

KAIROS: Processamento e Análise de Pluviogramas

*Ademir Paceli Barbassa¹; Jander Moreira²; Sidarta Fernandes³;
Mauricio Aparecido Rodrigues Silva Barbalho⁴*

Resumo – Apresenta-se neste artigo o Sistema Kairos para vetorização de pluviogramas. O sistema possui diversas funcionalidades como vetorização automática e semi-automática, detecção e correção de erros durante a execução do programa e pós-execução. As formas de verificação de falhas, além da visual, são por comparação entre as alturas calculadas pelo Kairos e as registradas manualmente no pluviograma, além de calcular o percentual de superposição de linhas. Os tempos despendidos para vetorização, em média, são estimados em 1,1 minuto para o modo automático e 1,7 minuto para o modo manual por pluviograma. O Kairos permitiu identificar falhas no traçado do pluviograma, pluviogramas com erros na altura registrada manualmente pelo leitorista, falhas do operador do Kairos e diferenças no casos de alturas pequenas ou nulas.

Abstract– This paper presents Kairos, a computational system designed for pluviogram vectorization. The software has several functionalities, such as automatic and semi-automatic vectorization, and also detection of flaws, both during execution and in post-execution times. Along with the inspection of the vectorization, flaws can be detected by the comparison between the precipitation registered on the pluviogram and that calculated from the vectorized data. The percentage of superposition of the lines are also available for analysis. The mean processing time for the vectorization a a single pluviogram has been estimated in 1.1 minutes when in full-automatic mode and 1.7 minutes when human intervention is required (semi-automatic mode). The Kairos System allows spotting abrupt interruptions in the line on the graph, mis-registrations manually noted on the pluviogram, mistakes made by the Kairos operator during vectorization, and differences when registered heights are small or even null.

Palavras-chaves - vetorização, gráficos climatológicos, pluviogramas

1 INTRODUÇÃO

Há, no Brasil, uma grande dificuldade de acesso a dados climatológicos, causada tanto pela indisponibilidade de dados dessa natureza quanto pelo precário armazenamento. Somados a estes fatores estão a descontinuidade da coleta, a mudança de prioridade e mesmo descaso para com os registros em diferentes administrações. Mesmo com tais restrições, há coletas disponíveis, sendo estas compostas por grande volume de registros em papel, coletados por décadas. Atualmente este processo de coletas está sendo substituído por registros digitais.

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 Graduando do Curso de Bacharelado em Ciência da Computação, UFSCar, Rod. Washington Luis, Km 235, 13566-260, São Carlos, SP; sidafer.22@gmail.com

⁴ Graduando do Curso de Engenharia de Computação da UNICEP, Rua Miguel Petroni, 5111, 13570-300, São Carlos, SP; maubarbalho@hotmail.com

A dificuldade de lidar com os registros em papel levou ao desenvolvimento de um software para a representação vetorial de dados registrados em pluviogramas impressos. O sistema Kairos é um sistema computacional semi-automático para extração de dados de pluviogramas, análises hidrológicas e climatológicas. Por meio deste sistema foi realizado o resgate de registros impressos da Estação Climatológica operada pelo Instituto Nacional de Meteorologia (INMET) e instalada na Universidade Federal de São Carlos (UFSCar) e também de registros do INMET de Brasília.

O software permitiu extrair e representar matematicamente os registros impressos de forma automática, reduzindo substancialmente o esforço do operador. Facilitou a leitura e interpretação do pluviograma pelo operador, dados recursos como de zoom de áreas específicas e ampliação do tamanho real do pluviograma, potencializando a visualização da imagem. Viabilizou correção simultânea à vetorização, facultando refazer parcialmente ou totalmente a extração de dados. Possibilitou, também, a confirmação da qualidade do trabalho do operador pelo armazenamento das informações originais e da linha vetorizada. Finalmente, permitiu a vetorização de forma praticamente instantânea, quando feita automaticamente, com tempo de processamento total dependente da presença de singularidades no gráfico, caso que pode demandar a interferência do operador.

O critério usado para avaliar a qualidade de uma vetorização é simplesmente avaliar se a linha vetorizada se sobrepõe ao traçado original da pena do pluviógrafo, o que é verificado automaticamente pelo sistema. O Kairos solucionou várias dificuldades de análise do pluviograma.

Este trabalho apresenta as mais recentes evoluções do Kairos, em relação a incrementos na interface e nas análises que o sistema é capaz de realizar. Analisou-se a performance do software com base na vetorização de 4326 pluviogramas, exemplificando como o Kairos está programado para identificar inexistência e erros de informações escritas do pluviogramas, erros inerentes ao processo de registro e deformação do papel e falhas do operador durante o processo de vetorização.

2 CARACTERIZAÇÃO DO SISTEMA KAIROS

O desenvolvimento de software para captura de pluviogramas iniciou-se em 1990, com o uso de mesa digitalizadora para vetorização de pluviogramas. Foram capturados 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).

Neste trabalho dá-se continuidade no desenvolvimento do Kairos, apresentado em Barbassa et al (2009). Naquele artigo são discutidas as características, vantagens e limitações da vetorização de pluviogramas via mesa. Apresentam-se também revisão sobre o processamento e análise de imagens, a segmentação de imagens, o reconhecimento de gráficos e a sua vetorização.

O Kairos é um sistema dedicado à extração de dados de leitura a partir de pluviogramas impressos digitalizados. O sistema assiste o operador humano no processo de obtenção de dados, que envolve uma sequência de etapas, iniciando-se com 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 são 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.

A aquisição das imagens é usualmente feita por escâneres, os quais tradicionalmente possuem resoluções variam de 100 a 500 dpi e profundidade de cor de cerca de 16 milhões de cores no formato RGB. O sistema tem bom desempenho no tratamento de imagens com resoluções de até 200dpi e define tal resolução como a máxima sugerida, visto que resoluções maiores não parecem acrescentar informações significativas em termos de precisão e, portanto, não trariam efeitos vantajosos para a qualidade dos dados obtidos.

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). A segmentação da imagem corresponde à separação dos dados da imagem. 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 sistema considera, assim, imagens coloridas, optando-se por uma classificação utilizando o algoritmo de K-médias, configurado para três classes. Experimentalmente determinou-se que o espaço L^*a^*b se mostra adequado para representar os dados a serem classificados.

O processo de vetorização é realizado de forma extrair a informação impressa fidedignamente. A vetorização da linha do gráfico é um processo semiautomático, 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.

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.

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.

Uma questão essencial é a do armazenamento das informações. 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. Além da “leitura” do gráfico, são armazenados dados importantes, como data, operador, registros feitos no papel etc.

O fluxo de operações do sistema Kairos é mostrado também na Figura 1. 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.

Ilustra-se pelo fluxograma, figura 2, a utilização do Kairos Visualizador, que é uma ferramenta com funções interessantes tanto para análises pluviométricas, quanto para detecção e visualização de erros. A princípio, como mostrado, há diversas funcionalidades que o usuário pode escolher.

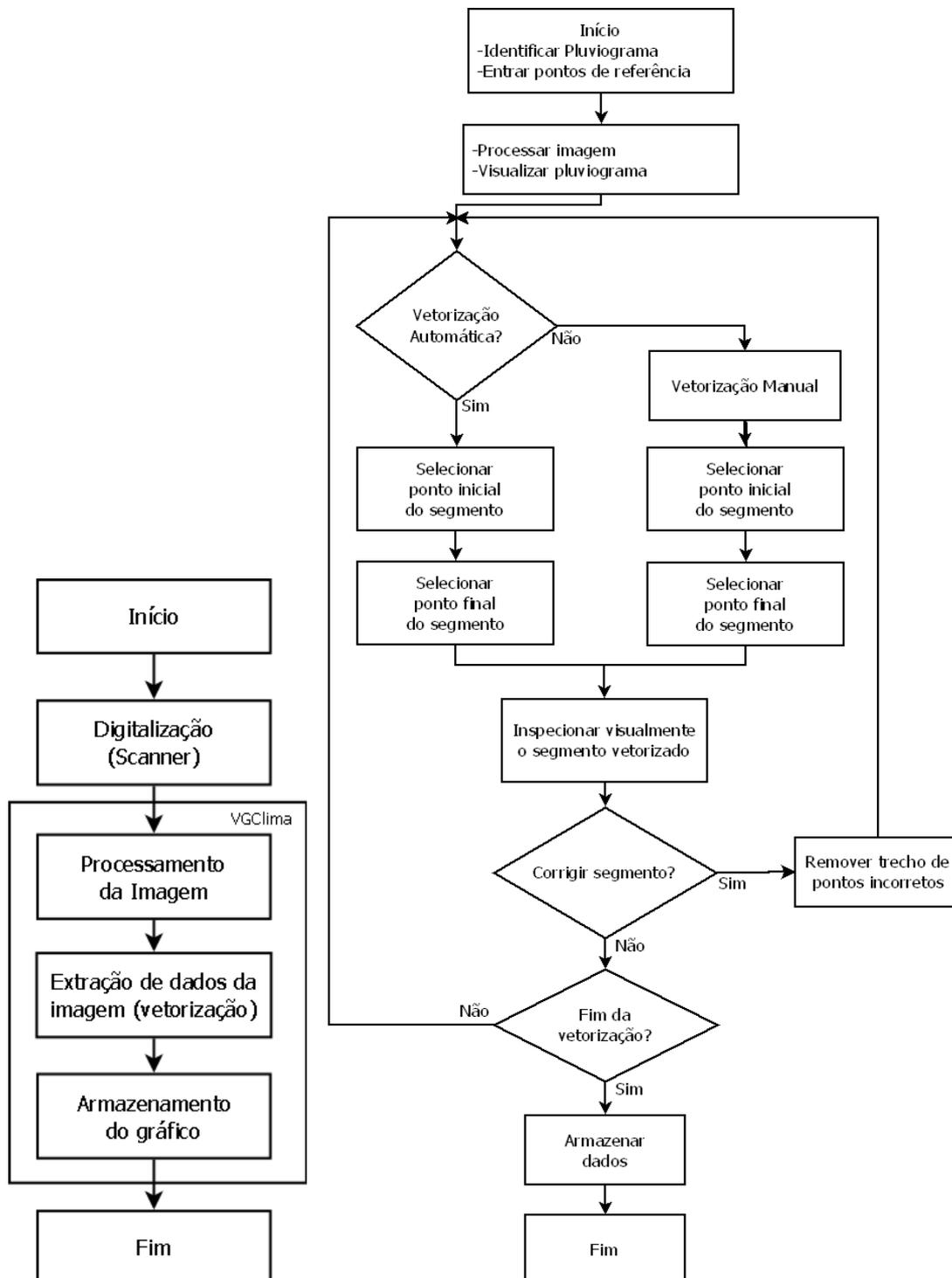


Figura 1 - Fluxo de operações e do sistema computacional Kairos.

A primeira funcionalidade é uma importante ferramenta de análise visual de erros, pois permite observar a imagem do pluviograma, com sua linha vetorizada traçada por cima, tornando

possível detectar imperfeições na vetorização. Em seguida, o programa oferece o histograma e o histograma acumulado para as chuvas calculadas do pluviograma aberto, ferramenta útil na análise e compreensão das chuvas. É importante mencionar que nessas duas abas, hietograma e hietograma acumulado, é permitido ao usuário exportar esses dados na forma de lista de pontos (intervalos de tempo, acompanhados de qual valor de chuva foi registrado em sua coordenada), ou de lista de pontos acumulado (idem, porém com valores acumulados de chuva), respectivamente.

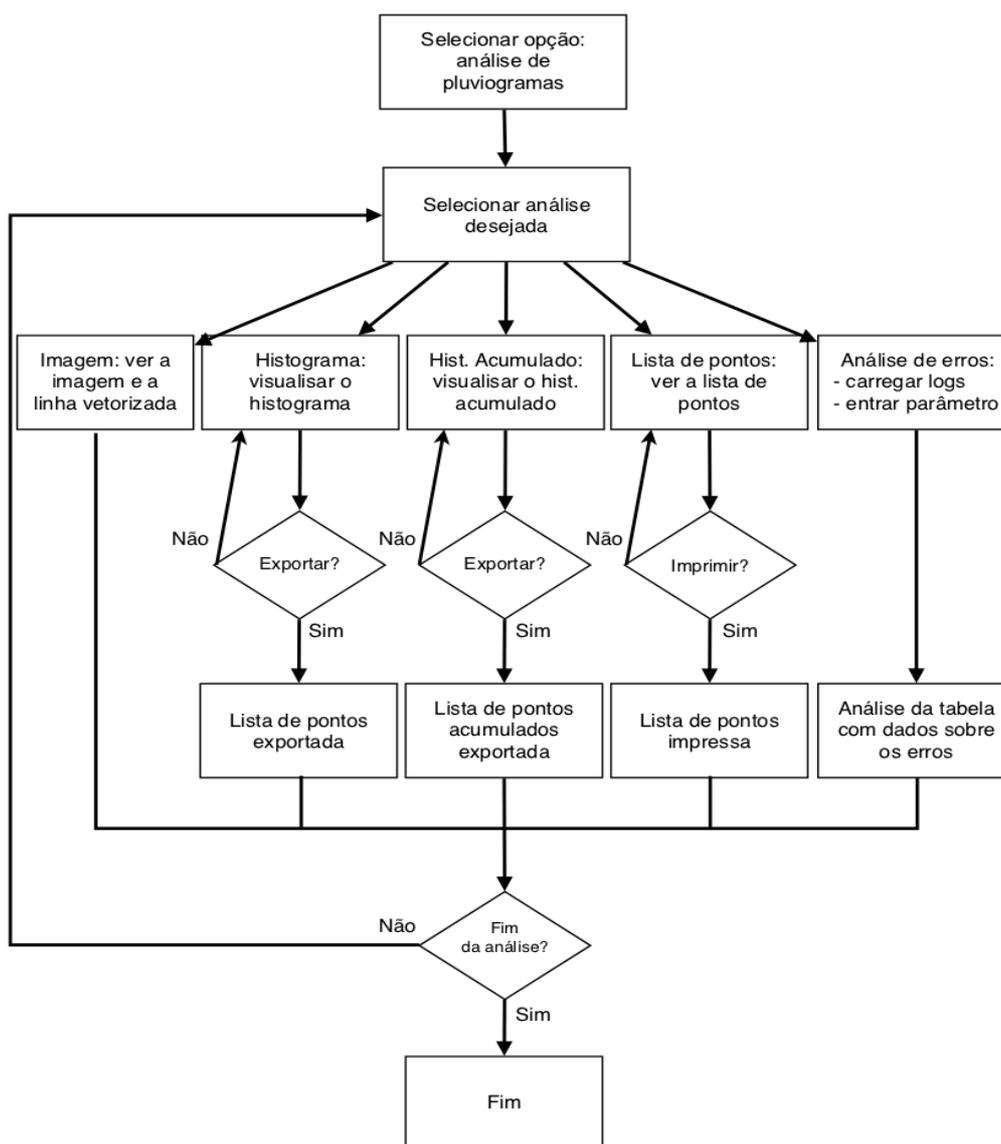


Figura 2 – Fluxograma com as funcionalidades do Kairos Visualizador

Também é permitido ao usuário, em outra funcionalidade do Kairos Visualizador, ver essa lista de pontos, citada anteriormente. E por último, uma função que foi implementada recentemente, permite uma análise profunda dos erros presentes na vetorização de pluviogramas. Nesta, o usuário deverá carregar todos os arquivos de log (arquivos com o mesmo nome da imagem, terminado em

.txt) referentes a quais pluviogramas deseja analisar. Em seguida, ele pode, se desejar, inserir um parâmetro para filtrar os dados a serem mostrados; este parâmetro deve ser um valor, para o qual serão mostrados todos os pluviogramas que apresentaram um valor maior de $| H_{\text{kairos}} - H_{\text{pluviograma}} |$. Caso não insira parâmetros, todos os pluviogramas serão incluídos no resultado. Esses pluviogramas a serem mostrados, aparecem em uma tabela, contendo todos os dados pertinentes a análise, como por exemplo: altura registrada pelo pluviômetro, altura registrada pelo Kairos e o erro em porcentagem, dentre outros. É com essa última funcionalidade que há a possibilidade de uma análise da qualidade de vetorização, e de erros presentes nas imagens, em larga escala e de maneira simples e objetiva.

2.1 Vetorização de modelos distintos de pluviogramas

A metodologia usada para análise das imagens permite a vetorização de pluviogramas com diferentes geometrias de sua grade de fundo, variadas cores, porém necessita realizar adaptações para reconhecimento automático para alguns tipos. Atualmente vetorizam-se automática ou manualmente os pluviogramas mostrados na figura 3.

2.1.1 Formas detecção e apresentação de “erros” pelo Kairos

O Kairos apresenta uma listagem com vários cálculos de “erros”, através dos quais podem-se detectar falhas. Pode-se executar o ordenamento de quaisquer “erros” do maior para o menor. Esta listagem pode ser confeccionada para tantos pluviogramas quantos se queira. A forma de apresentação é ilustrada na figura 4, com os seguintes cálculos:

- a) $(H_k - H_{ph})/H_{ph}$ (%) – este erro percentual entre a altura de chuva calculada pelo Kairos e a altura registrada manualmente pelo operador no pluviograma.
- b) Pontos vetorizados (%) – apresenta o percentual de pontos vetorizados sobrepostos à linha impressa no pluviograma. 100% indica total sobreposição.
- c) $H_{\text{pluviômetro}}$ – indica a altura medida no pluviômetro e a altura registrada manualmente pelo operador da estação climatológica
- d) $H_{\text{pluviôgrafo}}$ – indica a altura medida no pluviôgrafo e a altura registrada manualmente pelo operador da estação climatológica
- e) H_{Kairos} – refere-se a altura de chuva calculada pelo Kairos a partir da vetorização realizada.
- f) $H_{\text{kairos}} - H_{\text{pluviôgrafo}}$ – é a diferença em milímetro entre a altura de chuva calculada pelo Kairos e a altura de chuva registrada no pluviograma pelo operador.
- g) Idem item(f), com a diferença que o valor erro é absoluto

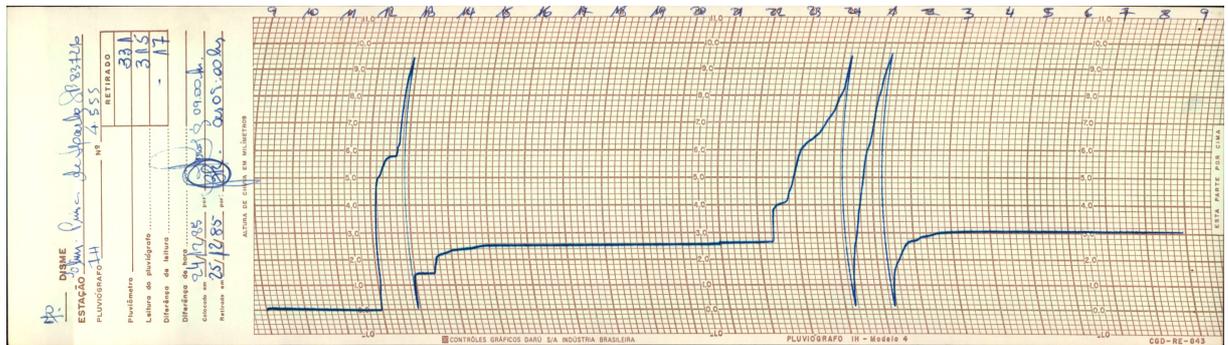
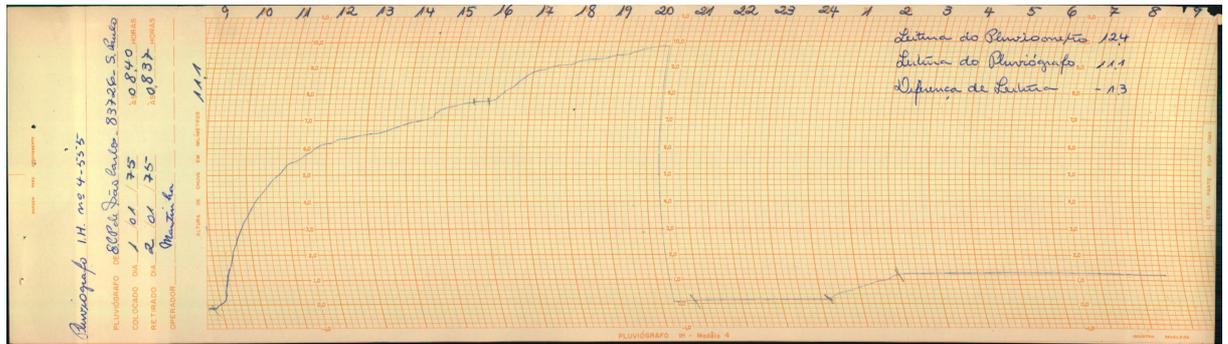
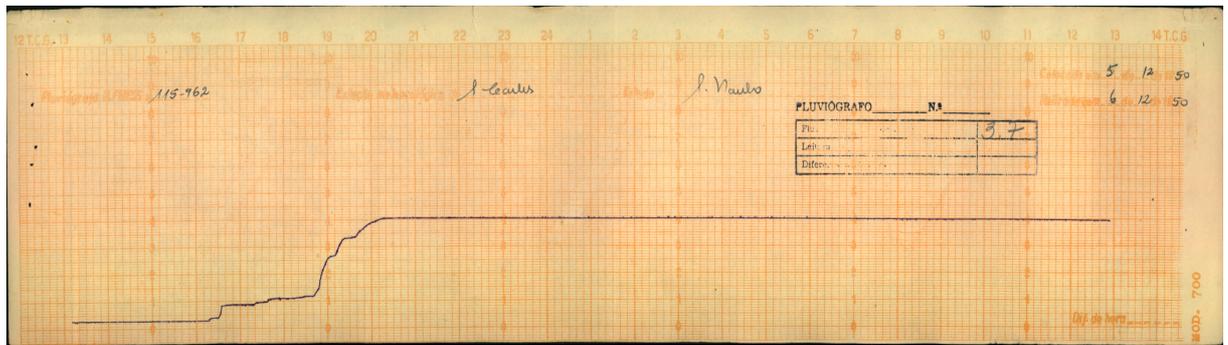


Figura 3 – Tipos de pluviogramas aceitos pelo Kairos no estágio atual de desenvolvimento

Nome do Arquivo	(Hk - Hpg) / Hpg (%)	Pontos Vetorizados (%)	Hpluviometro (Hpm)(mm)	Hpluviografo (Hpg)(mm)	Hkairos (Hk)(mm)	Hkairos-Hpluviografo (mm)	Hk - Hpg (mm)
C:\Users\user\D...	0,29	99	4,9	3,5	3,49	-0,01	0,01
C:\Users\user\D...	0,32	56	11,9	12,4	12,36	-0,04	0,04
C:\Users\user\D...	0,5	72	2,6	2	2,01	0,01	0,01
C:\Users\user\D...	0,5	48	39,5	28	27,86	-0,14	0,14
C:\Users\user\D...	0,52	80	8,1	7,7	7,74	0,04	0,04
C:\Users\user\D...	0,57	87	5,4	5,3	5,27	-0,03	0,03
C:\Users\user\D...	0,83	70	14,5	13,3	13,41	0,11	0,11
C:\Users\user\D...	0,86	71	15,1	14	14,12	0,12	0,12
C:\Users\user\D...	1,1	46	19,8	18,2	18,4	0,2	0,2
C:\Users\user\D...	1,37	70	11,3	11,7	11,86	0,16	0,16

Figura 4 – Tela com as diversas medidas de erros para todos os pluviogramas vetorizados

3 DETECÇÃO E CORREÇÃO DE FALHAS PELO KAIROS

O Kairos foi testado empregando os pluviogramas impressos utilizados 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).

Foram vetorizados 4326 pluviogramas a partir dos quais se exemplifica como o Kairos detecta, corrige e aponta falhas na linha impressa pelo pluviógrafo, falhas de informação manuscritas no pluviogramas, falhas do operador do Kairos, deformações no papel e na impressão dos pluviogramas, como se segue.

a) *Falhas na linha do traçada no pluviograma impõe dificuldade ao operador do Kairos saber qual o real traçado da linha.* Esta situação, ilustrada na figura 5, gera diferenças entre o que foi registrado manualmente no pluviograma e o calculado pelo Kairos, o que permite identificar o problema.

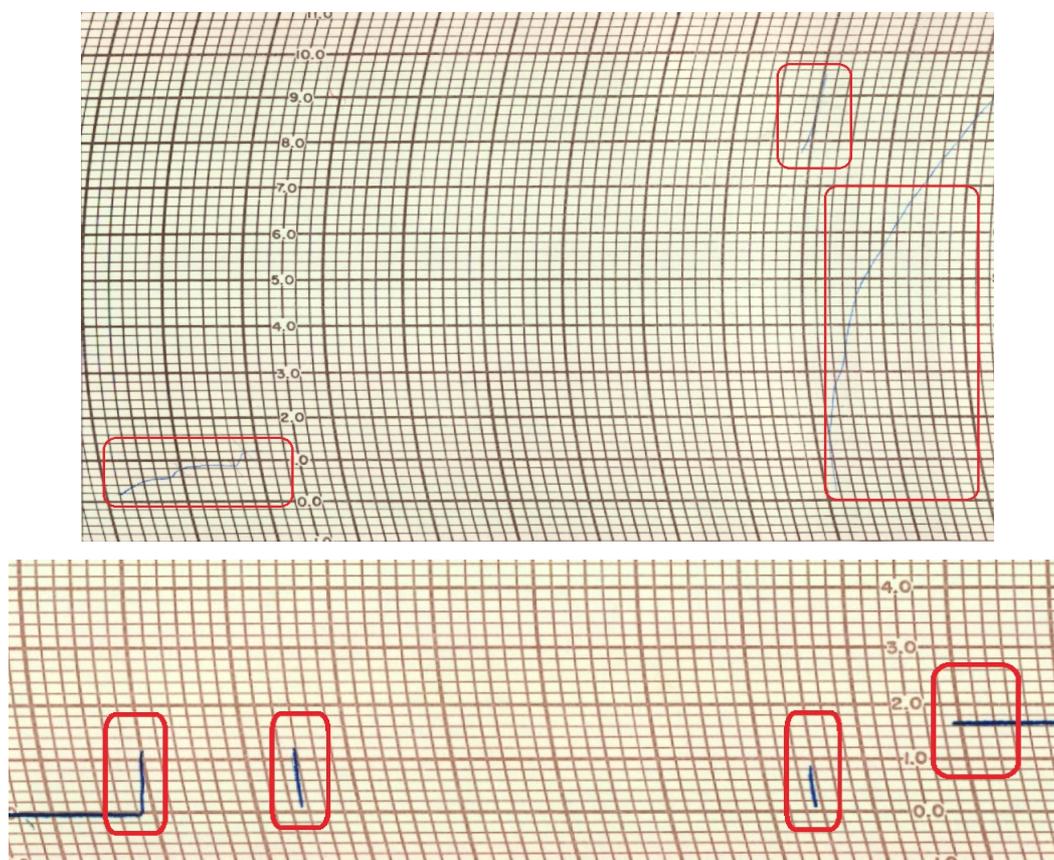


Figura 5 – Exemplos de trechos de pluviogramas com descontinuidades ou falhas na linha.

b) *Pluviograma com altura de chuva registrada manualmente no papel com erros ou inexistente.* Esta informação é registrada pelo leitorista que trabalha na Estação Climatológica, conforme se ilustra nos recortes da figura 6. Na ausência desta altura de chuva ou no caso de erro na altura de escrita no pluviograma (manualmente calculada), o Kairos indica o problema através da diferença entre a altura calculada pelo Kairos e a altura escrita erroneamente ou não escrita. O número de pluviogramas com estes problemas correspondeu a 113 (2,62%).



Figura 6 – Recortes de pluviogramas com inexistência ou falha nas alturas registradas nos pluviogramas pelo operador da Estação Climatológica.

c) *Falha do operador do Kairos na vetorização.* Esta situação pode ocorrer quando o Kairos está no modo manual de vetorização, conforme figura 7. Como nestes segmentos não há sobreposição entre a linha vetorizada e a linha no pluviograma, então o número de *pixels* em que não há sobreposição

se eleva, indicando estas ocorrências. Isto pode ser observado por análise visual. Dado que Kairos não possui funcionalidade para análise por intervalos menores, sendo os cálculos realizados para todo o pluviograma, não é possível detectar estas situações de forma automática.

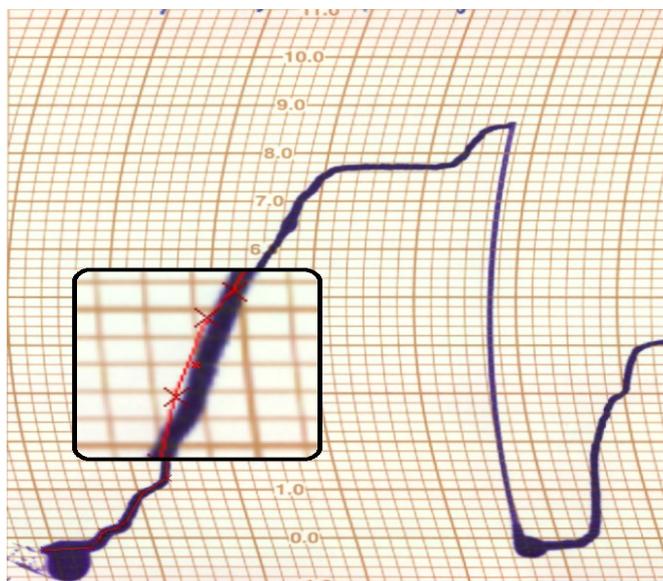


Figura 7 – Exemplo de falha do operador do Kairos na vetorização manual de pluviograma.

d) *Alturas de chuvas pequenas ou nulas*. Neste caso, o operador da estação verifica que não houve chuva, seja pela avaliação visual do pluviograma, seja por falha na interpretação. Entretanto devido à variação na linha do gráfico, por menor que seja, o Kairos detecta presença de precipitação, mesmo que pequena. Daí ocorre a detecção de “erro” ou diferença entre as alturas do Kairos e a registrada pelo operador. O número de pluviogramas com este tipo de falha atingiu 430, correspondendo a 9,95% dos 4326 pluviogramas vetorizados.

e) *Registro pelo leiturista de alturas não nulas sem reconhecimento da chuva pelo Kairos*. Há também o caso em que o valor da chuva registrada pelo operador foi maior que zero, porém o Kairos calculou valor de altura igual a zero. Estes erros correspondem a 43 pluviogramas (1%). Entretanto estas diferenças são inexpressivas, pois a altura precipitada é, geralmente, menor que um milímetro.

4 TEMPO GASTO PARA VETORIZAÇÃO

O tempo de vetorização depende da complexidade e da qualidade do pluviograma. No modo automático, incluindo todo o fluxo de trabalho (desde o carregamento da imagem até que os dados vetorizados sejam gravados em arquivo), o tempo estimado é de 1,1 minuto. Para gráficos com

singularidades, nos quais se exige a intervenção do operador do Kairos e uso do modo manual, o consumo de tempo é maior, tendo sido estimado em 1,7 minuto em média.

Registros mais apurados quanto ao tempo de vetorização de cada pluviograma serão realizados, permitindo estatísticas mais precisas quanto ao tempo consumido.

5 CONCLUSÕES

O sistema de vetorização Kairos foi testado em 4326 pluviogramas, sendo aprimorado nos seguintes aspectos:

- Possibilita a detecção de falhas pelo operador do Kairos e sua correção durante o processo de vetorização.
- Possibilita também que erros sejam detectados pós vetorização. Isto é feito pela comparação entre os valores de alturas pluviográfica e pluviométrica registrada nos pluviogramas manualmente e a altura computada pelo Kairos.
- Verifica também a sobreposição de linhas vetorizadas e linha impressa no pluviograma.

Desta forma sistema Kairos aponta erros humanos tanto do operador do Kairos quanto do operador da Estação Climatológica. Aponta também falhas e problemas de qualidade do pluviograma.

O Kairos está habilitado a vetorizar dois tipos pluviogramas, porém a metodologia deve ser estendida a mais modelos.

O tempo de vetorização médio tem sido estimado em 1,1 minuto para gráficos vetorizados automaticamente e em 1,7 minuto naqueles que exigem intervenção do operador.

Como continuidade resta adequação do Kairos para outros tipos de pluviogramas. Também necessita testar o cálculo automático de altura de chuva para diferentes durações.

Há ainda a possibilidade de vetorização de outros gráficos pluviométricos como barograma, anemograma e termohigrograma.

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.