4
Algoritmo de reconstrução de curvas

Como visto na introdução, o algoritmo proposto neste trabalho é constituído pelas seguintes etapas:

1. Agrupamento espacial dos pontos de amostragem
2. Processamento dos pontos de amostragem (redução de ruído, e ajuste da densidade dos pontos de amostragem através de superamostragem - processo comentado mais adiante)
3. Afinamento e podagem das células com pontos de amostragem em seu interior
4. Construção da curva a partir das células obtidas na etapa anterior

A seguir, cada uma das etapas do algoritmo será descrita em detalhes. Serão mostradas as estruturas de dados, as técnicas e os parâmetros utilizados em cada uma das fases do algoritmo. A reconstrução de uma curva paramétrica com ruído e densidade de amostragem não uniforme servirá como exemplo para ilustrar o algoritmo proposto. A equação geradora da curva de exemplo é a seguinte espiral:

$$(x(t), y(t)) = \left(\cos(t)\frac{e^{t/20}}{4}, \sin(t)\frac{e^{t/20}}{4}\right)$$

Foram utilizados neste exemplo 10.000 pontos de amostragem, sendo que o parâmetro t variou uniformemente de 0 a 10π, causando maior densidade de amostragem no centro da espiral. A cada ponto foi aplicado um ruído máximo de 0.02 em uma direção aleatória, uniformemente distribuído. Os pontos de amostragem da curva podem ser vistos na figura 4.1.

No exemplo utilizado para ilustrar o algoritmo de reconstrução, o grid foi dividido em 256^2 células. Neste caso, a dimensão do lado de uma célula é de aproximadamente 0.01, o que equivale à metade do ruído máximo dos pontos de amostragem. Na figura 4.2, podemos ver a ampliação de uma região do grid, além de alguns pontos de amostragem.

PUC-Rio - Certificação Digital Nº 0124793/CA

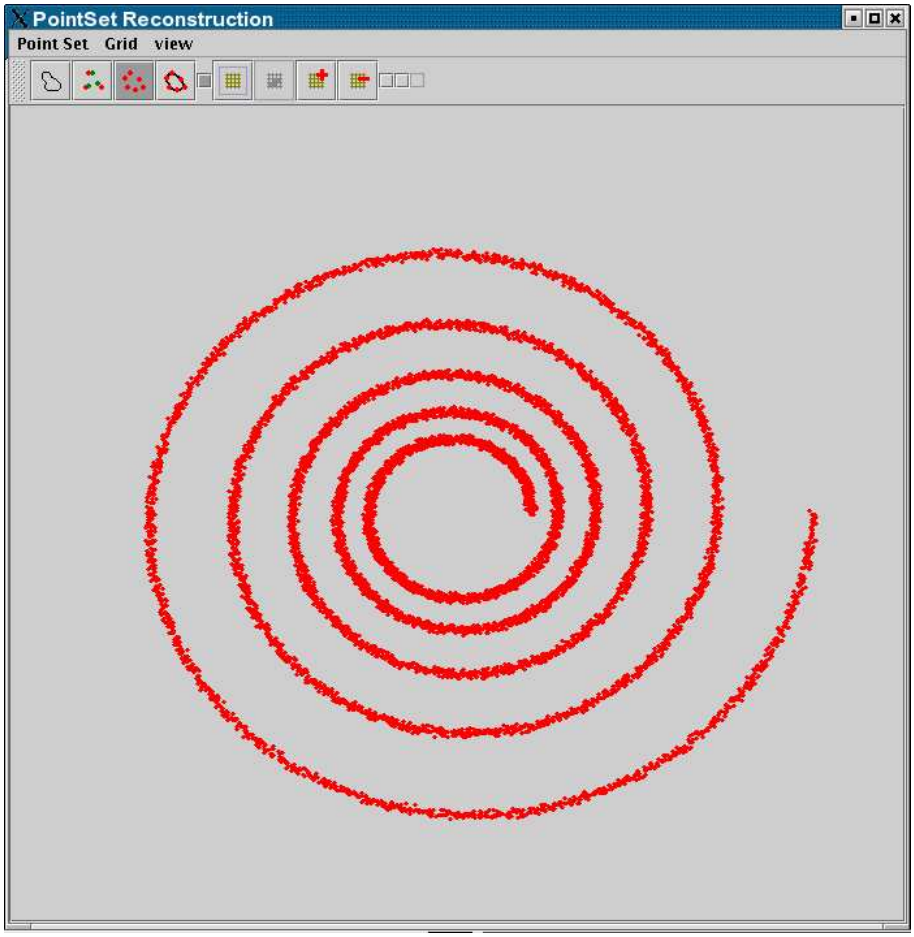

Figure 4.1: 10.000 pontos de amostragem com ruído

4.1 Agrupamento espacial dos pontos de amostragem

A etapa de agrupamento espacial dos pontos de amostragem consiste em armazenar os pontos em células de um *grid* retangular, onde o *grid* é um retângulo de centro (x_0, y_0), largura w e altura h, onde todos os pontos de amostragem estão contidos. O *grid* é dividido em m^2 células de dimensões iguais, sendo m linhas e m colunas de células. A célula (i, j) do *grid* representa o retângulo de centro $(i \times (w/m) + (w/2m), j \times (h/m) + (h/2m))$, largura w/m e altura h/m.

A estrutura de dados necessária para realizar tal tarefa pode ser bastante simples, como a utilizada na implementação proposta: o *grid* é representado por uma matriz de listas encadeadas, onde a lista encadeada na posição (i, j) da matriz armazena os pontos da célula (i, j). Esta implementação tem como vantagem permitir o acesso rápido aos pontos de uma célula, bem como acessar rapidamente as células vizinhas de uma determinada célula. Porém, é uma estrutura de dados que consome muita memória, pois o número de células do

PUC-Rio - Certificação Digital Nº 0124793/CA

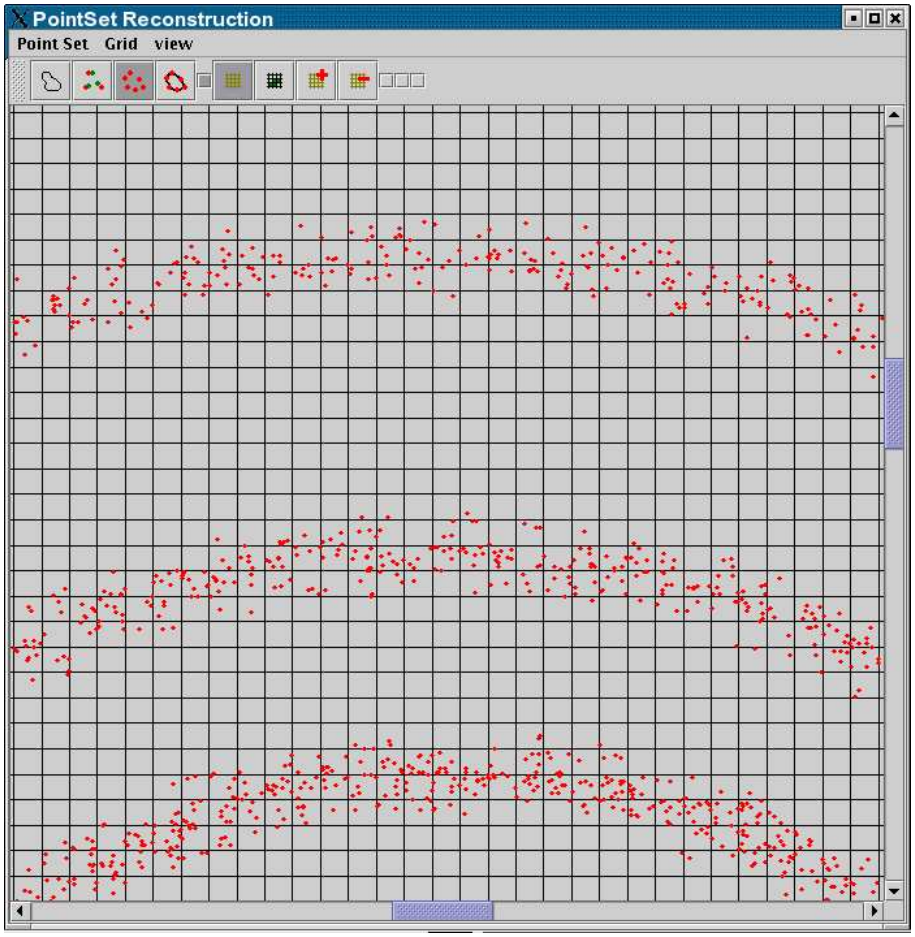

Figure 4.2: Células do grid

grid é de ordem $O(m^2)$.

Outras estruturas de dados seriam possíveis, como a utilização de *quad-tree*. Estas estruturas de dados representam a divisão espacial de um plano através de uma árvore, onde cada nó representa uma partição P do plano, e os filhos de um nó P são partições Q_i de P. Existem várias implementações possíveis para *quad-trees*. Uma alternativa é tentar balancear a árvore em relação aos pontos de amostragem, sendo que os nós de um mesmo nível devem possuir um número igual (ou quase igual) de pontos. Outra alternativa é dividir igualmente o espaço em cada uma das partições, de modo que, dado um ponto (x, y) no plano, se sabe de antemão como percorrer a árvore, sem precisar comparar as coordenadas de cada partição com o ponto que se está buscando. A implementação de *quad-trees* é mais complexa e mais lenta que a implementação utilizada neste trabalho. Mas caso seja necessário utilizar um número muito grande de divisões para o *grid*, a utilização de *quad-trees* pode ser mais adequada.

PUC-Rio - Certificação Digital Nº 0124793/CA

4.2 Processamento dos pontos de amostragem

A etapa de processamento dos pontos de amostragem tem como objetivo prepará-los para as etapas subseqüentes. Nesta etapa, é realizada a redução de ruído e o ajuste da densidade dos pontos de amostragem. A redução de ruído consiste em aplicar o operador $\tilde{P}_m$ a cada ponto de amostragem r_i, de modo que este se aproxime da curva induzida pelo conjunto dos pontos de amostragem através do algoritmo MLS. Já o ajuste de densidade é feito através da superamostragem do conjunto dos pontos de amostragem nas regiões onde a densidade de pontos não for suficiente para a etapa de afinamento digital. Ambas as etapas utilizam uma simplificação da projeção MLS como técnica para localizar a projeção de um ponto de amostragem na curva.

Implementação simplificada da projeção MLS

Como já foi mencionado anteriormente, a projeção MLS é um processo caro em termos computacionais, pois requer a utilização de métodos iterativos para a resolução de sistemas não lineares. Para reduzir este custo, foram feitas algumas simplificações sobre as projeções MLS, de modo que fosse possível utilizar apenas sistemas lineares. Note que a simplificação é necessária apenas na primeira etapa da projeção MLS, onde é achada uma reta de referência. Na segunda etapa, onde é achado um polinômio que aproxima a curva na vizinhança do ponto que está sendo projetado, os pesos atribuídos aos pontos de amostragem já estão fixados.

Nesta simplificação, uma aproximação inicial para o vetor tangente é encontrada fixando-se a reta de referência no ponto r (ou seja, t é fixado em zero), e calculando-se o autovetor associado ao menor autovalor da matriz de covariâncias descrita através da expressão (2-4). A seguir, a reta foi movida paralelamente, de modo que a média ponderada das distâncias de cada ponto de amostragem para a reta fosse nula. O peso θ_i de cada ponto de amostragem r_i para o cálculo da média ponderada foi dado em função da distância entre o ponto de amostragem e o ponto r a ser projetado na curva:

$$\theta_i = e^{-\|r_i - r\|^2/h^2}, \tag{4-1}$$

onde o parâmetro h ajusta a localidade do algoritmo, ou seja, quanto maior for h, menor será a região em torno de r considerada para o cálculo da projeção MLS.

PUC-Rio - Certificação Digital Nº 0124793/CA

A utilização do agrupamento espacial dos pontos de amostragem em células do *grid* ocupa um papel fundamental no ajuste da localidade. Note que a função peso decai para pontos distantes do ponto a ser projetado. Dependendo da distância de um ponto de amostragem para o ponto a ser projetado, o peso associado a este ponto de amostragem pode ser tão pequeno que o ponto de amostragem pode ser descartado. De fato, apenas os pontos de amostragem mais próximos ao ponto a ser projetados devem ser levados em consideração no cálculo de sua projeção. Por isso, podemos utilizar apenas os pontos de amostragem interiores às células ao redor da célula a qual pertence o ponto a ser projetado. O conjunto de células em torno da célula na qual se localiza o ponto r a ser projetado é denominado vizinhança. O parâmetro h é igual ao tamanho da diagonal do conjunto de células utilizado como vizinhança. Com este valor de h, garante-se que todos os pontos de amostragem com peso relevante estarão incluídos nestas células.

Para escolher uma vizinhança de tamanho adequado em cada região, são calculadas projeções com vários tamanhos de vizinhança. Para comparar as aproximações obtidas com as várias vizinhanças, calculou-se o somatório das distâncias dos pontos de amostragem à reta encontrada, ponderadas pela função peso de cada ponto de amostragem. Em seguida, calculou-se este mesmo erro para a reta perpendicular à reta encontrada e que passa pelo ponto que foi projetado. A medida comparativa entre os vários tamanhos de vizinhança é a razão entre os erros associados à reta obtida pela aproximação MLS e à reta perpendicular: quanto menor a razão, melhor é a aproximação. Os tamanhos de vizinhança variam de 1 célula até conjuntos de 16×16 células, cujo centro é a célula a qual pertence o ponto a ser projetado. O algoritmo é repetido até que se encontre uma razão entre os erros menor que 0.01, ou então até que o tamanho da vizinhança alcance 16×16. Se não for encontrado um tamanho de vizinhança onde a razão entre os error é menor que 0.01, considera-se o tamanho de vizinhança com menor razão entre os erros.

Depois que a reta de referência é encontrada, calcula-se o polinômio que aproxima a curva nas proximidades do ponto r. O cálculo do polinômio $p(x)$ é realizado através de um sistema linear de equações, sendo resolvido por decomposição LU. A projeção de r na curva é o ponto $(0, p(0))$, no sistema de coordenadas cujo eixo de x é a reta de referência anteriormente encontrada.

É importante ressaltar que esta aproximação, embora seja adaptativa, aumenta o erro na aproximação da projeção MLS. Mas este erro não é grande o suficiente para justificar a implementação de um método mais preciso e, por isso, mais caro computacionalmente. Observe que o objetivo da etapa de redução de ruído é fazer com que a maior parte dos pontos de

PUC-Rio - Certificação Digital Nº 0124793/CA

amostragem sejam movidos para a mesma célula ocupada por sua projeção MLS. O procedimento proposto acima atinge este objetivo e possui um custo computacional muito baixo quando comparado à uma implementação mais fina da projeção MLS.

Uma última observação é que a função peso $\theta(x) = e^{-x^2/h^2}$ pode ser aproximada para a função, $\theta(x) = 2(x/h)^3 - 3(x/h)^2 + 1$, se $x < h$, e $\theta(x) = 0$, se $x \geq h$. Note que, como x representa a distância entre dois pontos, não faz sentido em avaliar a função para valores de x menores que zero.

Redução de ruído

A redução de ruído é bastante simples: a idéia consiste apenas em projetar cada um dos pontos de amostragem sobre a curva, através de projeções MLS. Para que pontos com erro de aproximação elevado não sejam incluídos no novo conjunto de pontos, foi adotada a seguinte heurística: avalia-se a derivada no ponto $x = 0$ da função polinomial encontrada na projeção MLS. Note que este ponto equivale ao ponto de amostragem projetado na reta de referência. Como a reta de referência deve se aproximar da reta tangente, então espera-se que a derivada neste ponto seja próxima de zero. Um valor de derivada encontrado para descartar os pontos foi de 0.5, ou seja, pontos com derivada maior que 0.5 são descartados. Este valor foi achado empiricamente.

Superamostragem

A superamostragem consiste na geração de novos pontos de amostragem próximos à curva. O objetivo deste processo é aumentar a densidade de pontos de amostragem para as etapas seguintes do algoritmo de reconstrução. É importante saber como a curva será reconstruída a partir das células do grid para que se possa entender melhor o objetivo desta etapa: cada célula do grid com pontos de amostragem em seu interior será marcada como "célula cheia". O conjunto de células cheias será afinado e podado de modo que as células remanescentes deste processo representem a curva de reconstrução sob uma forma digital. Ao final, cada célula remanescente representará um vértice na curva de reconstrução, que será formada a partir da ligação dos vértices em células vizinhas.

É importante, então, que as células com pontos de amostragem em seu interior formem componentes conectados que representem cada uma das partes da curva de reconstrução (note que a curva pode conter várias componentes). Porém, o ruído e a densidade dos pontos de amostragem podem causar

PUC-Rio - Certificação Digital Nº 0124793/CA

"buracos" nas curvas, impedindo a reconstrução correta. Assim, o algoritmo de reconstrução deve acrescentar novos pontos de amostragem, de modo a reconstruir a curva corretamente. Uma questão importante é a seguinte: como saber se uma falha no conjunto de pontos de amostragem deve ser preenchida com pontos pela superamostragem ou se deve ser mantida como uma descontinuidade na curva. A resposta é que não é possível saber se um buraco é realmente um buraco ou uma falha na amostragem, como foi colocado por [7]. Neste algoritmo, uma falha será preenchida se o seu tamanho for menor que cinco vezes a largura da célula do *grid*, mas isto é decorrente do processo de reconstrução. Assim, o refinamento do *grid* deverá ser ajustado de modo que preserve os "buracos verdadeiros" da curva a ser reconstruída.

O algoritmo de superamostragem utiliza projeções MLS para avaliar novos pontos de amostragem. Para cada um ponto de amostragem inicial r_i (já com ruído reduzido), avalia-se H, a reta de referência local, que deve aproximar a reta tangente à curva nas proximidades de r_i. Os pontos de amostragem próximos a r_i são projetados na reta, dividindo a reta em vários segmentos, cujos extremos são as projeções dos pontos de amostragem. Para cada um dos segmentos obtidos, se o seu tamanho for maior que um limite pré-definido, adiciona-se um ponto p no centro do segmento. O ponto p é, então, avaliado no polinômio de aproximação da curva encontrado na projeção MLS. Este ponto é o novo ponto de amostragem. Note que este procedimento dividiu um segmento em dois, ou seja, dois novos segmentos foram criados. O processo repete-se recursivamente para esses dois novos segmentos, até que o tamanho dos mesmos seja menor que um limite. Este limite deve ser estabelecido de modo que se garante que a densidade de pontos de amostragem ao final do processo seja suficiente para as etapas seguintes do algoritmo de reconstrução. O limite escolhido no algoritmo implementado foi de 40% da diagonal de uma célula do *grid*.

Para evitar que pontos com erro muito grande sejam inseridos na superamostragem, pontos com derivada maior que 0.5 foram descartados. Note que espera-se que os pontos da superamostragem devem estar próximos da origem do sistema de referência local, e também que a derivada nas proximidades da origem do sistema de referência não deve variar muito.

Este processo de superamostragem está ilustrado na figura 4.3. Os o pontos vermelhos são os pontos de amostragem iniciais. Estes pontos são projetados em uma reta de referência local, formando-se, assim, diversos segmentos cujos extremos são as projeções destes pontos. Os extremos dos segmentos são os pontos azuis. Sempre que um segmento for maior que um tamanho máximo, um novo ponto é adicionado no centro do segmento (pontos

PUC-Rio - Certificação Digital Nº 0124793/CA

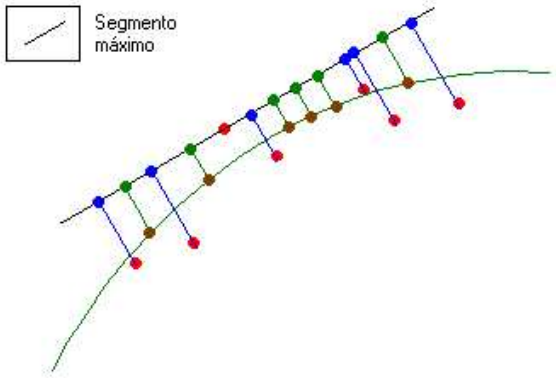

Figure 4.3: Superamostragem

verdes), e cada um destes novos pontos é projetado na curva. A projeção deste novo ponto é adicionado ao conjunto de pontos de amostragem (pontos marrons). Este algoritmo foi inspirado no trabalho de Alexa et al., em [1].

As figuras 4.4 e 4.5 mostram os pontos de amostragem originais em vermelho, e os pontos de amostragem processados, em azul.

PUC-Rio - Certificação Digital Nº 0124793/CA

4.3 Afinamento, podagem e reconstrução

O afinamento e a podagem foram realizados conforme explicado no capítulo 3. A única consideração a ser feita aqui é que foram executadas duas etapas de podagem para remover as ramificações da curva de pequeno comprimento, resultantes do processo de afinamento. As figuras 4.4 e 4.5 mostram o resultado da podagem e do afinamento no exemplo que está sendo acompanhado. As células amarelas e verdes representam o conjunto das células cheias antes do afinamento e da podagem, e as células verdes são as células que permaneceram neste processo.

Uma vez que o afinamento e a podagem foram realizados, o conjunto de células cheias passou a representar, sob a forma digital, o conjunto de curvas de reconstrução. Para transformar a curva digital em uma curva linear por partes é bastante simples: a cada célula da curva digital é associado um vértice da curva linear por partes. O vértice em questão é a projeção MLS do centro da célula correspondente. Basta, então, que os vértices sejam devidamente ligados pelas arestas da curva. As arestas são obtidas através da ligação dos vértices associados a duas células vizinhas. As figuras 4.6, 4.7 e 4.8 mostram o resultado final do processo de reconstrução.

Dois cuidados adicionais devem ser tomados para que as singularidades

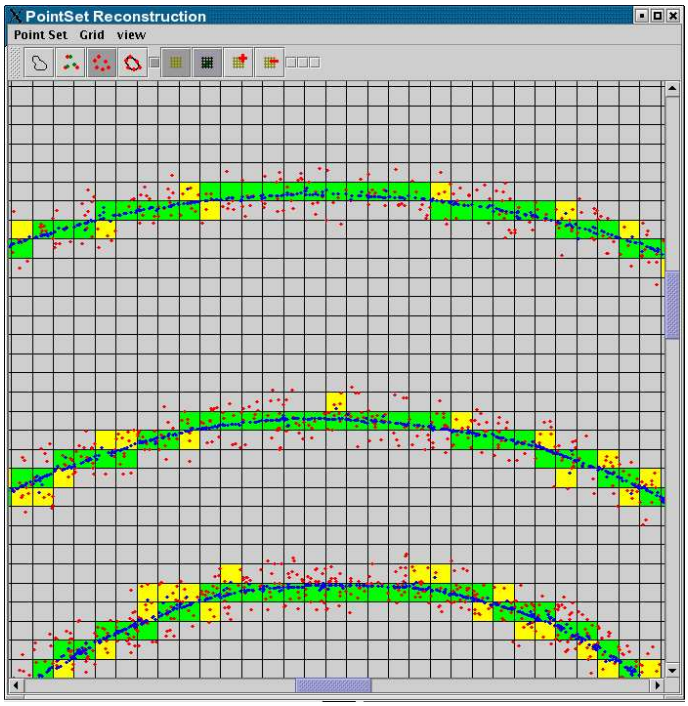

Figure 4.4: Processamento de pontos de amostragem e afinamento

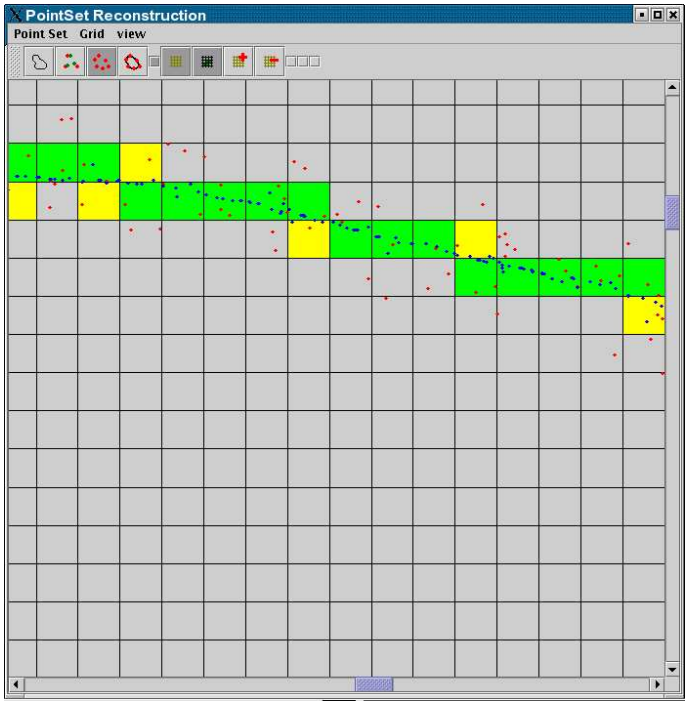

Figure 4.5: Processamento de pontos de amostragem e afinamento (zoom)

PUC-Rio - Certificação Digital Nº 0124793/CA

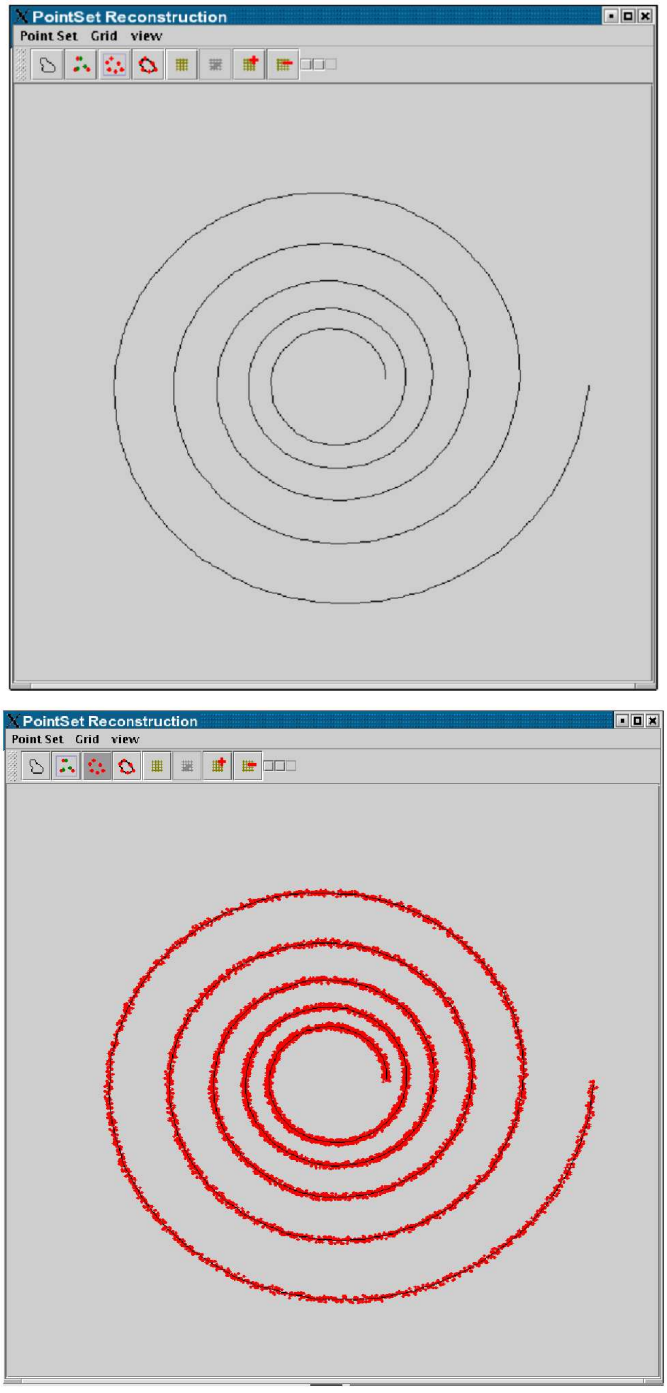

Figure 4.6: Resultado final

PUC-Rio - Certificação Digital Nº 0124793/CA

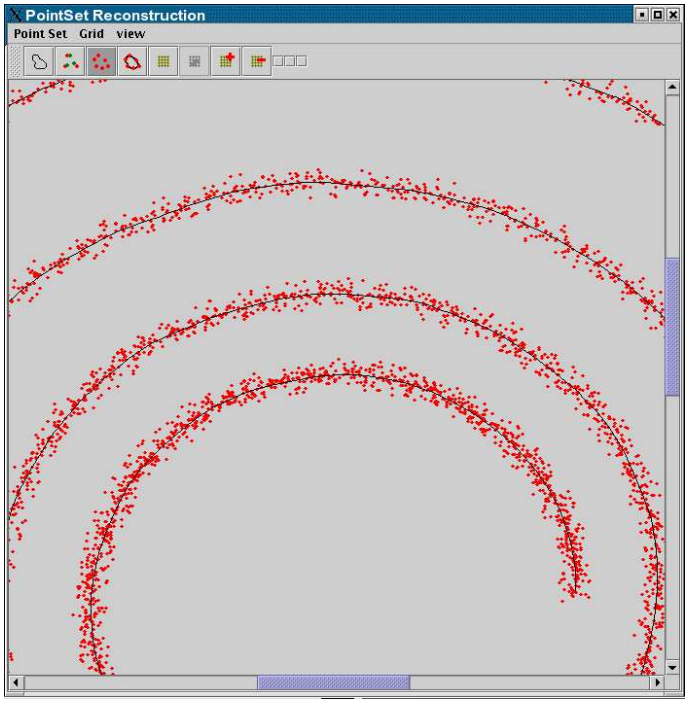

Figure 4.7: Resultado final (zoom)

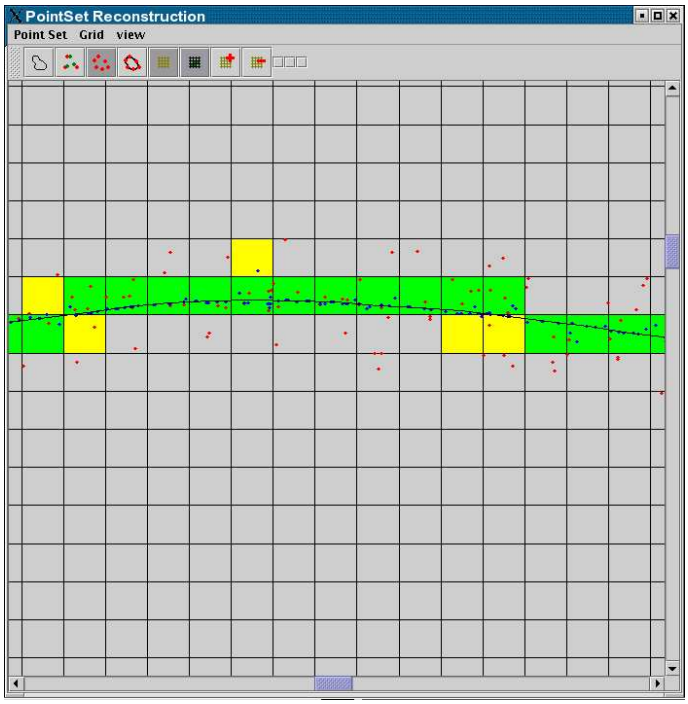

Figure 4.8: Resultado final (ilustrando a curva digital)

PUC-Rio - Certificação Digital Nº 0124793/CA

sejam tratadas. Nas singularidades, é possível que uma célula possua mais de duas células vizinhas. O caso de três células vizinhas entre si é tratado ligando-se apenas as células 4-conectadas. O caso de 4 células vizinhas entre si é tratado utilizando-se o vértice em comum destas 4 células como vértice da curva de reconstrução.

PUC-Rio - Certificação Digital Nº 0124793/CA