

Bài báo khoa học

DigiMeto: Công cụ trợ giúp đọc giản đồ tự ghi Khí tượng thủy văn

Phạm Lê Phương^{1*}, Lê Trung Hưng¹, Mạnh Xuân Huy², Triệu Văn Linh³, Nguyễn Xuân Hoài⁴, Hải Hồng Phan⁵, Vũ Hải⁶

¹ Trung tâm Thông tin và Dữ liệu KTTV; phuongpl80@gmail.com; hungtttl@gmail.com

² Trường Đại học Bách Khoa Hà Nội; huy.mx161818@sis.hust.edu.vn

³ Viện AI Academy Việt Nam; nxhoai@gmail.com; trieuvanlinh04101998@gmail.com

⁴ Học Viện Kỹ thuật Quân Sự; hongpth@lqdtu.edu.vn

⁵ Viện Điện tử – Viễn Thông, Trường Đại học Bách Khoa Hà Nội; hai.vu@hust.edu.vn

* Tác giả liên hệ: phuongpl80@gmail.com; Tel.: +84-912353680

Ban Biên tập nhận bài: 14/3/2021; Ngày phản biện xong: 23/4/2021; Ngày đăng bài: 25/5/2021

Tóm tắt: Bài báo trình bày phần mềm DigiMeto (Digitizing Metontological Graph) hỗ trợ tự động đọc một số giản đồ tự ghi khí tượng thủy văn (KTTV) như giản đồ mưa, âm, nhiệt độ và nước. Để xây dựng công cụ, các kỹ thuật xử lý ảnh và nhận dạng đã được áp dụng để có thể tự động tách đường ghi dữ liệu ra khỏi giản đồ. Dựa vào thông tin đã biết về cấu hình ô lưới và khoảng thời gian, giá trị tương ứng tại mỗi thời điểm sẽ được xác định. DigiMeto cho phép người dùng hiệu chỉnh các số tham số một khi kết quả tự động có sai sót. Kết quả là công cụ có thể tự động xuất giá trị của giản đồ trong mỗi ngày với mức độ chi tiết đến từng phút. Trong các đánh giá cho thấy công cụ chỉ yêu cầu thời gian dưới 3 phút cho toàn bộ các thao tác trên mỗi giản đồ. Độ chính xác đạt được ở các giản đồ âm, nhiệt trung bình có sai số dưới 2% so với giá trị đo do người thực hiện. Với các giản đồ phức tạp ghi giá trị do nhiều hơn một ngày thường có sai số lớn hơn. Một số hướng phát triển đã được thảo luận và đề xuất để công cụ có thể được tự động hóa hoàn toàn, giúp giảm thiểu hơn nữa các thao tác người dùng cuối.

Từ khóa: Giao diện đồ họa (GUI); Giản đồ tự ghi; Xử lý Ảnh; Tách biên.

1. Mở đầu

Dữ liệu KTTV của Việt Nam đã có lịch sử trên 100 năm. Việc quan trắc KTTV đã được tiến hành từ thời nhà Nguyễn (đầu thế kỷ 18). Đến thời Pháp thuộc đã có một số trạm quan trắc mà số liệu còn lưu trữ đến ngày nay tại Tổng cục KTTV. Sau năm 1954, công tác điều tra cơ bản KTTV của Việt Nam được phát triển một cách hệ thống. Trong số các tài liệu quan trắc tại các trạm KTTV, các loại giản đồ chiếm số lượng lớn như giản đồ mưa, âm, nhiệt, nước. Giản đồ hiện nay chiếm khoảng 1/3 khối lượng tài liệu lưu trữ, bảo quản tại Trung tâm Thông tin và Dữ liệu khí tượng thủy văn. Nhu cầu số hóa các loại giản đồ này ngày càng trở nên cấp bách. Một mặt, việc tổ chức lưu trữ thủ công như hiện nay rất khó kiểm soát và gây khó khăn trong việc tìm kiếm tài liệu. Mặt khác, việc số hóa tài liệu lưu trữ giúp bảo vệ tài liệu lưu trữ gốc khỏi bị hủy hoại do tác động của các yếu tố lý hóa. Đồng thời việc số hóa giúp tăng cường khả năng tiếp cận tài liệu của công chúng; làm cho quá trình khai thác số liệu KTTV hiệu quả hơn. Đây là những động lực chính để chúng tôi phát triển một công cụ phần mềm ứng dụng các kỹ thuật xử lý ảnh và nhận dạng tiên tiến để bước đầu thực hiện việc số hóa một số loại giản đồ phổ biến đang được lưu trữ tại Trung tâm thông tin và dữ liệu

KTTV Quốc gia. Việc xây dựng công cụ này nhằm đưa ra hướng giải quyết khả thi, phù hợp với nhiều loại giản đồ, chất lượng giản đồ và nguồn nhân lực cũng như quy trình bảo quản tài liệu số hóa các loại giản đồ KTTV tự ghi tại Việt Nam.

Vấn đề số hóa một số loại giản đồ khí tượng thủy văn đã được một số nhóm nghiên cứu trên thế giới quan tâm và phát triển các công cụ có giao diện đồ họa hỗ trợ. Nghiên cứu [1, 2] phát triển công cụ có tên là Pluviograph, PluvioReader tự động đọc các loại giản đồ nước. Trong các nghiên cứu này đã trình bày các kỹ thuật tiền xử lý, xác định vùng quan tâm trên giản đồ, kỹ thuật phân tách, kỹ thuật hậu xử lý và xuất kết quả đọc. Công cụ hỗ trợ các chế độ tự động toàn bộ và tự động từng phần. Trên Pluviograph, kết quả độ chính xác trên một số dữ liệu thử nghiệm đạt sai lệch 0,1778 mm (0,0152) với chế độ tự động và 0,1600 mm (0,0136) với chế độ có tương tác người dùng cuối. Công cụ NUNIEAU trong [3, 4] được phát triển để số hóa giản đồ mưa một số trạm khí tượng tại Pháp. Ngoài việc hỗ trợ các module xử lý ảnh và scan thông thường, công cụ NUNIEAU hỗ trợ các giản đồ tham chiếu, để có thể so sánh, đánh giá từ kết quả số hóa giản đồ. Thay vì đọc các giá trị trực tiếp trên giản đồ, NUNIEAU hỗ trợ các cách quan sát khác nhau để so sánh dữ liệu hiện thời với các giản đồ tham chiếu để đánh giá sự sai khác. Công cụ NUNIEAU cũng cho thấy các tác giả đã quan tâm đến các giản đồ với đặc điểm về kích thước, ô lưới khác nhau vì điều kiện ghi tại mỗi trạm KTTV là khác nhau. Ngoài ra, gần đây sản phẩm GScanth là công cụ cho phép số hóa các giản đồ nhiệt là kết quả của nghiên cứu số [5, 6]. Công cụ này hỗ trợ các bước chính như scan giản đồ, tiền xử lý dữ liệu ảnh và kiểm tra dữ liệu đọc tự động. Tuy nhiên, cũng giống Pluviograph, Gscanth chỉ làm việc với các loại giản đồ có 1 đường line, cấu hình lưới thẳng, và các kết quả đánh giá với số lượng ảnh hạn chế.

Tại Việt Nam, việc ứng dụng các kỹ thuật xử lý ảnh và trí tuệ nhân tạo trong ngành KTTV đã được quan tâm trong thời gian gần đây. Trong nghiên cứu [7, 8] các tác giả trích xuất thông tin về đường bờ biển và dự đoán biến động đường bờ biển tự động từ ảnh vệ tinh SENTINEL-1A sử dụng các kỹ thuật xử lý ảnh truyền thống. Ảnh radar SENTINEL-1A cũng được sử dụng trong nghiên cứu số [9] để xây dựng đặc tính hồ chứa. Ngoài ra, việc sử dụng ảnh radar, ảnh vệ tinh Himawari-8 trong dự báo thời tiết hoặc cảnh báo mưa lớn được trình bày trong nghiên cứu [10-12]. Trong lĩnh vực tài nguyên môi trường, các tác giả trong nghiên cứu [13] sử dụng dịch vụ GoogleEarth để đánh giá biến đổi thảm thực vật theo cả không gian và thời gian. Việc trích xuất thông tin từ hình ảnh kết hợp với các mô hình học máy cho thấy tiềm năng rất lớn trong việc khai thác dữ liệu khí tượng thủy văn. Tuy nhiên, việc số hóa các giản đồ tự ghi KTTV hiện nay mới chỉ dừng lại ở mức sử dụng phần mềm. Việc làm chủ công nghệ lõi, từ quá trình thu nhận ảnh đến quá trình xử lý, nhận dạng và đọc dữ liệu từ giản đồ sẽ cho phép tùy biến và tối ưu theo đặc thù các loại giản đồ KTTV của Việt Nam. Thông qua nghiên cứu này, chúng tôi sẽ tiếp cận theo hướng ứng dụng các kỹ thuật tiên tiến của xử lý ảnh và học máy để tối ưu cho không chỉ một loại giản đồ mà thích nghi với nhiều loại giản đồ và phù hợp với đặc thù dữ liệu giản đồ của Việt Nam.

Trong lĩnh vực trí tuệ nhân tạo tại Việt Nam, mặc dù kỹ thuật nhận dạng, số hóa trong văn bản đã được phát triển từ sớm [14-16], đến nay vẫn chưa có công cụ hỗ trợ số hóa giản đồ khí tượng thủy văn. Đối với số hóa tài liệu, có thể kể đến sản phẩm phần mềm VnDOCR 4.0 Professional là chương trình nhận dạng chữ Việt in, được phát triển bởi nhóm chuyên gia phát triển phần mềm tại Viện Công nghệ thông tin-Viện Khoa học và Công nghệ Việt Nam. Về kỹ thuật, các phần mềm số hóa tài liệu tiếng Việt chủ yếu phát triển dựa trên nền tảng mã nguồn mở Tesseract, với công nghệ Java/.NET, hỗ trợ nhận dạng cho các dạng ảnh PDF, TIFF, JPEG, GIF, PNG, và BMP [17]. Khả năng nhận dạng (ví dụ VietOCR) có thể đạt tới 95% đối với file ảnh có chất lượng tốt. Tại Trung tâm thông tin KTTV Quốc gia, có sử dụng dụng phần mềm chuyên ngành HYDSTRA (Úc) xử lý, lưu trữ số liệu thủy văn và một thiết bị kèm theo CalComp Digitizer để số hóa giản đồ mực nước. Việc phát triển một công cụ số hóa giản đồ tự ghi KTTV trong nước là hết sức cần thiết vì vừa đảm bảo làm chủ công nghệ

vừa có khả năng thích nghi, tùy biến với môi trường và điều kiện đặc thù của các loại giản đồ KTTV ghi tại Việt Nam.

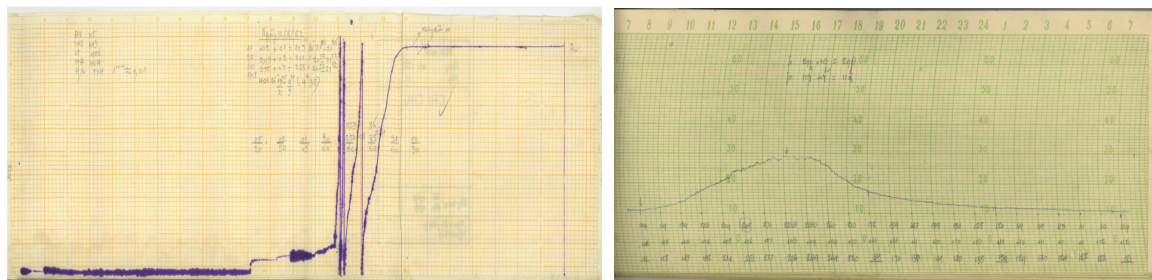
Ngày nay, với sự phát triển mạnh mẽ của cách mạng công nghiệp 4.0, trong đó các kỹ thuật nhận dạng và xử lý ảnh để phát triển các ứng dụng số hóa tài liệu/ văn bản đã có những bước tiến vượt bậc. Việc xây dựng công cụ DigiMeto cũng không nằm ngoài sự phát triển đó. Hướng tiếp cận chính khi xây dựng công cụ là sử dụng các kết quả của quá trình phân tích ảnh và nhận dạng tự động trong việc tách các đường line ghi trong giản đồ. Kết quả của quá trình này được hiển thị trên giao diện đồ họa người dùng cuối. Người dùng cuối sẽ thực hiện một số thao tác hiệu chỉnh hoặc tùy biến các tham số khi đọc số liệu từ một trong số các loại giản đồ mà công cụ hỗ trợ. Với phiên bản hiện thời, công cụ hỗ trợ đọc các loại giản đồ tự ghi thông dụng như giản đồ mưa, nhiệt, ẩm, áp và lượng nước. DigiMeto hoạt động với khả năng tính toán thời gian thực, giao diện đồ họa thân thiện với người dùng cuối. Công cụ có thể xử lý quy trình đọc, số hóa thông tin hoàn chỉnh đầu-cuối với dữ liệu đầu vào là ảnh scan của một loại giản đồ và đầu ra là các file csv lưu kết quả đọc theo khoảng thời gian đã được người dùng định nghĩa trước. Công cụ hướng tới mục tiêu rút ngắn đáng kể thời gian đọc thủ công hiện nay. Việc xây dựng công cụ đã mở ra hướng giải quyết khả thi cho các loại tài liệu KTTV đang lưu trữ tại Trung tâm KTTV quốc gia như các loại sổ ghi.

2. Phương pháp nghiên cứu

2.1. Dữ liệu nghiên cứu

Dữ liệu được sử dụng trong nghiên cứu bao gồm các loại giản đồ KTTV tự ghi được lưu trữ tại Trung tâm thông tin và Dữ liệu KTTV như: giản đồ mưa, ẩm, áp và nhiệt. Các loại giản đồ này đều có đặc điểm chung là dữ liệu thu thập của các trạm khí tượng đo cùng với thời gian trạm được thành lập từ những năm 1950 cho đến nay. Theo thời gian, số liệu giản đồ bị gián đoạn do thiết bị hỏng hoặc máy bị trục trặc. Tùy thuộc vào giai đoạn, các loại giản đồ có các hình thức và thiết kế không hoàn toàn giống nhau. Ví dụ trên các giản đồ mưa, các dòng kẻ màu xanh trong giai đoạn năm 1960 đến năm 1970. Trong các giai đoạn sau, giản đồ mưa có màu cam, độ phân giải các ô lưới cũng ở mức chi tiết hơn. Một số loại giản đồ có kích thước khác nhau qua các giai đoạn, hoặc có tỉ lệ thang chia có khác nhau. Ví dụ, giản đồ ẩm có loại thang chia là 5% (do Việt Nam sản xuất) hoặc 2% (do Liên xô cũ sản xuất). Màu sắc các giản đồ cũng có sự khác biệt, có loại có các màu cơ bản như giản đồ màu xanh đậm và xanh nhạt; có loại đen đậm và đen nhạt; có loại màu nâu đậm.

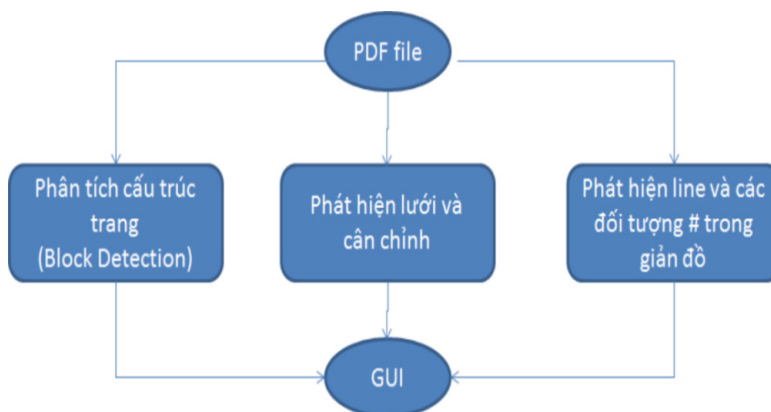
Trải qua thời gian, các giản đồ từ trước năm 70 có tình trạng vật lý kém, nhiều tờ giản đồ bị thủng, bị rách và còn có những tờ giản đồ bị oxy hóa gây ra mất thông tin. Trong khi đó, giai đoạn sau này, giản đồ mỏng, là nguyên nhân các giản đồ dễ bị nhòe khi gặp độ ẩm không khí cao và cũng rất dễ bị nhăn, cong trong quá trình vận chuyển. Chất lượng vật lý nhiều giản đồ ẩm có những vết đen trên giấy do vết oxy hóa. Rất nhiều giản đồ bị hiện tượng thay giấy in gặp lúc trời đang mưa, do đó rất có thể bị ướt sinh ra nét mực bị nhòe nét. Nét in trên giản đồ cũng có thể bị nhòe do bút mực của thiết bị đổ mực không đều, hoặc trong quá trình bảo quản cũng có thể sơ ý làm giản đồ bị ẩm sinh nhòe nét. Hình 1 minh họa một số hiện tượng vật lý làm suy giảm chất lượng của giản đồ. Những vấn đề trên gây khó khăn đối với các thuật toán nhận dạng và xử lý ảnh làm suy giảm tính xác của quá trình đọc tự động. Do đó, hướng tiếp cận của công cụ DigiMeto là thực hiện tự động một phần. Kết quả của thuật toán tự động được quan sát và hiệu chỉnh dựa trên giao diện GUI. Việc xây dựng công cụ DigiMeto được mô tả chi tiết như sau.



Hình 1. Chất lượng các giản đồ bị suy giảm do thời gian. Trái: nét mực in bị nhòe; Phải: đường line ghi dữ liệu bị mờ.

2.2. Phương pháp nghiên cứu

Nghiên cứu này tập trung phát triển ứng dụng đồ họa người dùng cuối (*Graphic User Interface*) trong một sơ đồ tổng thể quá trình số hóa giản đồ tự ghi như minh họa trong Hình 2. Công cụ DigiMeto nhằm mục đích trực quan hóa kết quả tự động phân tích và nhận dạng đồ thị qua các kỹ thuật xử lý ảnh. Công cụ DigiMeto đồng thời hỗ trợ các thao tác tinh chỉnh của người đọc đồ thị, để đảm bảo kết quả đọc là chính xác và theo yêu cầu thực tế về việc số hóa các loại giản đồ. Như phần 2.1 đã trình bày, chất lượng các loại giản đồ theo thời gian có nhiều loại đã xuống cấp; việc áp dụng các công cụ tự động thường không đạt được kết quả như mong muốn. Do đó, hướng tiếp cận chính trong nghiên cứu là thông qua các thao tác tinh chỉnh của người dùng cuối, hệ thống sẽ có khả năng điều chỉnh và khắc phục những hạn chế/tồn tại kết quả sinh từ kỹ thuật tự động. Chúng tôi nhận thấy đây là cách tiếp cận phù hợp với điều kiện thực tế khi triển khai ứng dụng. Công cụ được xây dựng trên nền tảng máy tính chạy hệ điều hành Window 10; Viết bằng ngôn ngữ lập trình C++, sử dụng một số bộ thư viện hỗ trợ đồ họa là Qt 5.15, bộ thư viện hỗ trợ xử lý ảnh OpenCV 4.5.1. Công cụ được dịch và đóng gói cho phép cài đặt như một ứng dụng độc lập trên các máy tính Windows PC bất kỳ với cấu hình tối thiểu là Intel Core i5; 8GB RAM; 500 GB HDD, hệ điều hành Windows 10.



Hình 2. Các khối module chính của ứng dụng DigiMeto.

Trong nghiên cứu này tập trung phát triển giao diện giao diện người dùng cuối (GUI). Về mặt chức năng, công cụ xây dựng gồm có 3 mức chức năng chính như trình bày trong Hình 3 về Biểu đồ phân cấp chức năng. Chức năng *Hiển thị* gồm các thao tác liên quan đến việc chọn thư mục, chọn giản đồ hiển thị trong màn hình chính; tinh chỉnh các tham số hình học về việc hiển thị biểu đồ (như thu phóng ảnh, xoay, cắt ảnh để phù hợp với loại giản đồ thị). Các chức năng này cũng được thể hiện thông qua thanh toolbar như trên giao diện GUI chính của công cụ. Chức năng tinh chỉnh tham số cho phép người dùng cuối điều chỉnh khoảng cách giữa các mốc thời gian, thời điểm bắt đầu ghi dữ liệu. Tinh chỉnh tham số của đường lưới đồ thị trong trường hợp đường lưới có dạng cong (hình parabol). Chức năng tạo báo cáo để ghi lại thông tin báo cáo theo mẫu; xuất kết quả ra file csv; nhập các thông tin

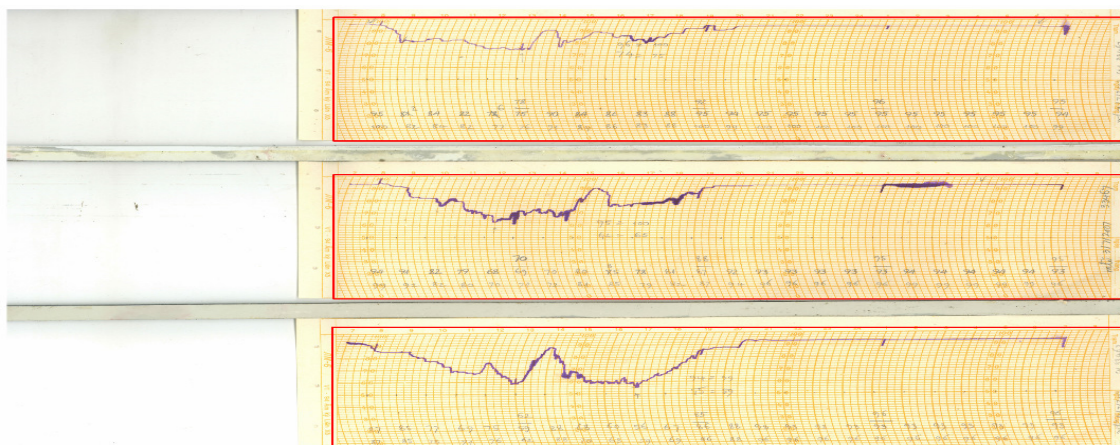
khác liên quan đến đồ thị đang đọc như tên trạm, ngày tháng đo. Các module của chương trình cũng được thiết kế thực hiện theo các chức năng này. Trong phần tiếp theo, một số module chính sẽ được trình bày trên hình 3.



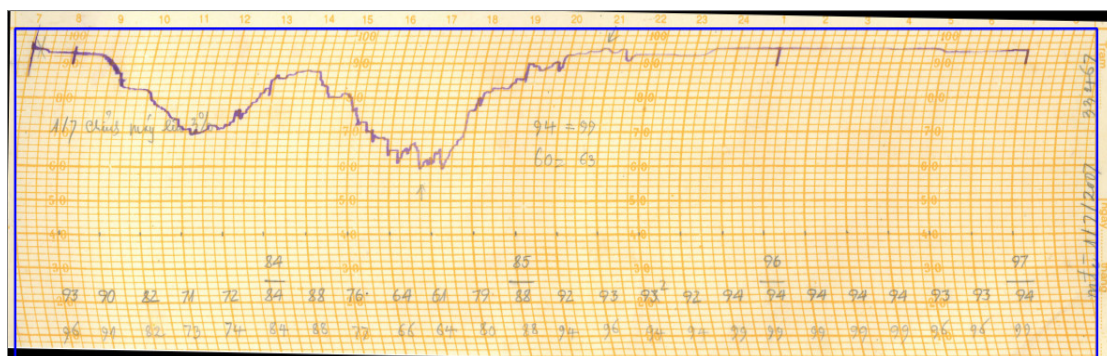
Hình 3. Biểu đồ phân cấp chức năng của ứng dụng.

2.3. Các kỹ thuật xử lý ảnh áp dụng trên giản đồ

Do đặc điểm các giản đồ có thể được sắp đặt và quét (scan) vào chung 1 file ảnh. Do đó cần có quá trình tách từng giản đồ theo mỗi ngày ra khỏi file ảnh ban đầu. Quá trình xử lý này sẽ áp dụng các kỹ thuật xử lý ảnh truyền thống để tách mỗi giản đồ ra thành các đối tượng riêng biệt từ hình ảnh scan có thể chứa nhiều biểu đồ trong một file ảnh. Như minh họa trong Hình 4, 1 file scan ảnh có thể chứa 03 giản đồ tương ứng với 03 ngày ghi dữ liệu. Kết quả sau bước tách giản đồ là các khối giản đồ là vùng khoanh màu đỏ đã được xác định. Trong bước này, mật độ các điểm đặc trưng sử dụng thuật toán Harris Corner [18] đã được áp dụng để xác định vùng đồ thị quan tâm trong ảnh. Sau khi tách xong, các kỹ thuật nắn và xoay đồ thị đã được áp dụng để đưa ảnh hiển thị trong một hệ tọa độ chuẩn thông qua tìm các tham số biến đổi Affine [19] (Hình 5). Tách đường line để thực hiện quá trình số hóa đồ thị (Hình 6). Các quá trình trên được thực hiện tự động hoàn toàn dựa trên các thuật toán cơ bản của xử lý ảnh như kỹ thuật Canny để tách biên [20]; Các bước này được thực hiện bằng ngôn ngữ Python và sử dụng thư viện hỗ trợ xử lý ảnh OpenCV [4].



Hình 4. Kết quả xác định các đối tượng là biểu đồ (khoanh vùng màu đỏ) từ hình ảnh gốc.



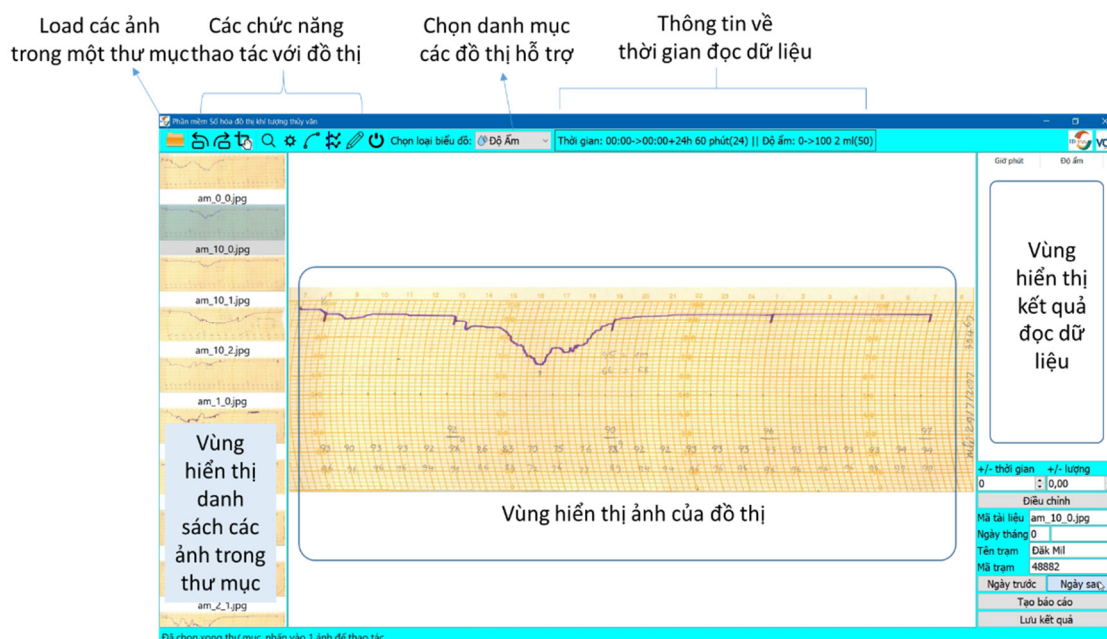
Hình 5. Kết quả cân chỉnh (xoay) biểu đồ về một hệ tọa độ trục chuẩn (viên đen ở góc trên bên phải và góc dưới bên trái minh họa kết quả của phép xoay để đồ thị được hiển thị vuông góc).



Hình 6. Kết quả tách đường line trong 1 biểu đồ âm.

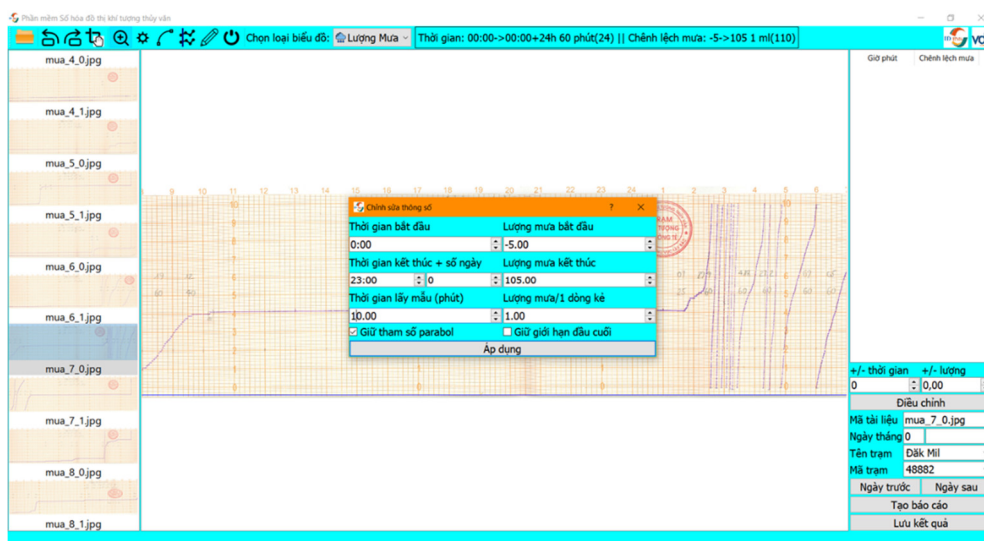
2.4. Xây dựng giao diện GUI

Sau khi đã thực hiện quá trình tiền xử lý, kết quả tách đường line và ảnh gốc được lưu vào trong 1 thư mục. Người dùng thực hiện chức năng load ảnh từ một thư mục. Giao diện GUI của Digimeto được minh họa trong Hình 7. Như minh họa, thanh toolbar thể hiện các chức năng chính của ứng dụng DigiMeto. Giao diện được chia làm 3 vùng chính: (1) Vùng hiển thị ảnh trong thư mục; vùng hiển thị ảnh giảm độ đang làm việc, với các giá trị đo tương ứng; Vùng hiển thị kết quả đọc trên toàn bộ giảm độ. Một số chức năng hỗ trợ như duyệt giảm độ theo ngày, thông tin về trạm, tạo lập báo cáo lưu kết quả cũng được hỗ trợ.



Hình 7. Giao diện chính của ứng dụng DigiMeto.

Đối với một số loại giản đồ, kết quả phát hiện điểm đầu (mốc thời gian) bắt đầu đọc dữ liệu còn chưa chính xác; hoặc xác định khoảng thời gian lấy mẫu (thay vì mặc định là 15 phút hoặc 1h, có thể đặt giá trị tùy chỉnh); hoặc điểm gốc tọa độ bắt đầu đọc dữ liệu; kỹ thuật viên có thể tinh chỉnh những tham số này qua giao diện như Hình 8.



Hình 8. Thao tác lựa chọn khoảng thời gian đọc (thời gian lấy mẫu tín hiệu) trên công cụ.

3. Kết quả và thảo luận

3.1. Tập dữ liệu đánh giá và độ đo

Trong các đánh giá bước đầu này, chúng tôi chỉ lựa chọn một tập nhỏ các ảnh từ các hình ảnh giản đồ đã được scan tại trung tâm KTTV. Mỗi loại dữ liệu lựa chọn 05 giản đồ; với các đặc điểm về độ khó và chất lượng giản đồ khác nhau. Ví dụ, Hình 9 minh họa độ phức tạp khi đọc giản đồ mưa. Trong đó, Hình 9a là một giản đồ mưa đơn giản thường gặp; giản đồ Hình 9b là một giản đồ mưa phức tạp trong đó có nhiều đường line sát nhau có thể làm ảnh hưởng tới kết quả đọc bằng phần mềm DigiMeto.



(a) – Một giản đồ mưa đơn giản thường gặp



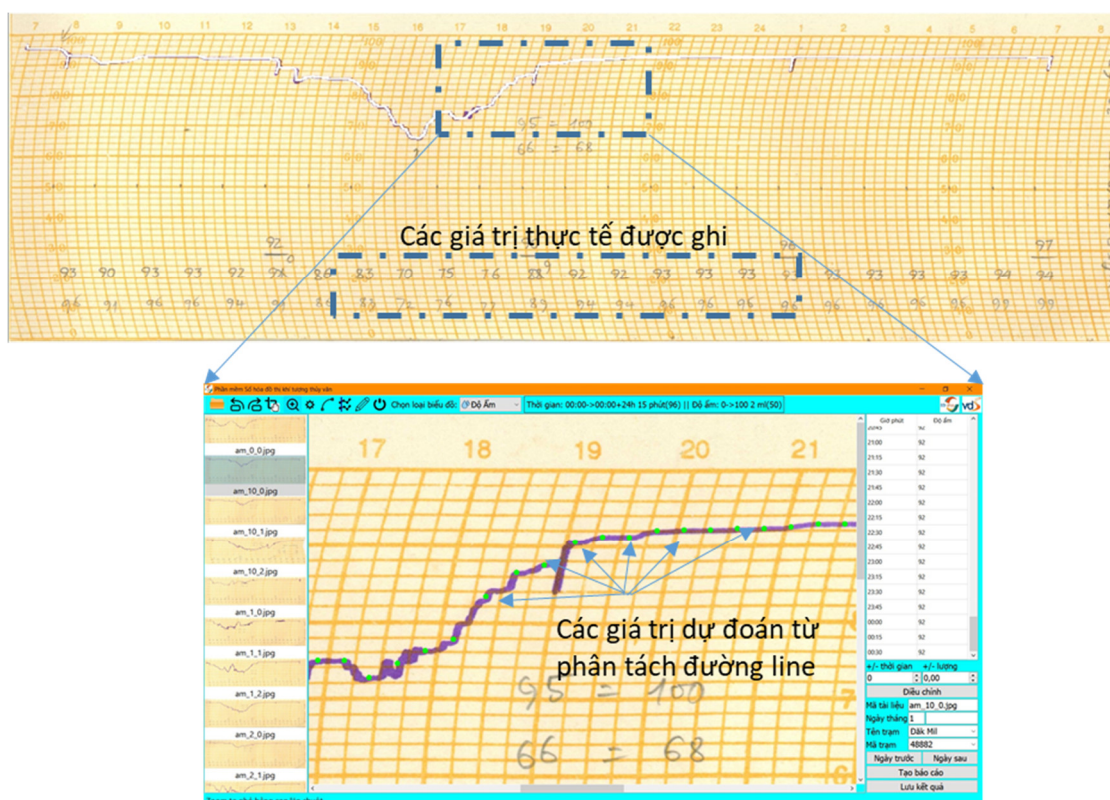
(b) – Một giản đồ mưa phức tạp với lượng mưa lớn ở thời gian sau 2h sáng

Hình 9. Minh họa độ khó của một loại giản đồ mưa được lựa chọn trong tập dữ liệu đánh giá.

Độ đo được sử dụng là sai lệch trung bình (*MAE*–*Mean Absoluted Error*) giữa giá trị đo bằng phần mềm DigiMeto và giá trị thực tế do kỹ thuật viên ghi trên bản đồ. Vì mỗi loại giản đồ có đơn vị đo khác nhau, chúng tôi tính toán giá trị tương đối (%) thay vì giá trị tuyệt đối như trong công thức *MAE* thông thường. Công thức tính *MAE* theo % được định nghĩa như sau:

$$MAE = \frac{1}{l} \sum_{i=1}^l \frac{|f(i) - \hat{f}(i)|}{f(i)} \quad (1)$$

Trong đó, *l* là tổng số mẫu (điểm) lấy dữ liệu; *f*(*i*) là giá trị của kỹ thuật viên, $\hat{f}(i)$ là giá trị đo bằng phần mềm. Hình 10 minh họa cách xác định giá trị *f*(*i*) và, $\hat{f}(i)$ đối với một ảnh giản đồ trong tập dữ liệu đánh giá. Ngoài ra, sai lệch về thời gian tại điểm đo cũng được đánh giá tương tự nhưng sử dụng giá trị tuyệt đối (theo đơn vị là phút).



Hình 10. Kết quả đọc tại một thời gian từ 17h đến 21h của đồ thị trong Hình 1.

3.2. Kết quả đánh giá

Kết quả đánh giá một số dữ liệu giản đồ ẩm, mưa, áp và nhiệt được thống kê trong Bảng 1 đến Bảng 4 tương ứng. Kết quả giá trị thống kê cho thấy, về mặt giá trị đo, sai lệch với các giản đồ nhiệt, ẩm và áp là không quá 2%. Trong khi đó, giản đồ mưa có sai lệch lớn hơn. Ngoài giá trị trung bình, các giá trị thống kê khác như trung vị, sai số lớn nhất cũng cho thấy độ ổn định của kết quả. Đặc biệt trường hợp Hình 9b khi giản đồ mưa phức tạp, lượng mưa lớn nên đoạn từ 7h có giá trị sai số lớn. Một số hướng giải quyết trong trường hợp này bao gồm các công cụ tinh chỉnh lựa chọn đường và giá trị gốc của mỗi đường có thể trợ giúp hiệu chỉnh các giá trị đo. Về sai lệch thời gian đo, trong các bảng thống kê cho thấy sai lệch không quá 5 phút. Tuy nhiên, một số giản đồ yêu cầu thời gian tinh chỉnh như xác định thời điểm bắt đầu, thời điểm kết thúc; đánh dấu mốc trên lưới. Những thao tác này yêu cầu thời gian thực hiện của kỹ thuật viên. Trung bình thời gian thực hiện cho mỗi giản đồ là 3 phút. Ngoài giá trị trung bình, các giá trị thống kê khác như trung vị, sai số lớn nhất cũng cho thấy độ ổn định của kết quả.

Bảng 1. Kết quả đánh giá sai lệch với các gián đồ Âm.

Ảnh	Đánh giá sai lệch	Giá trị thống kê		
		Trung bình ± STD	Trung vị	Sai số lớn nhất
am_2_1.png	Giá trị đọc (%)	0,79 ± 1	1,08	2,33
	Thời gian (phút)	0:05	0:04	0:15
am_1_1.png	Giá trị đọc (%)	0,77 ± 1	1,06	2,27
	Thời gian (phút)	0:07	0:02	0:20
am_10_1.png	Giá trị đọc (%)	0,55 ± 1	0,00	3,16
	Thời gian (phút)	0:03	0:02	0:14

Bảng 2. Kết quả đánh giá sai lệch với các gián đồ Mưa.

Ảnh	Đánh giá sai lệch	Giá trị thống kê		
		Trung bình ± STD	Trung vị	Sai số lớn nhất
mua_5_0.png	Giá trị đọc (%)	0.04 ± 0	0	1.00
	Thời gian (phút)	0:00:31	0:00	0:01
mua_6_0.png	Giá trị đọc (%)	0.04 ± 0	0	1.00
	Thời gian (phút)	0:00:29	0:00	0:01
mua_7_0.png	Giá trị đọc (%)	12.40 ± 46	0	59
	Thời gian (phút)	0:00:31	0:00	1:00

Bảng 3. Kết quả đánh giá sai lệch với các gián đồ Nhiệt.

Ảnh	Đánh giá sai lệch	Giá trị thống kê		
		Trung bình ± STD	Trung vị	Sai số lớn nhất
nhiet_2_1	Giá trị đọc (%)	1.75 ± 5	0.55	24,81%
	Thời gian (phút)	0:00:20	0:00	0:01
nhiet_10_1	Giá trị đọc (%)	1.19 ± 1	1.13	2,45%
	Thời gian (phút)	0:00:37	0:01	0:01
nhiet_0_0	Giá trị đọc (%)	2.45 ± 1	2.49	6,07%
	Thời gian (phút)	0:00:29	0:00	0:01

Bảng 4. Kết quả đánh giá sai lệch với các gián đồ Nước.

Ảnh	Đánh giá sai lệch	Giá trị thống kê		
		Trung bình ± STD	Trung vị	Sai số lớn nhất
nuoc_0	Giá trị đọc (%)	0.09 ± 0	0,10	0,10
	Thời gian (phút)	0:00:26	0:00	0:01
nuoc_1	Giá trị đọc (%)	0.13 ± 0	0,13	0,14
	Thời gian (phút)	0:00:29	0:00	0:01
nuoc_3	Giá trị đọc (%)	0.15 ± 0	0,15	0,15
	Thời gian (phút)	0:00:31	0:00	0:01

4. Kết luận

Trong nghiên cứu đã đề xuất các kỹ thuật để phát triển ứng dụng DigiMeto thực hiện số hóa một số loại giản đồ tự ghi ngành KTTV. Phần mềm sử dụng các kết quả từ các thuật toán xử lý ảnh và nhận dạng. Hiện nay phần mềm DigiMeto hỗ trợ 4 loại giản đồ chính là Mưa, Âm, Nhiệt, Áp. Qua đánh giá bước đầu với một số giản đồ cho thấy sai số trong quá trình đọc bằng phần mềm là dưới 2% với các giản đồ có chất lượng tốt. Một số tinh chỉnh do người dùng cuối thực hiện trong phần mềm đã giúp cho việc đọc dữ liệu trên các giản đồ được chính xác và tùy biến. Kết quả bước đầu này minh chứng tính khả thi khi sử dụng các công cụ của xử lý ảnh và học máy trong giải quyết bài toán số hóa tài liệu ngành KTTV. Trong thời gian tới, chúng tôi sẽ tiếp tục tối ưu quá trình xử lý để đảm bảo giảm thiểu thời gian và sự tương tác của người dùng cuối. Đặc biệt, chúng tôi sẽ tập trung giải quyết những hạn chế của ứng dụng đối với các giản đồ có chất lượng kém do đã lưu trữ qua thời gian dài.

Đóng góp của tác giả: Xây dựng ý tưởng nghiên cứu: P.L.P., H.H.P., H.V.; Lựa chọn phương pháp nghiên cứu: P.L.P., H.H.P., H.V.; Xử lý số liệu: L.T.H., M.X.H., T.V.L.; Phân tích mẫu: H.H.P., H.V., M.X.H., T.V.L.; Lấy mẫu: L.T.H.; Viết bản thảo bài báo: P.L.P., H.H.P., H.V.; Chỉnh sửa bài báo: P.L.P., H.H.P., H.V.

Lời cam đoan: Tập thể tác giả cam đoan bài báo này là công trình nghiên cứu của tập thể tác giả, chưa được công bố ở đâu, không được sao chép từ những nghiên cứu trước đây; không có sự tranh chấp lợi ích trong nhóm tác giả.

Tài liệu tham khảo

1. Susin, N.; Peer, P. Open–source tool for interactive digitisation of pluviograph strip charts. *Weather* **2018**, 73, 222–226. <https://doi.org/10.1002/wea.3001>
2. Burboa, A. PluvioReader: a software for digitizing weekly siphoning–type 1 pluviograph strip charts, *Computers and Geosciences*, 2020. <https://doi.org/10.1016/j.cageo.2020.104463>.
3. Pons, F.; Laroche, C.; Trmal, C.; Puechberty, R.; Baillon, M. “Hydrometry data rescue, a stake for the future”, in the Proceeding of 3rd European Conference on Flood Risk Management, 2016. <https://doi.org/10.1051/e3sconf/20160704021>.
4. Deidda, R.; Mascaro, G.; Piga, E.; Querzoli, G. An automatic system for rainfall signal recognition from tipping bucket gage strip charts. *J. Hydrol.* **2007**, 333, 400–412.
5. Jaklic, A.; Sajn, L.; Derganc, G.; Peer, P. Automatic digitization of pluviograph strip charts. *Meteorol. Appl.* **2016**, 23, 57–64. <https://doi.org/10.1002/met.1522>.
6. Radon, J. Digitizing of Analogue Data–Charts from Thermohygrographs. *J. Applied Comput. Sci.* **2017**, 25, 49–60.
7. Nhi, H.Y.; Thoa, L.T.K. Trích xuất đường bờ biển từ ảnh SENTINEL–1A khu vực thành phố Phan Thiết. *Tap chí Khí Tượng Thủy Văn* **2019**, 697, 20–25.
8. Tình, T.V.; Phong, D.H. Sử dụng ảnh viễn thám và Gis nghiên cứu biến động đường bờ biển mũi Cà Mau. *Tap chí Khí tượng Thủy văn* **2017**, 684, 35–40.
9. Hiệp, N.Q.; Hùng, N.A. Cách tiếp cận mới xây dựng đường đặc tính hồ chứa bằng việc sử dụng ảnh viễn thám radar SENTINEL–1. *Tap chí Khí tượng Thủy văn* **2019**, 706, 10–19.
10. Trung, L.B.; Toán, H.M.; Phong, N.B. Nghiên cứu sử dụng số liệu định vị sét kết hợp với ảnh radar để cảnh báo đợt mưa lớn từ 01–06/8/2017 trên khu vực Tây Bắc. *Tap chí Khí tượng Thủy văn* **2018**, 685, 48–53.
11. Ngà, P.T.T. Nghiên cứu tương quan giữa đặc tính mây và mưa lớn cho khu vực Hồ Chí Minh bằng dữ liệu vệ tinh Himawari–8 và GSMaP. *Tap chí Khí tượng Thủy văn* **2019**, 702, 21–30.
12. Công, T.T.; Quyền, L.N.; Giám, N.M.; Quyết, L.D. Ứng dụng số liệu ảnh mây vệ tinh Himawari trong dự báo và cảnh báo mưa dông cho khu vực Đồng bằng sông Cửu Long. *Tap chí Khí tượng Thủy văn* **2020**, 713, 1–13.

13. Phương, T.A.; Cường, T.M.; Chiền, P.V.; Phong, L.V.V. Nghiên cứu ứng dụng công nghệ Google Earth Engine đánh giá sự thay đổi thảm thực vật theo không gian và thời gian ở Bến Tre trong mùa khô giai đoạn 2016–2020. *Tap chí Khí tượng Thủy văn* **2020**, 713, 47–55.
14. Hoàng, L.M.; Tạo, N.A.; Mai, L.C.; Ứng dụng mô hình Markov ẩn trong nhận dạng chữ. *Tap chí khoa học và công nghệ* **2002**, 40, 31–40.
15. Phuong, P.A.; Tạo, N.Q.; Mai, L.C. Kết hợp các bộ phân lớp SVM cho việc nhận dạng chữ viết tay rời rạc. *Tap chí tin học và điều khiển* **2004**, 25, 1, 88–97.
16. Vu, T.T.; Luong, M.C.; Nakamura, S. A study of Phonetic Units for Vietnamese Speech Recognition, in the Proceeding of the IEEE International Conference on Computing and Communication Technologies RIVF (Research, Innovation and Vision for the Future), 2010, 1–4, Hanoi, Vietnam.
17. Hiếu, L.T.; Vũ, L.A.; Kiên, L.T. Áp dụng xác suất thống kê và quá trình máy tự học cho bài toán phân tách từ văn bản tiếng Việt. *Tap chí Khoa học và công nghệ Trường Đại học Duy Tân* **2013**, 1, 32–38.
18. Nixon, M.S.; Aguado, A.S. Feature Extraction and Image Processing. Newnes: Elsevier, Oxford, 2002.
19. Forsyth, D.A.; Ponce J. Computer Vision – A Modern Approach. Prentice Hall: Upper Saddle River, NJ, 2002.
20. Gonzalez, D.C.; Woods, R.E. “Digital Image Processing, Third Edition, Pearson Educational International Publisher, 2002.

DigiMeto: A tool supports Digitizing Metontological Graphs via Graphic User Interface

Pham Le Phuong^{1,*}, **Le Trung Hung**², **Manh Xuan Huy**³, **Trieu Van Linh**⁴, **Nguyen Xuan Hoai**⁴, **Hai-Hong Phan**⁵, **Vu Hai**⁶

¹ Hydro–Meteorological Information and Data Center; phuongpl80@gmail.com

² Hydro–Meteorological Information and Data Center; hungtttl@gmail.com

³ Hanoi University of Science and Technology; huy.mx161818@sis.hust.edu.vn

⁴ AI Academy Viet Nam; nxhoai@gmail.com; trieuvanlinh04101998@gmail.com

⁵ Military Technical Academy, Viet Nam; hongpth@lqdtu.edu.vn

⁶ School of Electronics and Telecommunications, Hanoi University of Science and Technology; hai.vu@hust.edu.vn

Abstract: In this paper, a software tool named DigiMeto (Digitizing Metontological Graph) is presented. This tool aims to support automatically reading hydro–meteorological graphs such as rainfall, pressure, humidity, temperature. To build this tool, recognition algorithms and image processing techniques are applied. First, the region–of–interest for each graph is cropped from the scanned image; Lines (containing data) and grid data of each type of the graph are separated from the background. Based on the pre–defined parameters of the grid configuration and a parameter of interval time, the corresponding values at points are determined. DigiMeto is built with a Graphic User Interface (GUI). A number of supporting functions built in DigiMeto allow users to update the outputs once the automatical results are incorrect. The experimental results confirmed that this tool can help reduce reading time comparing with conventional way with manual procedures. Accuracy rate of the humidity, pressure and thermohygrographs is averagely less than 2% error compared with the measurements by technical staff. This tool requires only 3 minutes for all operations to obtain the final results. With more complex graphs (e.g., the graphs are measured more than one day), there is usually a greater error. Future research directions are discussed and suggested so that this tool can be fully automated and minimal end–user’s operations.

Keywords: Digitizing Metontological Graphs; Pluviograph; Computer vision; Image Processing; Graphic User Interface.