

SISTEMA COMPUTACIONAL PARA REPRESENTAÇÃO VETORIAL DE PLUVIOGRAMAS IMPRESSOS

Ademir Paceli Barbassa¹; Jander Moreira²; Alexandre Zollinger Chohfi³; Lorena Avelina Rojas Gutierrez⁴

RESUMO – Um sistema para a extração de dados vetoriais a partir de registros impressos, em particular pluviogramas, é apresentado. Denominado Kairos, apresenta funcionalidades para o processamento de imagens digitais obtidas por meio de escâner. O sistema viabiliza o resgate, análise e o armazenamento de registros climatológicos. Apresentando-se como um sistema para a vetorização de imagens, possibilita o tratamento dos dados de forma automática na maior parte do tempo, minimizando a necessidade de intervenção manual do operador. Por meio de recursos como ampliação da imagem ou partes dela e da correção total ou parcial da vetorização já realizada, assiste o operador, reduzindo o esforço e o tempo dispendidos. Testes iniciais mostraram que o sistema é flexível e aplicável em grande escala.

ABSTRACT – This paper presents a new computer system, which is able to extract data from printed records, in this case pluviograms. This system, called Kairos, employs image processing techniques on scanned images. It also allows the possibility of recovering, analysis and storage of climatological records. The graphics present in the images are vectorized automatically most of the time, reducing the need of manual intervention by a human operator. Resources such as local and global zoom and full or partial correction of the vectorized data helps the operator, minimizing the dispended effort and time. A initial set of tests showed that the system is flexible and the application in large scale is viable.

Palavras-chaves: vetorização, gráficos climatológicos, pluviogramas

1 INTRODUÇÃO

O registro de dados climatológicos no Brasil, de forma geral, é bastante escasso e, quando existe, tem seu acesso dificultado por diversas razões. Entre elas, por se tratar de informações que devem ser coletadas por décadas seguidas e por ter condições de conservação e armazenamento muitas vezes precárias. Há, ainda, a conhecida descontinuidade, mudança de prioridade e mesmo descaso para com os registros em diferentes administrações. Outro aspecto, não menos importante, é a grande quantidade de gráficos, o que demanda um esforço humano e de equipamento, além de considerável dispêndio de tempo para trabalhá-los de forma convencional.

¹) Professor Associado do DECiv, UFSCar, Rod. Washington Luis, Km 235, 13566-260, São Carlos, SP; barbassa@ufscar.br

²) Professor Associado do DC, UFSCar, Rod. Washington Luis, Km 235, 13566-260, São Carlos, SP; jander@ufscar.br

³) Bolsista IC/FAPESP, graduando do Curso de Bacharelado em Ciência da Computação, UFSCar, Rod. Washington Luis, Km 235, 13566-260, São Carlos, SP; azchohfi@comp.ufscar.br

^e) Bolsista REUNI, Mestranda do PPG-EU, UFSCar, Rod. Washington Luis, Km 235, 13566-260, São Carlos, SP; lorenavelina@gmail.com

Os dados existentes trazem informações relativas a precipitação, como pluviogramas, dos quais se podem obter relações de intensidade, duração e frequência, requisitos básicos de modelos de previsão de enchentes os mais variados e no dimensionamento de redes de drenagem pluvial e reservatórios de controle de enchentes. Podem também ser usados para determinação de chuvas de projeto por várias formulações. Estas são apenas algumas das aplicações desta variável climatológica do ponto de vista do hidrológico.

A vetorização destes gráficos pode ser realizada por meio de softwares comerciais que disponibilizam recursos de vetorização de imagem. Podem-se citar o AUTOCAD e os sistemas de informações geográficas como o SPRING, o ARCVIEW, o MICROSTATION, CORELDRAW etc. O problema relacionado ao seu uso, quando não o custo de aquisição e licença, é a dificuldade de operação, além de não terem sido projetados para este fim, pois os gráficos são “estruturas” simples, porém com singularidades e particularidades que, tanto quanto possível, devem ter soluções facilitadas. Assim, o emprego destes sistemas não parece ser o mais viável para o proposto.

Foi incentivada, desta forma, a realização do desenvolvimento de um sistema específico, com possibilidade de projetar estruturas de armazenamento, de avaliação e possibilitar o acesso e a formatação dos registros. Em outras palavras, permitir análises conforme aplicação desejada para uma dada área. Este desenvolvimento incorpora, além da disponibilização de um banco de dados, a integração da aplicação tecnológica da área de processamento de dados com outras, como engenharia civil e engenharia ambiental, por exemplo.

Objetiva-se apresentar o sistema computacional semi-automático, denominado Kairos (do grego *καιρός*, *clima*), para extração de dados de pluviogramas, para análises hidrológicas e climatológicas e resgate de registros impressos da Estação Climatológica operada pelo Instituto Nacional de Meteorologia (INMET) e instalada na Universidade Federal de São Carlos (UFSCar). Especificamente pretende-se: demonstrar o processamento do conteúdo das imagens; apresentar as soluções para tratar singularidades, como falhas, borrões e linhas indistintas frequentemente encontradas nos gráficos; e avaliar erros entre os registros originais e os vetorizados. A extração de dados da imagem do pluviograma será ao final, uma “interpretação” ou “leitura” eletrônica de forma confiável e representativa, pois é passível de conferência e verificação entre o que foi registrado pelo pluviógrafo e o que foi extraído pelo Kairos.

Os itens seguintes referem-se aos conceitos empregados, aos materiais e métodos e à descrição do software e de algumas aplicações.

2 SÍNTESE DA REVISÃO BIBLIOGRÁFICA

Este projeto compreende o tratamento de registros climatológicos com vista à análise hidrológica e climatológica. Como parte do processamento necessário para a análise requer o

tratamento dos gráficos (registros) impressos em papel, esta síntese da revisão bibliográfica abrange o tratamento da imagem digital obtida pela captura dos gráficos.

2.1 Captura via mesa digitalizadora

Já houve, em 1990, um esforço, com apoio da FAPESP, para a captura de pluviogramas de 27 anos de dados da Estação Climatológica, os quais foram usados para cálculo da equação de chuvas de São Carlos, conforme FCTH (2001) e Barbassa (1991).

Na época, usou-se mesa digitalizadora para captura dos gráficos, que era a tecnologia mais avançada disponível (a mesa está em perfeitas condições de uso, exceto pelo mouse que não é mais fabricado). Este método mostrou-se muito eficiente, quando comparado com análise manual dos pluviogramas, pelos seguintes motivos:

- Ampliavam-se enormemente as possibilidades de análise dos gráficos, pois era possível representar um minuto, por exemplo, por uma dezena de pontos, o que implica em décimos de minuto, o que é impossível manualmente.
- Os arquivos digitais ocupavam pouco espaço, embora à época fossem centenas de disquetes de 5 ¼ polegadas.
- Reduziam-se os erros humanos, embora não fosse possível eliminá-los completamente.

Como desvantagens daquele método citam-se:

- O tempo gasto para a digitalização do operador do programa via mesa/micro.
- Dificuldade ou impossibilidade de conferência visual e resgate do original. Podia-se, entretanto, conferir a altura de chuva do dia registrada no pluviograma, com aquela obtida do pluviograma vetorizado.

O programa para captura consistia, sucintamente, em vetorizar a linha do gráfico diretamente dele e na seleção de pontos enviados pela mesa ao micro para representação da linha. O arquivo produzido já era o do gráfico vetorizado, após mudanças de coordenadas da mesa para coordenadas dos gráficos com as unidades hidrológicas registradas. O tratamento das várias “singularidades” dos gráficos era feito pelo operador durante a captura na mesa.

Embora alguns cuidados tenham sido tomados para armazenamento, os registros digitais dos pluviogramas, após serem transferidos de uma mídia para outra, acabaram se perdendo.

2.2 Processamento e análise de imagens

Imagens digitais podem ser processadas de forma a permitir extrair delas informações relevantes a uma dada aplicação. De forma geral, a imagem digital corresponde a uma malha regular de duas dimensões, na qual cada cruzamento corresponde a um pixel. Cada pixel é representado por um atributo, que pode ser apenas a intensidade da iluminação ou uma cor. No primeiro caso, as imagens são ditas monocromáticas e representadas por diferentes níveis de cinza,

variando uniformemente do preto ao branco. Quando é feita a representação de cores, a cada pixel é associada uma cor específica, usualmente dada pela combinação aditiva dos estímulos cromáticos nas frequências correspondentes às cores vermelho, verde e azul (RGB), embora haja outras representações. Esta forma matricial de representação é chamada *raster*. (Castleman, 1995; Gonzales e Woods, 2007).

O processamento de imagens dá ferramentas para melhorar as condições da imagem, realçar determinados atributos, segmentar objetos ou áreas de interesse, remover imperfeições e alterar sua representação, entre outras possibilidades.

Para este trabalho, a segmentação das imagens para se obter a linha do gráfico e a restauração para se melhorar o resultado desta segmentação são os principais recursos empregados, além do processo de vetorização em si. Os principais pontos de interesse são revistos nas seções seguintes.

2.2.1 Segmentação de imagens

A segmentação de imagens é o processo de separar os objetos presentes na matriz de pixels. Esta segmentação pode utilizar abordagens balizadas por critérios de diferença ou de semelhança. Variações abruptas nos atributos de pixels vizinhos podem ser atribuídas à transição entre um objeto e outro, caracterizando segmentação por diferenças de atributos locais. Dentro desta classe enquadram-se os métodos de detecção de bordas, principalmente. De um ponto de vista diferente, objetos podem ser entendidos como conjuntos de pixels que compartilham um mesmo atributo, o seja, os que apresentam atributos similares entre seus pixels. Técnicas de segmentação por semelhança envolvem, por exemplo, limiarização, classificação e crescimento de regiões. (Gonzales e Woods, 2007)

Uma opção que se mostra adequada para o tratamento dos gráficos é dada pela classificação dos pixels segundo suas componentes cromáticas. Para tanto, apenas informações cromáticas são consideradas, manipulando o espaço de cores inicial da aquisição (RGB) para o espaço L^*a^*b . Enquanto o espaço RGB, mesmo baseado nas características dos receptores de cor das retinas humanas, não apresenta separabilidade distinta de cores nem linearidade nas diferenças perceptuais, o espaço L^*a^*b separa atributos cromáticos das intensidades luminosas, permitindo que a classificação por cores seja independente de suas características de intensidade. Assim, para a classificação, foi empregada a técnica de classificação por K-médias apenas sobre os atributos cromáticos (matiz e saturação) dos pixels da imagem.

2.2.2 Reconhecimento de gráficos e vetorização

O termo reconhecimento de gráficos se refere, de forma geral, à análise de documentos contendo grande quantidade de gráficos, como mapas, desenhos técnicos, plantas ou outros

esquemas (Tombre e Tabbone, 2000). A conversão dos componentes dos desenhos, como linhas e arcos, da forma *raster* para a forma vetorizada consiste em uma parte essencial deste processo. Na primeira as linhas presentes no gráfico são formadas por pixels de atributo igual (ou similar) em posições vizinhas, compondo uma forma conectada. O processo de vetorização consiste, assim, no processamento da imagem digital *raster* para localizar os elementos nela existentes e representá-los utilizando uma descrição vetorial, ou seja, por listas de vetores.

Técnicas de vetorização em si já são bastante conhecidas e o problema é considerado resolvido (Tombre e Tabbone, 2000). Os problemas intrínsecos da metodologia é que ela é geral e proporciona imprecisões em diversos casos, principalmente nas extremidades de linhas e em pontos onde há junções. Este problema de precisão é crítico em muitos casos e novos estudos ainda são necessários para contornar tais problemas de forma satisfatória.

O problema de conversão da representação de pixels em uma imagem binária para a forma vetorizada é abordado por etapas. O primeiro passo é a localização das linhas na imagem original, que geralmente é a fase em que, comumente, se utiliza um processo de obtenção do eixo medial. O passo seguinte é a aproximação da linha por um conjunto de vetores, o que é feito por uma aproximação poligonal restrita por diferentes tipos de critérios. O terceiro passo, que representa um pós-processamento, envolve refinar a representação vetorial e descrever de forma mais apurada as junções existentes e corrigir falhas como discontinuidades.

Uma imagem, que supostamente contém gráficos, precisa ser adequadamente processada para localizar o conjunto de linhas existente, cada linha caracterizada por uma cadeia de pixels. Para tanto, é razoável que se considere a base da linha como o eixo medial da forma (linha) presente na imagem *raster*.

O problema da esqueletização, que é a redução das linhas para que sejam representadas pelo seu eixo medial, apresenta uma questão crítica em relação às extremidades e junções, além da necessidade adicional, para os gráficos considerados, do tratamento de borrões causados pela pena sobre o papel, que é uma situação relativamente comum. Estes detalhes podem ser observados na Figura 1, (item 3.1, que descreve o objeto de estudo), em que há borrões gerados pelo espalhamento da tinta da pena e também pela movimentação rápida da pena, que descaracteriza a linha.

No caso dos gráficos, um interesse mais pronunciado está na obtenção do esqueleto da linha do gráfico. Assim, algoritmos como os baseados nas proposições de Rosin e West (1989) ou de Wall e Danielsson (1984) podem ser utilizados. No caso do primeiro, que é um algoritmo *split-and-merge*, há a vantagem de não necessitar de parâmetros ou limiares arbitrários; o segundo, por sua vez, é rápido e eficiente, embora necessite de um parâmetro de limiar e não proporcione tanta precisão quanto o primeiro. O método de Rosin e West (1989) tem como princípio a quebra da curva em segmentos cada vez menores, recursivamente, até que um critério de erro seja satisfeito ou

um comprimento mínimo seja atingido. Os pequenos segmentos definidos são posteriormente organizados para análise, mantendo aqueles que são relevantes, segundo um critério de significância, definido como a razão entre o desvio máximo e o comprimento do segmento. Wall e Danielsson (1984) propuseram um método iterativo que utiliza um único parâmetro, que é um limiar para a razão entre a superfície algébrica e o comprimento do segmento.

Conforme mencionado, uma primeira abordagem é fazer a esqueletização das linhas, ou seja, determinar seu esqueleto. Esta alternativa peca por poder potencialmente deixar “rebarbas” nas linhas afinadas e mover as junções e extremidades para posições não adequadas.

Existem duas abordagens básicas para vetorização com base na obtenção do eixo medial em imagens binárias: afinamento iterativo (por erosão) e baseado em transformada de distância (Lam et al., 1992). O afinamento iterativo é produzido pela remoção sucessiva de pixels da borda de uma forma, reduzindo sua espessura, mas mantendo sua estrutura de forma. O afinamento baseado em distância considera as linhas de elevação formadas pelos centros de todos os círculos máximos incluídos na forma original e correspondem, portanto, ao conjunto de pontos equidistantes das bordas da forma considerada. Outras referências importantes sobre métodos de afinamento podem ser vistas da revisão de Lam et al. (1992). A abordagem por afinamento por erosão foi a alternativa escolhida, dada a velocidade de processamento compatível com os gráficos utilizados e a qualidade do afinamento produzido.

Além da metodologia tradicional, que envolve a análise da imagem em dois passos (vetorização e pós-processamento), há que se ainda considerar alternativas de análise de gráficos em um único passo. Neste enfoque, o tratamento da imagem é feito diretamente, utilizando técnicas como a transformada de Hough para retas e arcos, análise de linhas retas diretamente sobre os pixels (Chiang et al, 1998) ou vetorização global de retas (Song et al, 2000). Uma visão geral sobre estas abordagens é apresentada por Song et al (2002).

3 MATERIAIS E MÉTODOS

O método foi delineado a partir de um objeto de estudo que requereu soluções específicas, relativas ao desenvolvimento do sistema, tratamento de imagens e estruturas finais para possibilitar análises e armazenamento de quantidade razoável de informações, conforme se segue.

3.1 Objeto de estudo

O desenvolvimento pretendido é feito empregando-se os registros impressos na Estação Climatológica atualmente instalada na Universidade Federal de São Carlos, campus São Carlos-SP, e operada pelo Instituto Nacional de Meteorologia (INMET). Esta é uma Estação Climatológica principal, atualmente com registros contínuos de precipitação, umidade e temperatura do ar, pressão atmosférica, velocidade, rajada e direção dos ventos. Esta estação funciona na cidade de São Carlos

desde 1951 e já teve seus dados pluviográficos analisados, conforme Barbassa (1991). Na Figura 1 ilustra-se um pluviograma. Não se conhece o número exato de gráficos existentes desta estação, mas pode-se estimar seu número, ainda que grosso modo. Considerando que há 114 dias com 0,1 mm ou mais por ano, conforme consulta ao INMET (2008), e um intervalo de tempo de 1951-2007, ou seja 57 anos, haveria um total de 6498 pluviogramas.

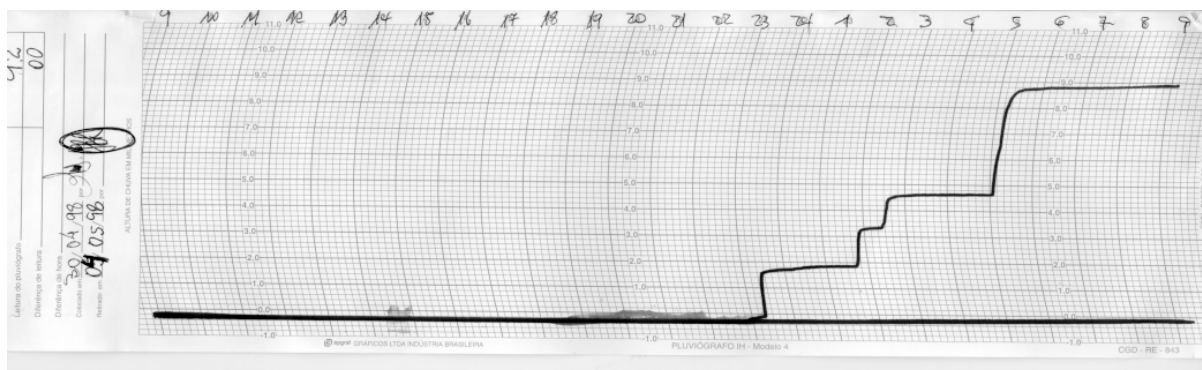


Figura 1. Pluviograma registrado na Estação Climatológica de São Carlos – SP.

Citam-se, a seguir, algumas particularidades dos gráficos dos pluviogramas:

- Há modelos com ou sem curvaturas na grade de fundo do gráfico.
- Apresentam como particularidades linhas “duplicadas”, linhas borradas, falhas do traço etc.

3.2 Desenvolvimento do sistema

O sistema computacional tem como finalidade implementar funcionalidades que permitam o processamento das imagens dos gráficos em papel e produzam como resultado a descrição vetorial dos dados contidos nos gráficos. Assim, os dados devem ser armazenados não somente como imagens dos gráficos originais, mas também formando um banco de dados que viabilize as análises climatológicas e hidrológicas.

O tratamento computacional que assiste o operador humano no processo de obtenção de dados envolve uma sequência de etapas. Como o tratamento é feito de forma digital, o primeiro passo é a digitalização dos registros impressos, o que usualmente é feito por um escâner. A posse da imagem digital que representa o gráfico permite, então, manipulá-la por meio de rotinas apropriadas e determinar as informações de interesse. Diversas técnicas podem ser empregadas para isolar, na imagem, o objeto de interesse, que no caso deste trabalho é a linha desenhada pela pena sobre o gráfico. A linha deve ser interpretada, então, para determinar os valores reais registrados sobre o papel.

3.2.1 Definição da plataforma de implementação

A programação do sistema requer versatilidade na implementação, velocidade de processamento, facilidade na estruturação de interfaces gráficas e facilidade na interação com um banco de dados. Em princípio, a utilização do sistema operacional Microsoft Windows, que possui grande difusão, e da linguagem de programação C#, também da Microsoft, foi definida. Ambos proporcionam um ambiente de desenvolvimento adequado ao projeto, segundo os critérios indicados.

Associada à escolha da linguagem para a implementação do sistema, há também a questão do projeto de uma interface gráfica adequada ao processamento de grande volume de dados e gráficos. Este ponto também é relevante na determinação da plataforma de desenvolvimento.

3.2.2 Digitalização das imagens dos gráficos impressos

A aquisição das imagens é usualmente feita por um escâner, o qual determina a resolução da aquisição (dada em pixels por polegada ou dpi, *dots per inch*) e a profundidade de cor (referente à quantização dos valores que determinam as cores na imagem). Tradicionalmente resoluções variam de 100 a 500dpi e a profundidade de cor utiliza cerca de 16 milhões de cores no formato RGB. Mas estas grandezas têm influência no tratamento da imagem. A digitalização do gráfico gera, então, uma imagem *raster*, que é aquela representada como uma matriz de pixels, cada um com um atributo de cor específico.

O sistema tem bom desempenho no tratamento de imagens com resoluções de até 200dpi e define tal resolução como a máxima sugerida. Tratamentos adicionais, porém, estão em estudo, o que permitirá que algumas das dificuldades encontradas nas imagens de maior resolução possam ser suplantadas.

3.2.3 Segmentação da imagem

Como as imagens em si não possuem qualquer informação sobre seu conteúdo além das cores dos pixels, um processo de segmentação precisa ser aplicado às imagens, de forma a separar os diversos componentes de interesse (Gonzales e Woods, 2007). Para que a segmentação seja realizada é preciso que haja um conhecimento prévio de seu conteúdo, com fim de se proceder ao tratamento adequado.

A segmentação da imagem, nesta etapa, corresponde à separação dos dados da imagem, ainda em formato *raster*. Os vários pixels da imagem devem ser classificados como pixels pertencentes aos objetos de interesse: a linha do gráfico, a grade pré-impressa do gráfico e o fundo (neste caso o papel). Um fator que os distingue de forma bastante interessante e dá subsídios para a segmentação é a cor. O papel é geralmente claro, com cor tendendo ao branco; as linhas de grade usualmente se apresentam mais claras, com cores marrom, verde ou laranja (entre outras);

finalmente a linha do gráfico é preta ou azul escuro. Mesmo em gráficos monocromáticos, há significativa diferença de intensidade entre as cores dos três objetos, embora informações sobre o matiz não existam.

Considerando-se, assim, que haja disponibilidade de imagens coloridas, uma abordagem mais abrangente pode ser utilizada, tendo-se optado por uma classificação utilizando o algoritmo de K-médias, configurado para três classes.

A classificação, quando realizada diretamente sobre as componentes espectrais RGB da imagem, que é usual na captura por escâneres, tende a não apresentar bons resultados, pois estas componentes não separam adequadamente as informações cromáticas das de intensidade luminosa. Foi feita, assim, a opção por um espaço alternativo, na qual esta separação pode ser obtida. Experimentalmente determinou-se que o espaço L^*a^*b se mostra adequado para representar os dados a serem classificados.

Certamente as condições do papel e do processo de aquisição das imagens em si acabam por introduzir pequenos problemas no processo de classificação. Em geral, alguns pixels acabam classificados incorretamente e passam a pertencer à linha do gráfico, introduzindo algum erro. Certa imprecisão em algumas regiões de transição de uma cor para outra ou em regiões onde a linha do gráfico e as linhas de grade acabam se mesclando são exemplo de áreas da imagem onde a classificação pode não ser totalmente correta. Este fato acontece por um aparente borramento causado pela tinta da pena sobre o papel, que mostra comportamento de espalhamento diferente quando há a presença da tinta das linhas de grade. Estas ocorrências, entretanto, são bastante locais e deixadas para serem tratadas pelas etapas posteriores da segmentação. Um tratamento por filtro da mediana tem se mostrado suficiente, na grande maioria dos casos para eliminar pequenas classificações incorretas.

3.2.4 Vetorização

O processo de vetorização deve ser realizado de forma extrair a informação impressa fidedignamente. A vetorização da linha do gráfico é um processo semi-automático, em princípio. Há intervenção do operador em vários momentos, como na identificação dos valores reais representados no gráfico e sua correspondência com os pixels da imagem. Questões como resolução da imagem, escala, rotações e translações também estão envolvidas.

Alguns destes processos podem ser automatizados, em maior ou menor grau. Assim, o sistema incorpora recursos para processamento autônomo da imagem, segundo o enfoque de minimizar o esforço do operador.

Nas situações em que a linha não apresenta borrões grandes e seja distinta de linhas próximas, esta vetorização pode ser feita automaticamente. Em outras palavras, o sistema consegue seguir a linha e extrair os dados de forma independente do operador. Há, porém, situações em que

linhas próximas, como as que ocorrem na situação de chuva intensa, na qual a pena faz diversas subidas e descidas no gráfico, tornam o eixo medial da imagem não representativo. Para estas situações, uma vetorização manual foi desenvolvida, permitindo que o operador interfira na determinação dos pontos de forma assistida pelo sistema.

A parte automática da vetorização é realizada pelo rastreamento da linha que compõe o eixo medial da linha do gráfico, que apenas indica o ponto inicial e o final do segmento a ser vetorizado. Este processo leva à obtenção dos dados do gráfico necessariamente dentro da linha e, portanto, corretas para este contexto. O algoritmo que analisa a linha do gráfico é capaz de ser auto-suficiente em algumas decisões, como é o caso, por exemplo, de corrigir pequenas interrupções na linha. Havendo, porém, problemas em que não seja possível determinar qual caminho a linha deve seguir (caso de bifurcações geradas por falhas no processo de segmentação), a vetorização automática é interrompida e a atenção do operador é novamente requerida.

O processo manual é simples e direto: o operador seleciona os pontos inicial e final de um trecho e o sistema determina a junção de ambos por uma linha direta. Na prática é feita a união por um segmento de reta, o qual deve ser adaptado para uma curva adequada à curvatura apresentada pelas linhas de escala do gráfico.

3.2.5 Armazenamento dos dados

A questão do armazenamento das informações é essencial para o sistema como um todo. A recuperação e processamento dos dados para a análise climatológica e hidrológica, como meta final do sistema, exige que haja uma organização pertinente.

Os dados são armazenados já na forma vetorizada, ou seja, valores de tempo e precipitação são registrados diretamente, após a conversão das coordenadas dos pixels da imagem para valores reais do gráfico.

3.2.6 Extração de informações e verificação de erro

Propôs-se desenvolver soluções para extração de séries de dados e permitir a recuperação de todos os parâmetros registrados em função do tempo para qualquer gráfico. Tem-se como premissa básica que os registros originais poderão, a qualquer tempo, ser recuperados, assim como os dados deles extraídos. Dessa forma, a qualquer tempo também, é possível verificar os erros.

4 RESULTADOS E DISCUSSÃO

O sistema Kairos foi desenvolvido na plataforma Microsoft C# executável em qualquer microcomputador com sistema operacional Windows. Apresenta-se neste item o fluxo de operações do sistema computacional Kairos, algumas aplicações e faz-se uma análise de suas potencialidades.

4.1 Diagramas de fluxo de operações

Inicialmente o usuário deve digitalizar o pluviograma impresso, o que pode ser feito com qualquer escâner, porém a velocidade dele será restritiva se o número de pluviogramas for grande. Neste caso, será necessário um equipamento rápido. Sugere-se definir a resolução de imagens entre 100 e 200dpi. Esta é a primeira operação, conforme fluxo geral mostrado na Figura 2.

O Kairos solicita ao usuário abertura da imagem que se deseja trabalhar. As imagens de pluviograma são, normalmente, numeradas automaticamente no momento do escaneamento, então é necessário identificá-la. A identificação de cada imagem é feita no início do seu processamento, pelo dia em que o pluviograma foi registrado na estação climatológica. Neste momento facultase ao operador entrar com todos os outros dados constantes do cabeçalho do pluviograma.

O fluxo de operações do sistema Kairos é mostrado também na Figura 2. Após abrir a imagem do pluviograma de interesse, solicita-se a entrada de três pontos de referência definidos pelas coordenadas: (1) tempo inicial, altura pluviométrica zero (origem do pluviograma); (2) tempo máximo e altura de chuva zero (máximo valor da abscissa) e (3) tempo zero e máxima altura (máxima altura de chuva). Estas informações são as bases para transformações de escala, correções de concavidade e correção de posição da imagem.

Disponibiliza-se no Kairos, após processar a imagem, a vetorização automática ou semi-automática da linha do pluviograma em segmentos ou totalmente. Esta decisão é tomada pelo operador ao analisar visualmente o pluviograma mostrado na tela. A vetorização automática possibilita percorrer trechos retos, curvas, quedas, mudanças de direção etc., uma vez definidos os pontos inicial e final do segmento. A vetorização manual é feita em segmentos de reta apenas, tantos quantos necessários, para transpor singularidades. Após cada segmento vetorizado é possível refazê-lo parcial ou totalmente. Exemplos de uso destes recursos são apresentados no item aplicação do Kairos.

Como última operação, executa-se o armazenamento do pluviograma original identificado, juntamente com pluviograma vetorizado, de modo que, a qualquer instante, seja possível verificar se a extração das informações foi suficientemente precisa e adequada ao uso que se venha a ter.

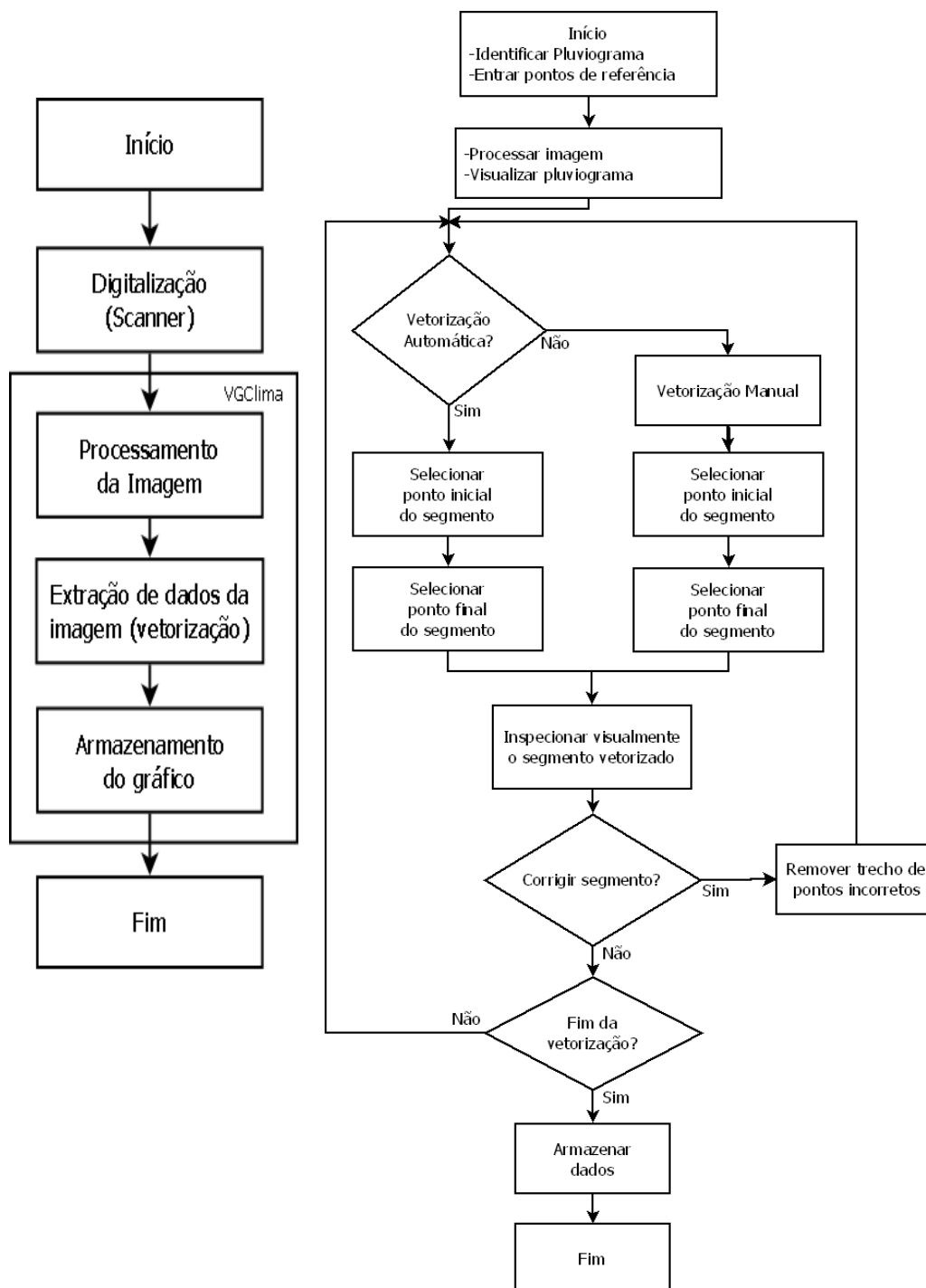


Figura 2 - Fluxo de operações geral e do sistema computacional Kairos.

4.2 Aplicação do Kairos

O Kairos foi projetado para extrair informações de pluviogramas com variados graus de dificuldades de interpretação, pois fornece recursos que facilitam a visualização da linha do gráfico, como se exemplifica neste item. Estas dificuldades estão relacionadas ao estado de conservação dos

pluviogramas impressos, da presença de borrões e de traçados indistintos que ocorrem, por exemplo, em chuvas intensas, pequenas falhas, entre outras singularidades.

Realizadas estas tarefas iniciais, processa-se a imagem pelo Kairos de forma a permitir a vetorização da linha traçada no pluviograma. O operador visualiza na tela o pluviograma carregado e decide se vai vetorizar automaticamente todo o pluviograma, ou automaticamente um segmento e manualmente outro. Caso característico de vetorização automática é mostrada na Figura 3, cujo pluviograma é “bem comportado”, sem borrões, sem discontinuidades ou linhas tão próximas que o traçado de uma emenda com o da outra. Os número 1, 2 e 3 da Figura 3 são os pontos de referência. A vetorização automática é a primeira disponibilizada, conforme fluxo de operações da Figura 2. Para isto é necessária a definição dos ponto inicial e final do segmento, que no caso, corresponde a todo o pluviograma.

Para possibilitar selecionar de forma segura os pontos de referência e também os pontos inicial ou final de um segmento, foi construída uma ferramenta de zoom para auxiliar esta tarefa. Ao clicar sobre qualquer ponto do pluviograma com a tecla shift pressionada, a área onde está o cursor do mouse é mostrada em uma quadrícula ampliada várias vezes. Este recurso é mostrado na Figura 4.

O Kairos corrige automaticamente falhas de tinta do pluviograma, falhas estas tão pequenas que não permitem desvio do traçado original do pluviograma.

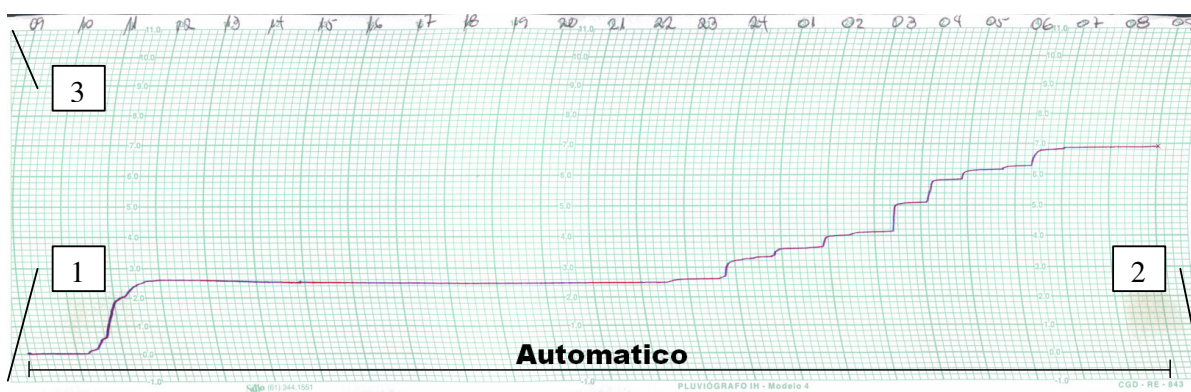


Figura 3– Pluviograma, sem cabeçalho, vetorizado automaticamente, com evidencia de pontos de entrada, com linhas original (azul) e vetorizada (vermelha) sobrepostas.

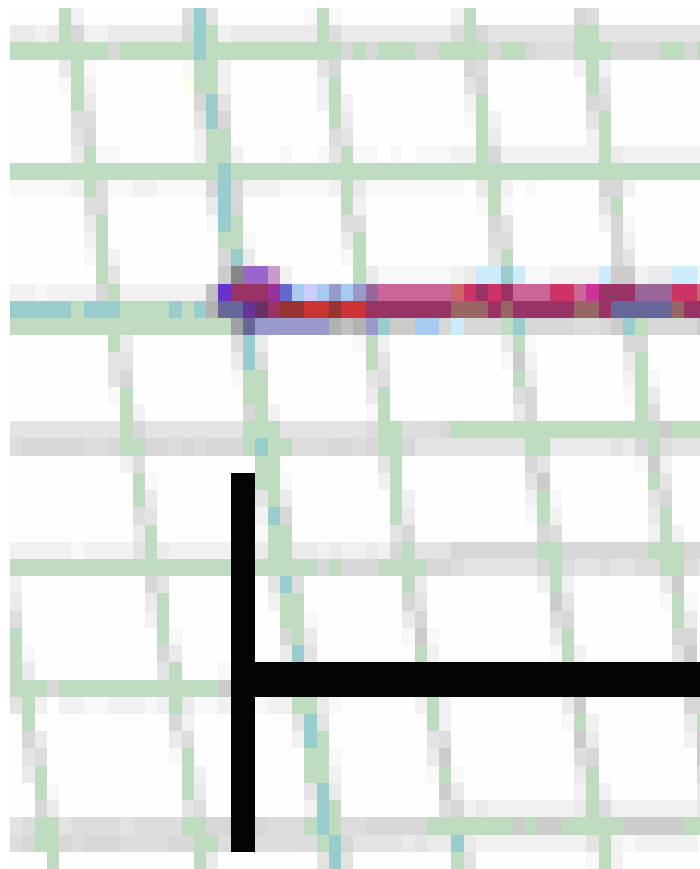


Figura 4- Quadrícula com zoom para entrada de pontos do pluviograma, obtido clicando shift e botão do mouse com o curso sobre o ponto de interesse.

No caso o pluviograma visualizado na tela pelo operador apresentar alguma singularidades, o operador pode optar por vetorização manual alternada com automática. É o caso de borrão no pluviograma, de forte chuva onde há picos no pluviograma, como mostrado na Figura 5. A vetorização manual foi exigida no segmento onde as linhas ascendentes e descendentes são muito próximas, indistintas, e nos segmentos de linha descendentes onde há falhas no traçado original. Neste caso a ferramenta de zoom é útil para facilitar a identificação dos picos, como ilustrado Figura 6.

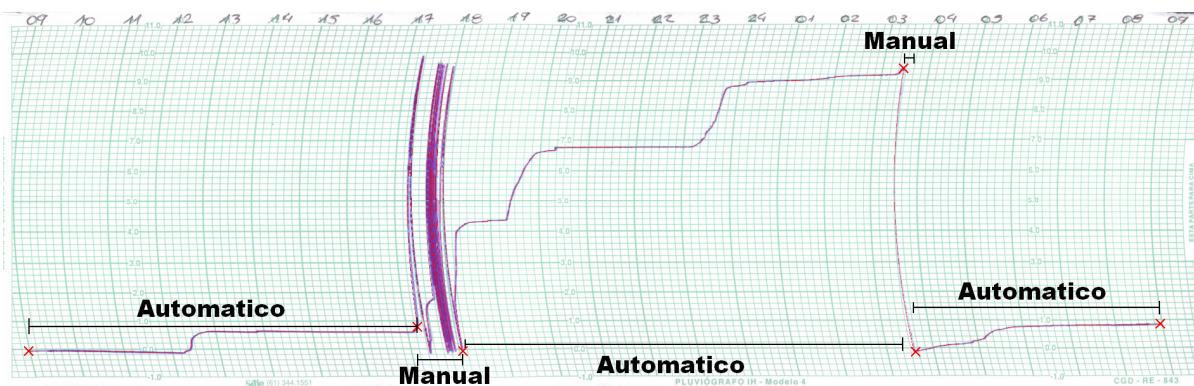


Figura 5– Pluviograma (sem cabeçalho). com segmentos vetorizados ora manual, ora automaticamente.

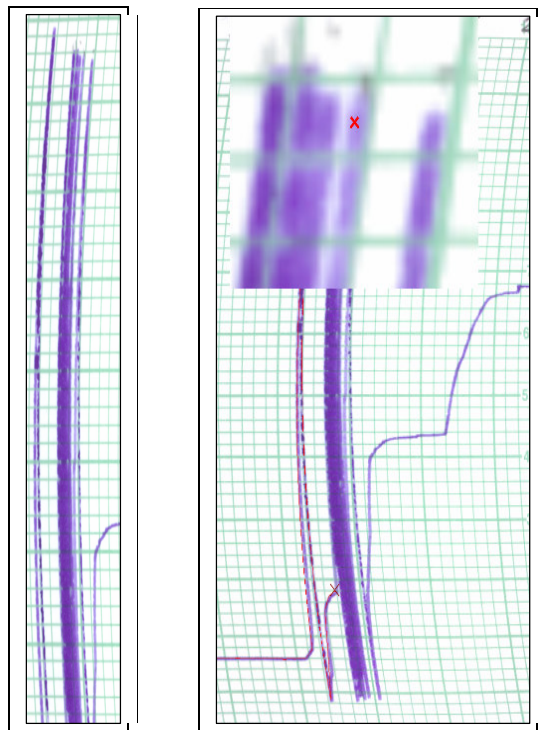


Figura 6 – Parte de pluviograma com linhas ascendentes e descendentes indistintos.

4.3 Precisão de extração de informações da imagem

Este é um aspecto da maior importância, pois a confiabilidade no funcionamento do Kairos depende da qualidade com que é feita a extração da informação da imagem e sua representação matemática. Considera-se satisfatória uma vetorização em que a linha vetorizada se sobreponha ao traçado original da pena do pluviógrafo. Há dois recursos que permitem ao operador avaliar se isto é satisfeito:

- A informação vetorizada é mostrada na imagem do pluviograma original à medida que esta é executada. Este recurso permite que o operador visualize se as linhas original e vetorizada se sobrepõem. Caso não sejam sobrepostas, é facultada a correção imediata, seja parcial ou totalmente. A Figura 3 e a Figura 5 são exemplos de uso deste recurso.

5 CONCLUSÕES

O sistema de extração de informações de gráficas, denominado Kairos, foi desenvolvido para vetorizar pluviogramas. O fluxo de operações apresentado e os testes em escala piloto executados permitem afirmar o seguinte:

- Consegue-se, em grande parte, extrair e representar matematicamente os registros impressos de forma automática, reduzindo a necessidade do operador ao mínimo.

- Facilita-se a leitura e interpretação do pluviograma pelo operador graças aos recursos de zoom de áreas específicas, como também a ampliação do tamanho real do pluviograma, os quais potencializam a visualização da imagem.
- Facilita-se a correção simultânea à vetorização, facultando refazer parcialmente ou totalmente a extração.
- Possibilita-se, pelo armazenamento das informações originais e da linha vetorizada, a confirmação da qualidade do trabalho do operador.
- Realiza-se a vetorização de forma praticamente instantânea, quando feita automaticamente. O tempo de processamento depende da presença de singularidades no gráfico, pois pode demandar a interferência do operador.

Considera-se que o Kairos, neste estágio de desenvolvimento, soluciona várias dificuldades de análise do pluviograma. As etapas seguintes serão as aplicações em larga escala e a incorporação de análises hidrológicas. Ao Kairos serão também incorporadas funcionalidades para vetorizar outros registros climatológicos impressos.

Agradecimentos: Agradecem-se ao Instituto Nacional de Meteorologia (INMET) pela acesso ao registros e a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pelo apoio financeiro.

Referências bibliográficas

- Barbassa, A. P. *Simulação do efeito da urbanização sobre a drenagem pluvial na cidade de São Carlos* -- Tese de Doutorado. 327 págs. (Hidráulica e Saneamento). SHS/EESC/USP, 1991.
- Castleman, K. R. *Digital image processing*. Prentice-Hall, 1995.
- Chiang, J.; Tue, S.; Leu, Y.. A new algorithm for line image vectorization. *Pattern Recognition*, 31:1541-1549, 1998.
- Chow, V.T. *Handbook of Applied Hydrology*. New York: McGraw Hill, 1964.
- FCTH. **Sistema de Informações para Gerenciamento de Recursos Hídricos do Estado de São Paulo**. São Paulo, DAEE, 2001. CD-ROM
- Gonzalez, R.C. & Woods, R.E. *Digital Image Processing*. Pearson Prentice Hall, 2007.
- Lam, L.; Lee, S.-W.; Suen, C.Y. **Thinning methodologies – a comprehensive survey**. IEEE Transactions on PAMI, 14(9): 869-885, 1992.
- Rosin, P.L. & West, G.A. **Segmentation of edges into lines and arcs**. *Image and Vision Computing*, 7(2): 109-114, 1989.
- Song, J.; Cai, M.; Lyu, M.R.; Cai, S. **Graphics recognition from binary images: one step or two steps**; *Proceedings of the International Conference on Pattern Recognition*, 2002.

- Song, J.; Su, F.; Tai, C.; Chen, J.; Cai, S. **Line net global vectorization: an algorithm and its performance evaluation.** *Proceedings of CVPR 2000*, Vol.2, 383-388, 2000.
- Tombre, K.; Tabbone, L. **Vectorization in Graphics Recognition: To Thin or not to Thin.** In *Proceedings of the 15th International Conference on Pattern Recognition*, Barcelona (Spain), volume 2, pages 91--96, Sept. 2000.
- Wall, K. & Danielsson, P. **A fast sequential method for polygonal approximation of digitized curves.** *Computer, Vision and Image Processing*, 28: 220-227, 1984.
- Wilken, P.S. *Engenharia de Drenagem Superficial*. São Paulo: CETESB. 1978. 232p.