

# Detecção de Canaletas em Imagens de Géis de Eletroforese utilizando Filtro de Kalman

Alessandra A. Paulino\*<sup>†</sup>, Maurílio Boaventura\* ,

Depto de Ciências de Computação e Estatística, IBILCE, UNESP,  
São José do Rio Preto, SP

E-mail: alessandra.paulino@gmail.com, maurilio@ibilce.unesp.br,

Carlos D. Maciel\*, Vilma A. de Oliveira\*.

Depto de Engenharia Elétrica, EESC, USP,

São Carlos, SP

E-mail: maciel@sel.eesc.usp.br, vilmao@sel.eesc.usp.br

**Resumo:** O objetivo deste trabalho é apresentar um método automático e eficiente para detecção de canaletas em imagens de géis de eletroforese utilizando projeções e filtro de Kalman.

## 1 Introdução

O estudo da genética de populações tem como um dos seus objetivos estabelecer relações de semelhança entre indivíduos ou identificação de grupos com as mesmas características. Analisando o DNA ou RNA de uma amostra é possível encontrar padrões genéticos similares, os quais podem dar suporte à inclusão de um indivíduo em um grupo de características conhecidas. Uma das formas de análise é com a técnica de eletroforese.

Eletroforese é um processo de separação eletroquímica no qual moléculas biológicas tais como proteínas ou fragmentos de DNA e RNA são impelidos a migrar ao longo de uma substância específica (gel de agarose ou gel de poliacrilamida) sob a influência de um campo elétrico. Por exemplo no caso do DNA, de forma mais detalhada, uma molécula deste é quebrada em muitos fragmentos pela ação de enzimas específicas. Estes fragmentos são dispersos em um gel de poliacrilamida ou um gel de agarose, no qual um campo elétrico é aplicado. Cada fragmento tem carga elétrica e massa molecular distintas, o que faz com que elas fiquem dispostas em diferentes posições ao

longo do gel. Os fragmentos de moléculas com menor peso ou maior carga se movem mais rapidamente que os fragmentos com maior peso ou menor carga. Transcorrido um período de tempo o processo é interrompido e o gel passa por uma coloração, o que possibilita a observação da posição em que a molécula parou. Um exemplo de imagem de gel de eletroforese é mostrado na Figura 1.

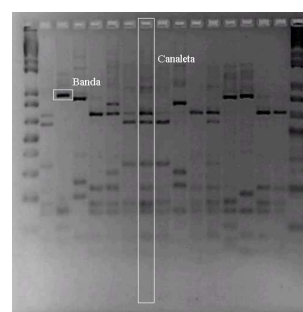


Figura 1: Imagem de um gel de eletroforese

Os corredores verticais são chamados de canaletas. O material contido nas canaletas, que são as posições nas quais as moléculas param, é chamado de banda.

O processo de eletroforese é bastante aplicado em seqüenciamento de DNA e em estudos de variação qualitativa e quantitativa da separação de proteínas ou ácidos nucléicos obtidos de diferentes fontes. Cientistas utilizam eletroforese para obter informações de substâncias em estudo, bem como uma comparação da composição de amostras, ou quantificação do montante e propriedades dos diferentes elementos presentes em uma coleção de

\*Beneficiário de auxílio financeiro da CAPES - Brasil

<sup>†</sup>Bolsista de Mestrado CNPq - Brasil

amostras.

A comparação entre duas canaletas em uma imagem de gel de eletroforese é geralmente um processo complexo devido a subjetividade da percepção visual humana. Baseados no mesmo material duas pessoas podem chegar a conclusões diferentes. Com o progresso das técnicas de análise de imagens, o computador está se tornando uma ferramenta importante para cientistas e profissionais na aquisição e interpretação de dados. Assim, para analisar e comparar canaletas computacionalmente precisa-se que essas canaletas sejam identificadas. A identificação dessas é um processo facilmente executável por um especialista. Ao analisar poucas imagens com 15 ou 20 canaletas, pode-se separá-las manualmente utilizando algum editor de imagens. Quando o número de imagens cresce, essa solução manual se torna um problema, tanto do ponto de vista do trabalho quanto do tempo utilizado para sua execução.

O objetivo principal deste trabalho é automatizar essa detecção e identificar as canaletas de forma eficiente para facilitar a comparação de forma automática. Neste trabalho utilizam-se imagens com e sem pré-processamento para retirada de ruído de fundo. Para o pré-processamento aplicou-se um algoritmo baseado em equações diferenciais parciais não lineares proposta por Barcelos, Boaventura e Silva Jr [1] para a eliminação de ruído. Ainda na fase de pré-processamento retirou-se o fundo da imagem utilizando Transformação Top-hat por fechamento mostrada em [4].

Neste trabalho apresenta-se uma implementação de três algoritmos para detecção de canaletas em imagens de géis de eletroforese propostos por [3], assim como alguns resultados obtidos pela implementação dos procedimentos de pré-processamento e de detecção de canaletas.

## 2 Pré-Processamento

O primeiro passo no pré-processamento de imagens de géis de eletroforese foi o de eliminar o ruído através de suavização matemática. O processo de suavização utilizado baseia-se no conceito matemático de imagem, que pode ser definida como sendo uma função  $I$  dada em cada ponto  $x \in \mathbb{R}^p$ ,  $p = 2$  ou  $3$ . Dada uma

imagem  $I(x)$ , um dos problemas mais comuns é remover interferências indesejadas. Seja  $I$  uma imagem observada e  $u$  a imagem reconstruída. Estas funções podem ser definidas como funções de  $\Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ , que associa a cada pixel  $(x, y) \in \mathbb{R}^2$  seu nível de cinza  $I(x, y)$  ou  $u(x, y)$ .

Em [1] foi proposta a seguinte equação diferencial parcial parabólica para restauração de imagens:

$$\begin{aligned} u_t &= g|\nabla u| \operatorname{div} \left( \frac{\nabla u}{|\nabla u|} \right) - \lambda(1-g)(u-I), \\ &\quad x \in \Omega, \quad t > 0, \\ u(x, y, 0) &= I(x, y), \quad x \in \Omega \\ \left( \frac{\partial u}{\partial n} \right) |_{\partial \Omega \times \mathbb{R}_+} &= 0, \quad x \in \partial \Omega, \quad t > 0 \end{aligned}$$

onde  $I(x, y)$  é uma imagem a ser processada,  $u(x, y, t)$  é sua versão suavizada na escala  $t$ ,  $G_\sigma$  é o núcleo da convolução, e  $(|G_\sigma * \nabla u|)$  é a estimativa local de  $\nabla u$ . A função  $g(s) \geq 0$  é uma função não crescente satisfazendo  $g(0) = 1$  e  $g(s) \rightarrow 0$  quando  $s \rightarrow \infty$ . Neste trabalho foi considerada a função  $g(s) = \frac{1}{1+Ks^2}$ , onde  $s = |G_\sigma * \nabla u|$ . Observa-se que, nas regiões homogêneas da imagem, tem-se  $|\nabla G_\sigma * u|$  pequeno o que implica em  $g \sim 1$ , por outro lado, para as regiões de contorno onde  $|\nabla G_\sigma * u|$  é grande, tem-se  $g \sim 0$ , e, desta forma, tem-se que a função  $g$  funciona como um detector de bordas. A equação apresentada anteriormente mostrou ser bastante eficiente no processo de restauração de imagens (eliminação de ruídos e segmentação), sem entretanto perder bordas e, portanto, preservando as características originais da imagem. Isso justifica a escolha por esse algoritmo.

A Transformação Top-hat, mostrada em [4], também utiliza ferramentas matemáticas. A partir dos operadores básicos da Morfologia Matemática (erosão e dilatação), define-se o Gradiente Morfológico por dilatação e erosão ( $\rho$ ) como a diferença aritmética entre a dilatação e erosão de uma imagem  $I$  por um elemento estruturante  $B$ :

$$\rho(I) = \delta_B(I) - \varepsilon_B(I)$$

onde:

- $\rho$  representa o gradiente por dilatação e erosão;
- $\delta_B$  é a dilatação pelo elemento estruturante  $B$ ;
- $\varepsilon_B$  é a erosão pelo elemento estruturante  $B$ .

O gradiente morfológico é invariante à operação de complementação, sendo, portanto uma transformação auto-complementar.

A abertura  $\gamma$  de uma imagem  $I$  por um elemento estruturante  $B$ , é definida pela operação de erosão de  $I$  por  $B$ , seguido pela operação de dilatação com elemento estruturante transposto  $B'$ , a qual é determinada pela expressão

$$\gamma_B(I) = \delta_{B'}[\varepsilon_B(I)].$$

O fechamento  $\phi$  de uma imagem  $I$  por um elemento estruturante  $B$ , é definido pela operação de dilatação de  $I$  por  $B$ , seguido pela operação de erosão com elemento estruturante transposto  $B'$ , o qual é determinado pela expressão

$$\phi_B(I) = \varepsilon_{B'}[\delta_B(I)].$$

Com esses conceitos pode-se introduzir a Transformação Top-hat. Uma forma de considerar a detecção de bordas consiste em usar uma combinação entre uma imagem original e a imagem correspondente aberta (operação de abertura), ou usar a imagem fechada (operação de fechamento) e a imagem original. Dessas combinações são elaborados os top-hats. O top-hat por abertura de uma imagem  $I$  é definido como a diferença aritmética entre a imagem original  $I$  e a imagem aberta  $\gamma$ , e é determinado por

$$WTH(I) = I - \gamma(I).$$

A transformação top-hat por fechamento de uma imagem  $I$  é definida pela diferença aritmética entre o fechamento  $\phi$  da imagem original e a imagem original  $I$

$$BTH(I) = \phi(I) - I.$$

Neste trabalho utiliza-se a transformada Top-hat de fechamento como auxílio para a correção do fundo. Maiores detalhes sobre essa transformação podem ser obtidos em [2] e [4].

A Figura 2 mostra a imagem original e a Figura 3 mostra a mesma imagem, porém após o tratamento (pré-processamento) recebido.

### 3 O Problema da Detecção de Canaletas

O problema da detecção de canaletas consiste em detectar a quantidade e a posição das

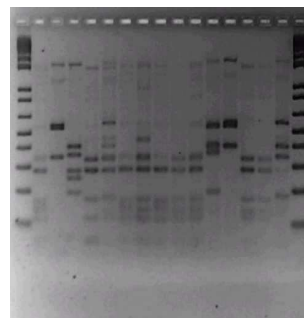


Figura 2: Imagem original

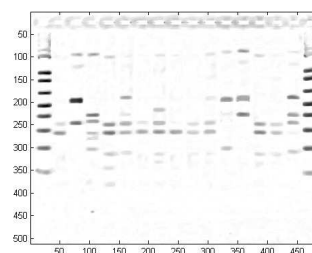


Figura 3: Imagem após o pré-processamento

canaletas e separá-las, gerando imagens distintas tantas quantas forem as canaletas existentes na imagem.

Alguns aspectos importantes merecem atenção, pois ajudam a reduzir a complexidade do problema. O primeiro é o fato de que as canaletas são geralmente corredores verticais, ou seja, são faixas verticais no gel e os compartimentos que as contêm possuem a mesma largura. Um problema aparece quando um dos compartimentos não recebe material, pois isso interrompe a regularidade do gel. Um outro aspecto importante é o fato de que as coordenadas  $y$  dos pixels pertencentes às bandas não são essenciais para a localização de uma canaleta. Baseados nesse aspecto e para simplificar o problema toma-se a projeção da imagem no eixo  $x$ , reduzindo assim sua dimensão. A projeção pode ser tomada visto que a posição da banda no eixo  $y$  não é importante, mas sim sua posição no eixo  $x$ .

Em geral as imagens de géis de eletroforese são de bandas escuras e fundo claro. Os algoritmos da forma como são apresentados neste trabalho são aplicados à imagens invertidas, ou seja, bandas claras e fundo escuro. A implementação recebe a imagem de entrada com as canaletas verticalmente dispostas, com bandas escuras, como mostrado na Figura 1.

Lembrando que usualmente uma imagem possui intensidades de pixels no intervalo  $[0, 255]$ , o programa computacional inverte a imagem, normalizando-a se necessário. Neste caso, normalizar significa colocar os valores da função imagem no intervalo acima citado.

### 3.1 Projeção

Como mencionado anteriormente, a projeção de uma imagem pode ser utilizada para diminuir a complexidade do problema. As duas principais técnicas para calcular a projeção de uma imagem  $I$ , de dimensão  $M \times N$ , onde  $M$  é o número de linhas e  $N$  é o número de colunas, no eixo  $x$  são a projeção pela soma e a projeção pelo máximo.

A primeira toma a soma dos elementos de cada coluna da imagem:

$$P(x) = \sum_{i=1}^N I(i, x), \forall x = 1, \dots, M.$$

A projeção pelo máximo é feita tomando o valor máximo de cada coluna.

$$P(x) = \max_{i=1, \dots, N} \{I(i, x)\}, \forall x = 1, \dots, M.$$

Uma desvantagem da projeção pela soma é que as bandas que contêm pouco material não se tornam evidentes e podem ser excluídas durante o processo de filtragem. Outra desvantagem é que o ruído pode se tornar muito visível na projeção, já que todos os pixels da coluna contribuem na soma, e pode ser erroneamente tomado como canaleta.

### 3.2 Máximos Locais

Como as imagens foram normalizadas de forma que as intensidades dos pixels estejam no intervalo  $[0, 255]$ , os máximos são atingidos onde existem bandas. A projeção apresenta muitos máximos locais que podem indicar a presença de canaletas na imagem. A Figura 4 mostra a projeção obtida pelo máximo da imagem apresentada na Figura 1.

Para reduzir o número de máximos locais obtidos na projeção, suaviza-se a função utilizando o filtro da média que consiste em calcular o valor de  $f_i$  de coordenada  $i$  como a média entre  $P_{i-1}$ ,  $P_i$ ,  $P_{i+1}$ , onde  $P$  é a projeção pelo máximo e  $f$  é a projeção suavizada.

Em geral uma filtragem (ou suavização da função) não é suficiente para reduzir o número

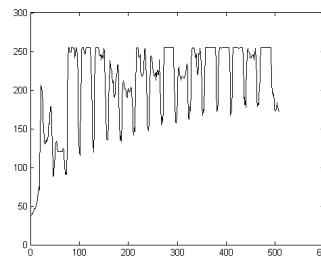


Figura 4: Projeção pelo máximo

de máximos locais ao desejado. Uma alternativa é aplicar o filtro mais de uma vez, sucessivamente, até atingir uma quantidade de máximos próxima da esperada.

Baseado no fato de que as canaletas geralmente são equidistantes, Machado et. al. em [3] propuseram o seguinte algoritmo iterativo para decidir quando suspender a aplicação do filtro:

1. Filtre a projeção com o filtro especificado.
2. Encontre os máximos locais.
3. Calcule as distâncias entre cada par de máximos consecutivos e encontre a média.
4. Se existir qualquer distância que não seja um múltiplo inteiro dessa média, vá para o passo 1. Senão, finalize.

Se existe uma canaleta que não possui material, a distância entre as canaletas anterior e posterior será o dobro do valor da média. Essa relação inteira entre as distâncias e a média deve ter alguma tolerância.

A Figura 5 mostra a projeção depois de submetida ao algoritmo anterior, mostrada junto com a imagem teste. Os pontos brancos são os máximos locais da projeção.

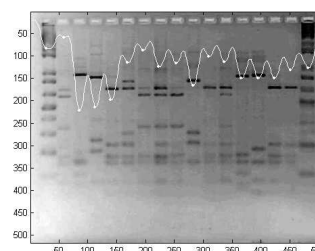


Figura 5: Projeção pelo máximo.

Como o resultado da filtragem ainda depende muito da tolerância utilizada nesse algoritmo, estão sendo estudadas outras alternativas para a suavização da função projeção, como por

exemplo o Filtro de Kalman, um processo recursivo para solucionar problemas lineares relacionados à filtragem de dados discretos.

### 3.3 Algoritmos

Algumas vezes a posição correta da canaleta nas imagens não é exatamente a posição do máximo correspondente detectado no processo de filtragem. Isso pode ocorrer tanto por efeito do filtro quanto pela aparência do gel, que pode apresentar diferentes concentrações de material, levando a projeções não simétricas em relação ao centro da canaleta. Três algoritmos foram propostos em [3] para detectar a posição correta da canaleta.

A partir da projeção filtrada ( $f$ ), encontra-se seus máximos e mínimos na forma

$$a_0, b_1, a_1, b_2, \dots, b_n, a_n$$

onde  $b_i$  é a  $i$ -ésima coordenada do  $i$ -ésimo máximo, que é precedido e sucedido por dois mínimos respectivamente denotados por  $a_{i-1}$  e  $a_i$ .

Assim, tem-se o primeiro algoritmo:

#### Algoritmo 1

1. Encontre os valores de  $n$  máximos  $b$  e  $n+1$  mínimos  $a$ .
2. Para  $i = 1 : n$  faça

- (a) Calcule a área entre dois mínimos consecutivos

$$A_i = \sum_{j=a_{i-1}}^{a_i} f_j.$$

- (b) Calcule o primeiro momento entre dois mínimos consecutivos

$$m_i = \sum_{j=a_{i-1}}^{a_i} j f_j.$$

- (c)  $c_i = m_i / A_i$ .

$c_i$  será a coordenada  $x$  da canaleta  $i$ .

Esse algoritmo leva em consideração o centro de área entre dois mínimos consecutivos, considerando toda a distribuição do material ao invés de somente o máximo. Mas nos casos em que existe uma canaleta que não possui material, a posição calculada pode cair fora da canaleta, ou seja, pode cair nas canaletas vizinhas. Isso acontece porque a área da canaleta que não possui material também é computada.

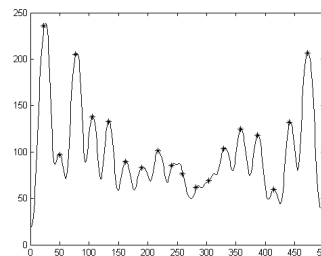


Figura 6: Função projeção e posição das canaletas identificadas pelo Algoritmo 1.

A Figura 6 mostra o gráfico da projeção de uma imagem e os pontos \* são as posições da canaleta identificadas pelo Algoritmo 1. Note que nem sempre essa posição é a mesma da posição do máximo.

Um outro algoritmo simples para calcular a posição correta da canaleta não levando em consideração o valor da projeção ou a área na região é dado abaixo.

#### Algoritmo 2

1. Encontre os valores de  $n$  máximos  $b$  e  $n+1$  mínimos  $a$ .

2. Para  $i = 1 : n$  faça

$$c_i = ((a_{i-1} + a_i) / 2 + b_i) / 2.$$

$c_i$  será a coordenada  $x$  da canaleta  $i$ .

Esse algoritmo leva em consideração apenas as posições dos mínimos e dos máximos, considerando a simetria que existe entre dois mínimos consecutivos. Possui também a mesma desvantagem do anterior no caso de canaletas sem material.

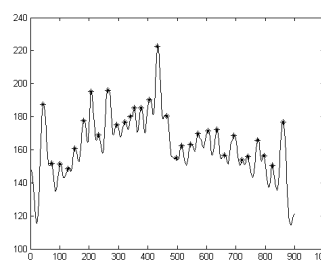


Figura 7: Função projeção e posição das canaletas identificadas pelo Algoritmo 2.

A Figura 7 mostra o gráfico da projeção de uma outra imagem e as posições da canaleta identificadas pelo Algoritmo 2.

O Algoritmo 3 visa melhorar a identificação de canaletas em imagens que apresentam ausência de material em alguma delas.

### Algoritmo 3

1. Encontre os valores de  $n$  máximos  $b$  e  $n+1$  mínimos  $a$ .

2. Para  $i = 1 : n$  faça

- (a) Encontre o mínimo  $d_i$  mais próximo de  $b_i$  fazendo  $d_i = \min(b_i - a_{i-1}, a_i - b_i)$ .
- (b) Calcule a área entre o segmento que vai de  $(b_i - d_i, f(b_i - d_i))$  até  $(b_i + d_i, f(b_i + d_i))$  e chame-a de  $A_i$ .
- (c) Calcule o primeiro momento da área considerada no passo anterior e chame-a de  $m_i$ .
- (d)  $c_i = m_i/A_i$ .

$c_i$  será a coordenada  $x$  da canaleta  $i$ .

Esse algoritmo elimina o problema de erro na detecção da posição das canaletas vizinhas a uma canaleta sem material porque leva em consideração apenas a área acima da reta que passa pelos dois mínimos. A Figura 8 mostra um resultado obtido com o Algoritmo 3.

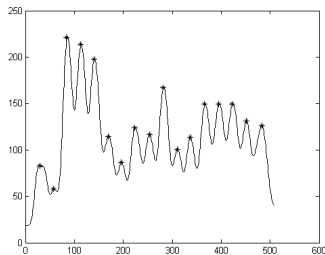


Figura 8: Função projeção e posição das canaletas identificadas pelo Algoritmo 3.

## 4 Resultados

Nesta seção são apresentados alguns resultados obtidos com a implementação computacional realizada para os procedimentos de pré-processamento e detecção das canaletas.

A Figura 9 mostra um exemplo de imagem na qual o Algoritmo 1 foi aplicado. A imagem mostrada é pré-processada e invertida. As linhas brancas verticais que cortam a imagem são as posições das canaletas identificadas pelo Algoritmo 1.

A imagem pré-processada mostrada na Figura 10 exemplifica uma aplicação do Algoritmo 2. Pode-se ver na Figura 10 que ape-

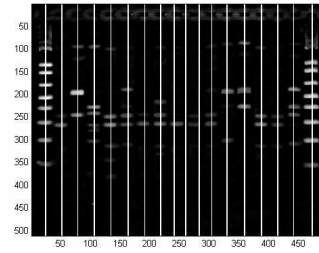


Figura 9: Imagem pré-processada com as marcações das posições das canaletas obtidas pelo Algoritmo 1.

nas uma das canaletas não foi identificada e as demais foram identificadas corretamente.

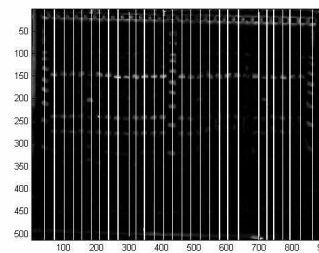


Figura 10: Imagem pré-processada com as marcações das posições das canaletas obtidas pelo Algoritmo 2.

A imagem mostrada na Figura 11 é um exemplo de imagem na qual o Algoritmo 3 foi aplicado.

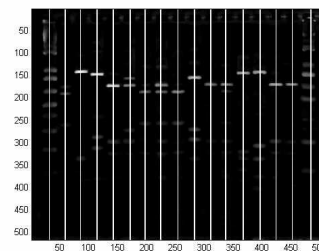


Figura 11: Imagem pré-processada com as marcações das posições das canaletas obtidas pelo Algoritmo 3.

A Figura 11 que foi obtida de uma imagem pré-processada e da aplicação do Algoritmo 3, mostra todas as canaletas identificadas corretamente. A implementação aplicada na imagem original, sem o pré-processamento proposto, não identificou a segunda canaleta.

O objetivo final é o de obter arquivos sepa-

rados em que cada um contenha uma canaleta. A Figura 12 mostra algumas canaletas já separadas, obtidas de uma imagem pré-processada. Esta imagem foi invertida para realizar o procedimento de detecção, mas a separação propriamente dita foi aplicada na imagem sem inversão do fundo. Assim, a Figura 12 mostra algumas das canaletas da Figura 1, prontas para utilização computacional.

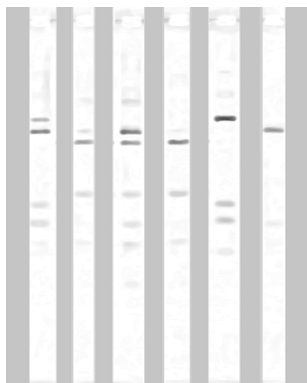


Figura 12: Canaletas já separadas de uma imagem pré-processada sem inversão do fundo.

## 5 Conclusões e Propostas Futuras

Nas imagens testadas obteve-se um melhor resultado nas imagens nas quais foram aplicadas as técnicas de pré-processamento, ou seja, nas imagens pré-processadas as canaletas foram identificadas de forma correta enquanto nas originais houve perda de canaletas.

Os três algoritmos para a detecção das canaletas apresentaram um resultado similar entre si para imagens nas quais todas as canaletas possuíam material.

No processo de eletroforese algumas vezes o campo elétrico não é distribuído de forma uniforme, gerando uma deformação nas laterais da imagem chamada de efeito "sorriso". Um dos trabalhos futuros é desenvolver um algoritmo para corrigir esse efeito.

Uma proposta de trabalho futuro é, ainda, desenvolver um algoritmo utilizando sistemas inteligentes, redes neurais por exemplo, para classificar automaticamente a qualidade de uma imagem como boa, regular ou ruim.

A tolerância no filtro citada na seção 3.2 é um ponto importante do processamento, pois uma pequena mudança nesta pode implicar

uma grande redução ou aumento do número de canaletas detectadas. Assim, como os resultados dependem dessa tolerância, serão feitas alterações no processo de filtragem, utilizando o filtro de Kalman para suavizar a função projeção de forma que não seja necessário o controle da tolerância. Serão realizados testes exaustivos em diversas imagens para comparação com resultados obtidos por outros autores.

## 6 Agradecimentos

Agradecemos ao CNPq e à CAPES pelo auxílio financeiro.

## Referências

- [1] C. A. Z. Barcelos, M. Boaventura, and E. C. Silva, Jr "A Well-Balanced Flow Equation for Noise Removal and Edge Detection," *IEEE Trans. on Image Processing*, Vol. 12, No. 7, July 2003.
- [2] R. C. Gonzalez, and R. E. Woods, "Processamento de Imagens Digitais," Ed. Edgard Blucher, São Paulo - SP, 2000.
- [3] A.M.C. Machado, M.F.M. Campos, A.M. Siqueira, and O.S.F. de Carvalho, "An Iterative Algorithm for Segmenting Lanes in Gel Electrophoresis Images," *Proceedings of the XI SIBGRAPI*, pp. 140-146, 1997.
- [4] N. G. Medeiros, E. A. da Silva, and J. R. Nogueira "Segmentação Morfológica de Imagens utilizando o Gradiente Morfológico Multi-Escala", *Revista Brasileira de Cartografia*, Presidente Prudente, V. 01, No. 54, pp. 77-85, 2002.
- [5] X. Ye, C. Y. Suen, M. Cheriet, and E. Wang, "A Recent Development in Image Analysis of Electrophoresis Gels," *Proceedings of Vision Interface, Trois-Rivières*, pp. 432-438, May 1999.