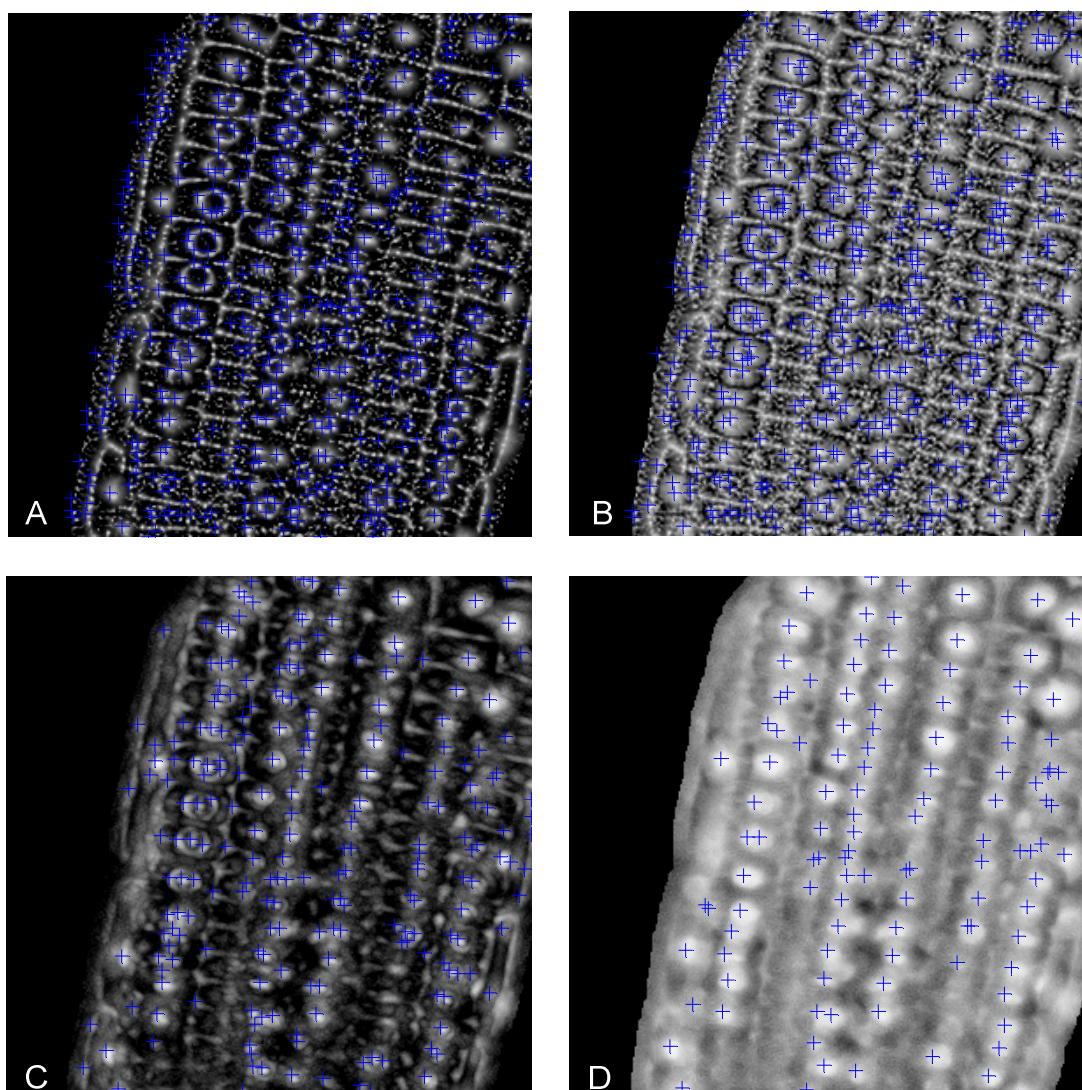


Figura 4.3 - Máximos locais de: A- Filtro de Moeda; B- Filtro de Íris; C- Filtro em anel adaptativo; D- Filtro SBF

Como se pode observar pelos resultados da Figura 4.3, os filtros de moeda (A) e de íris (B) possuem máximos locais nos núcleos, mas também possuem muitos máximos locais que não pertencem a qualquer núcleo. No caso do filtro de anel adaptativo (C), o número de máximos locais é muito inferior aos outros dois filtros; no entanto, a muitos dos núcleos estão associados mais do que dois máximos. Comparativamente o filtro SBF (D) é o que possui o menor número de máximos locais além de associar apenas um máximo local a maior parte dos núcleos.

4.3 Identificação dos Núcleos

O segundo passo do algoritmo proposto trata da localização e selecção dos possíveis núcleos. Para tal, começa-se por determinar os máximos locais na imagem realçada com o filtro SBF, que se mostra nos diversos exemplos da Figura 4.4.

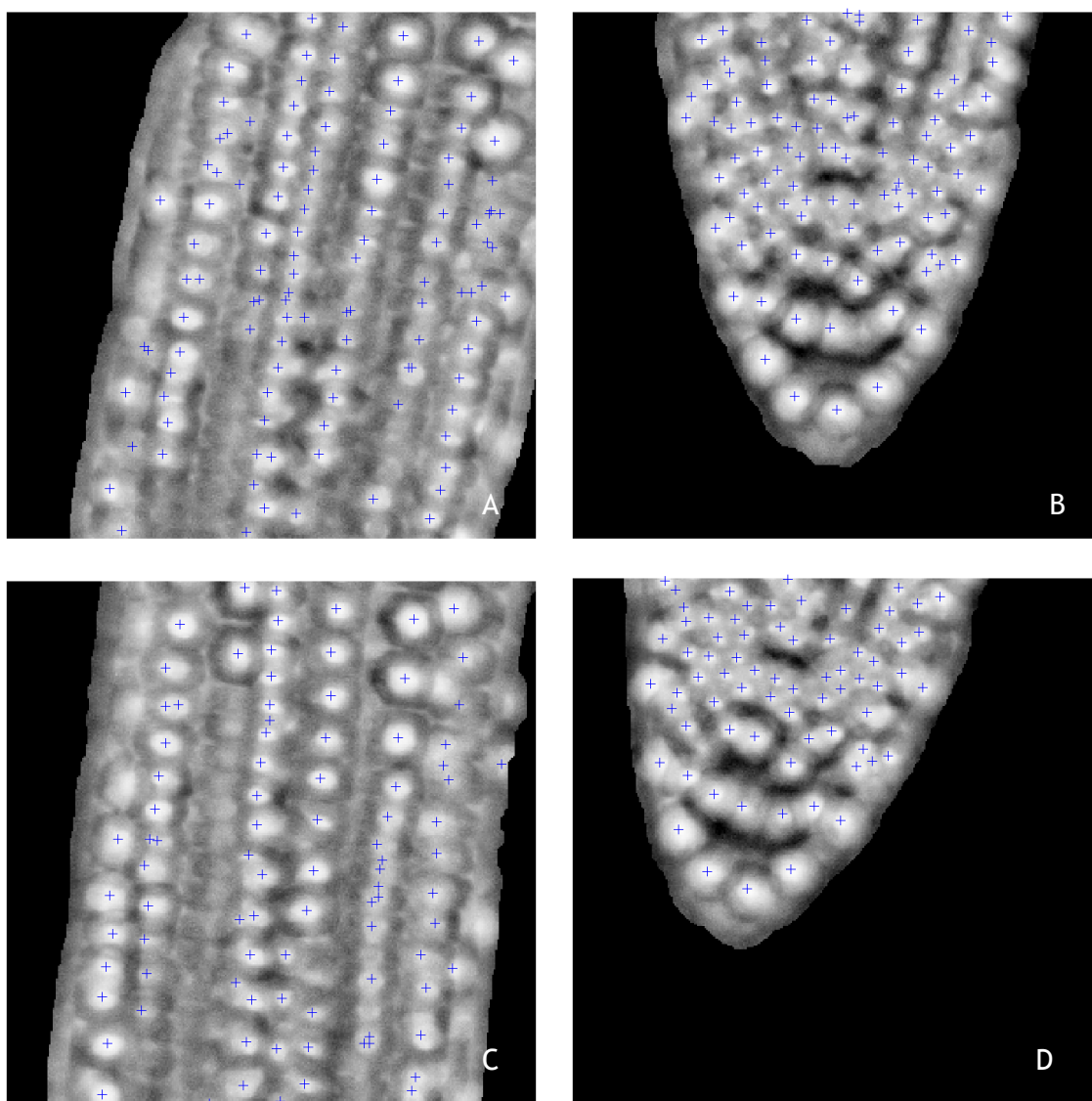


Figura 4.4 - Máximos Locais

Os resultados obtidos para esta etapa do processo foram conseguidos usando limiares únicos automaticamente determinados para cada imagem. Comparando com as imagens iniciais, pode-se observar que os máximos identificados se aproximam muito dos locais onde efectivamente se podem encontrar os núcleos.

No passo seguinte, procede-se à determinação dos pontos seguindo-se a estimação da elipse que melhor se adapta a cada conjunto de N pontos associados a um máximo local.

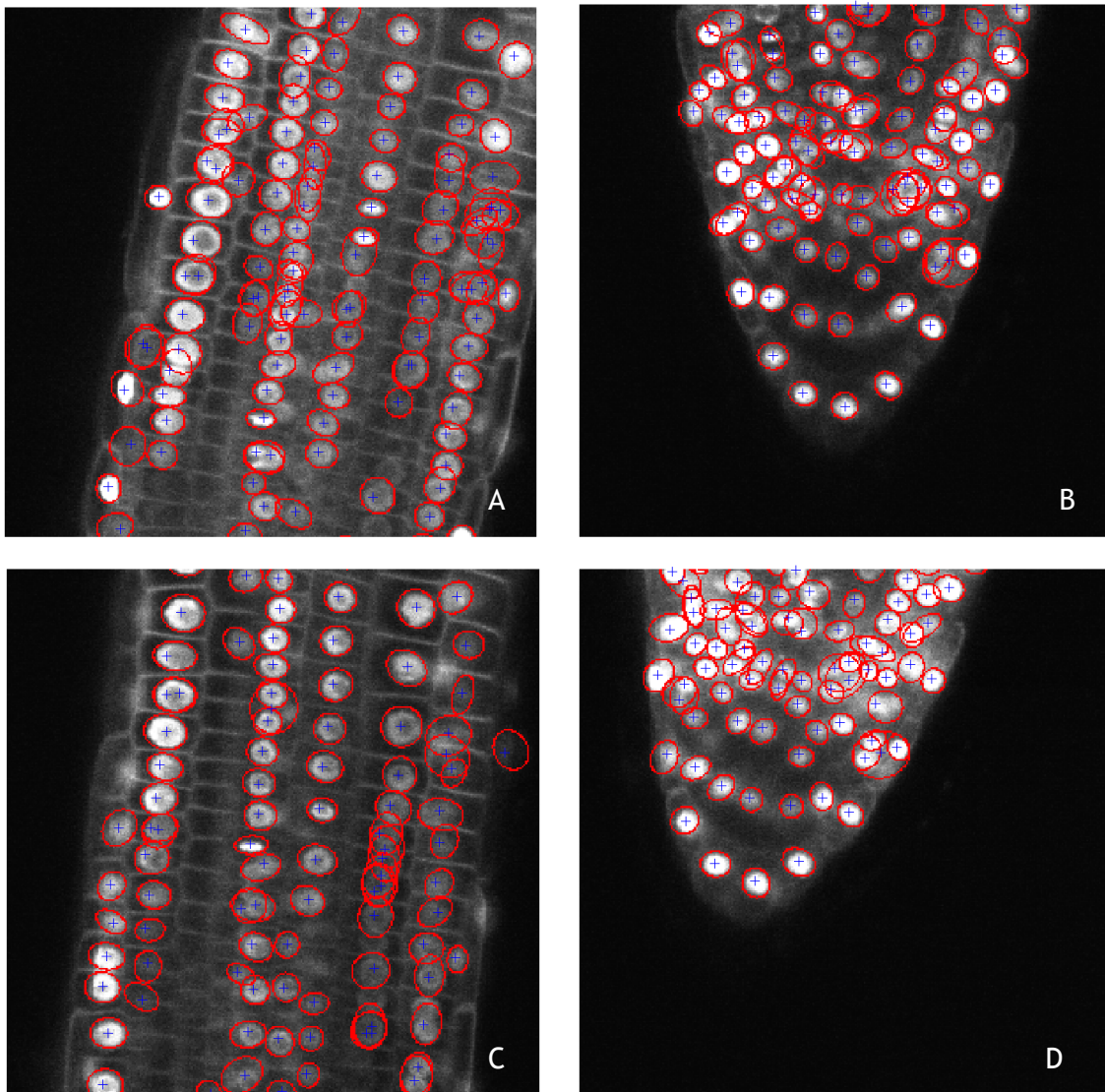


Figura 4.5 - Elipses equivalentes

As elipses assim estimadas são usadas para delinear as áreas de potencial existência de núcleos. Nas imagens da Figura 4.5 pode-se observar que as elipses delimitam relativamente bem as regiões onde estão os núcleos, apesar de ainda existirem muitas regiões com áreas sobrepostas.

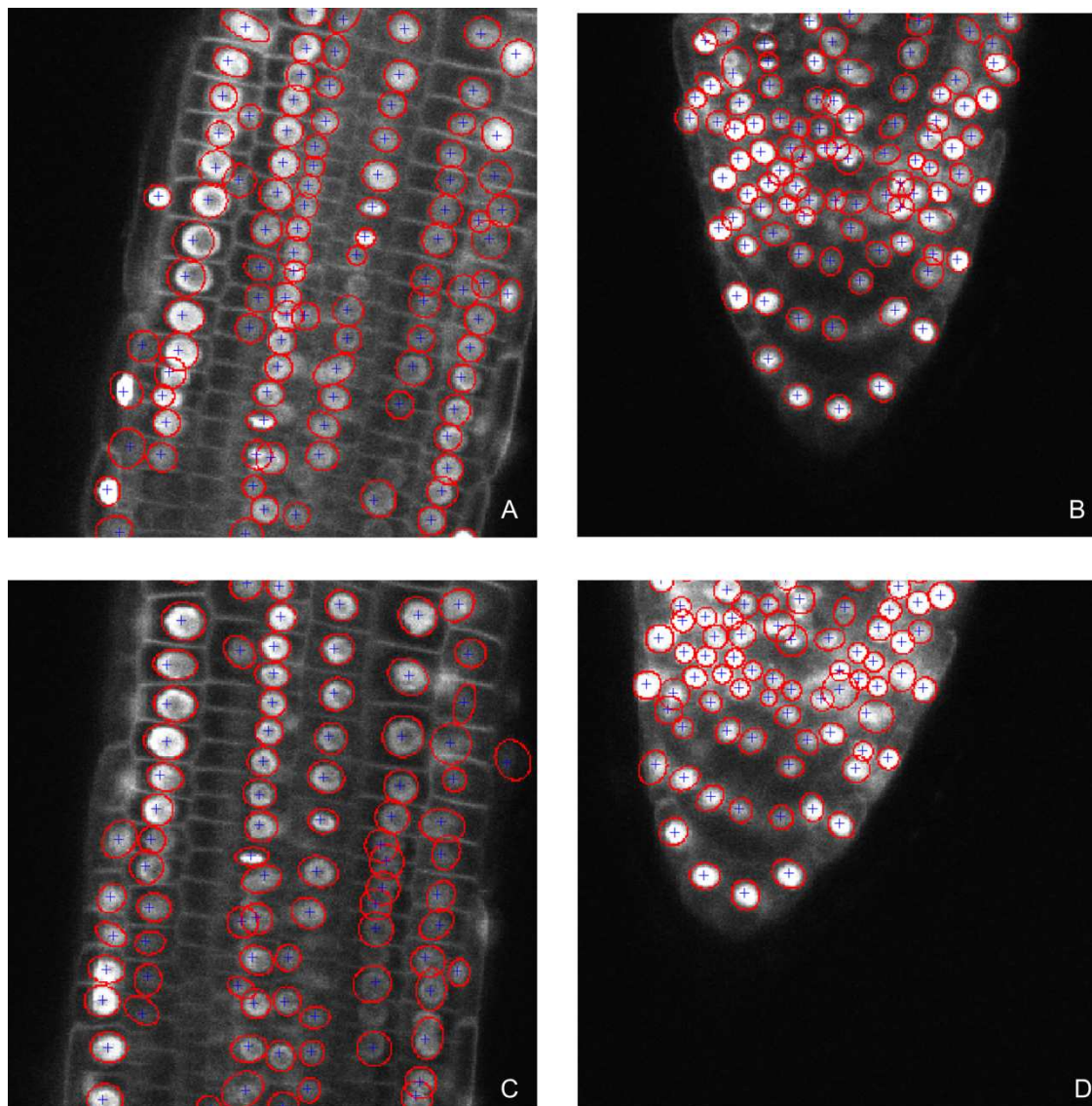


Figura 4.6 - Imagens Finais

Como se pode observar pelas imagens da Figura 4.6, após a selecção das elipses, as áreas resultantes correspondem com bastante sucesso às áreas onde se podem encontrar os núcleos e também à forma dos núcleos. No entanto ainda existem algumas zonas sobrepostas, possivelmente devido à influência dos núcleos que se encontram em camadas diferentes.

4.4 Conclusão

Neste capítulo foram mostrados os resultados produzidos pelo método proposto no Capítulo 3, aplicado a 4 imagens, de forma a ilustrar o que havia sido descrito anteriormente. Começou-se por comparar os vários filtros de convergência para realçar as imagens de estudo, apresentando as razões da escolha do filtro SBF.

Depois foram mostrados os resultados de cada passo do processo da selecção dos núcleos das células nomeadamente a determinação dos máximos locais, a adaptação de elipses aos pontos do contorno que resultam do processo de filtragem, de forma a identificar possíveis regiões onde se encontram os núcleos e finalmente a selecção dos locais onde se podem encontrar os núcleos.

Capítulo 5

Conclusão

5.1 Conclusão

Neste trabalho foi proposto um método para delimitar automaticamente os núcleos das células de imagens, de microscopia confocal, de secções da raiz da planta *Arabidopsis Thaliana*.

O primeiro passo do algoritmo descrito no Capítulo 3, consistia em realizar uma filtragem com vista a realçar os núcleos. Uma vez que os núcleos têm uma forma arredondada foram testados quatro filtros de convergência de modo a encontrar qual o filtro que melhor se adequava às imagens de estudo. Desta comparação concluiu-se que o filtro de banda deslizante (SBF) era o que obtinha os melhores resultados, pois conseguia ajustar-se mais eficazmente aos núcleos e homogeneizar o contraste entre o interior e o exterior dos mesmos, que consistia no principal problema que não permitindo o uso dos máximos locais da imagem como forma de detecção dos possíveis locais para os núcleos.

Posteriormente, usando os pontos do contorno da região de suporte resultante durante a filtragem, foram geradas elipses que se adaptavam a essas regiões, com a finalidade de identificar as áreas onde os núcleos poderiam estar localizados.

Finalmente, nos locais onde havia uma grande concentração de regiões sobrepostas, seleccionaram-se os núcleos com base na sua média geométrica, calculada entre a imagem original e a imagem realçada. Partindo do ponto com maior média removeram-se os pontos que produziam áreas de intersecção superiores a 50% da área do núcleo em estudo.

Verificou-se que o algoritmo desenvolvido permitia a obtenção de resultados bastante satisfatórios na delimitação das regiões onde se encontram os núcleos.

Os erros encontrados na identificação dos núcleos através deste método devem-se sobretudo às zonas onde existem sobreposição de núcleos de camadas diferentes e às paredes das células que se encontram bem definidas em muitas das imagens que foram disponibilizadas para este trabalho.

Referências

- [1] Hidefumi Kobatake e Shigeru Hashimoto, "Convergence Index Filter for Vector Fields", *IEEE Transactions on Image Processing*, Vol 8, N°8, Agosto 1999
- [2] Carlos S. Pereira, Hugo Fernandes, Ana Maria Mendonça e Aurélio Campilho, "Detection of Lung Nodule Candidates in Chest Radiographs", *IbPRIA 2007*, 170-177, 2007
- [3] Gerhard Zinser e Dymitr Komitowski, "Segmentation of Cell Nuclei in Tissue Section Analysis", *The Journal of Histochemistry and Cytochemistry*, Vol 31, nº1, 94-100, 1983.
- [4] Norberto Malpica, Carlos Ortiz de Solórzano, Juan José Vaquero, Andrés Santos, Isabel Vallcorba, José Miguel García-Sagredo e Francisco del Pozo, "Applying Watershed Algorithms to the Segmentation of Clustered Nuclei", *Cytometry*, Vol 28, Issue 4, 289-297, 1997.
- [5] T. Mouroutis, S. J. Roberts, A. A. Bharath, "Robust Cell Nuclei Segmentation Using Statistical Modelling", *Bioimaging*, Vol 6(2), 79-91, 1998.
- [6] Faguo Yang e Tianzi Jiang, "Cell Image Segmentation with Kernel-Based Dynamic Clustering and an Ellipsoidal Cell Shape Model", *Journal of Biomedical Informatics*, Vol 34, 67-73, 2001.
- [7] C. Wählby, I.-M. Sintorn, F. Erlandsson, G. Borgefors e E. Bengtsson, "Combining intensity, edge and shape information for 2D and 3D segmentation of cell nuclei in tissue sections", *Journal of Microscopy*, Vol 215, 67-76, Julho 2004.
- [8] Xiaowei Chen, Xiaobo Zhou, Stephen T.C Wong, "Automatic Segmentation, Classification, and Tracking of Cancer Cell Nuclei in Time-Lapse Microscopy", *IEEE Transactions on Biomedical Engineering*, Vol 53, nº 4, Abril 2006.
- [9] Xiaodong Yang, Houqiang Li e Xiaobo Zhou, "Nuclei Segmentation Using Marker-Controlled Watershed, Tracking Using Mean-Shift, and Kalman Filter in

Time-Lapse Microscopy”, *IEEE Transactions on Circuits and Systems - I: Regular Papers*, Vol 53, nº 11, Novembro 2006.

- [10] Peter Kovesi, “MATLAB and Octave Functions for Computer Vision and Image Processing”,
<http://www.csse.uwa.edu.au/~pk/Research/MatlabFns/index.html>, Acesso em 14 de Junho de 2008.
- [11] CVonline: The Evolving, Distributed, Non-Proprietary, On-Line Compendium of Computer Vision, <http://homepages.inf.ed.ac.uk/rbf/CVonline/>, Acesso em 14 de Junho de 2008.