HNUE JOURNAL OF SCIENCE DOI: 10.18173/2354-1059.2018-0001

Natural Sciences 2018, Volume 63, Issue 3, pp. 3-9

This paper is available online at <http://stdb.hnue.edu.vn>

**PHỤC HỒI ĐỐI TƯỢNG TỰ ĐỘNG TRONG VIDEO**

Đặng Thành Trung1 ,Nguyễn Thị Ngọc Anh2 và Vũ Đình Thuấn3

*1Khoa Công nghệ Thông tin, Trường Đại học Sư phạm Hà Nội*

*2Khoa Tự Nhiên, Trường Cao đẳng Sư phạm Thái Bình*

*3Phòng Nghiên cứu khoa học và Hợp tác Quốc tế, Trường Cao đẳng Sư phạm Thái Bình*

**TÓM TẮT**

Phục hồi ảnh số/ video số đều được xem là một trong những nghiên cứu quan trọng trong lĩnh vực thị giác máy tính, mục đích chính của bài toán này là nhằm phục hồi hoặc loại bỏ một số đối tượng trong ảnh cũng như video sao cho người xem không thể phát hiện được. Sự khác biệt giữa ảnh và video là việc phục hồi được thực hiện dựa trên không chỉ các thông tin trong miền không gian mà cả các thông tin trong miền thời gian (một chuỗi các ảnh thay vì một ảnh). Thông thường, các kỹ thuật phục hồi video thường có độ phức tạp cao hơn nhiều so với phục hồi trong ảnh vì một số lý do như: ảnh hưởng của nguồn sáng, sự di chuyển của camera, nền ảnh thay đổi, … Trong bài báo này, chúng tôi đề xuất một kỹ thuật phục hồi video được phát triển mở rộng dựa trên một kỹ thuật phục hồi ảnh. Thuật toán đề giới thiệu hai cải tiến chính: giảm không gian vùng tìm kiếm và tích hợp chiều thời gian cho độ đo tương tự nhằm đảm bảo chất lượng phục hồi cũng như thời gian tính toán là chấp nhận được. Một số kết quả thực nghiệm cho thấy thuật toán đề xuất là khá tốt, có nhiều tiềm năng phát triển trong tương lai.

***Từ khóa:*** phục hồi ảnh, phục hồi video, xóa đối tượng tự động, phục hồi vùng ảnh.

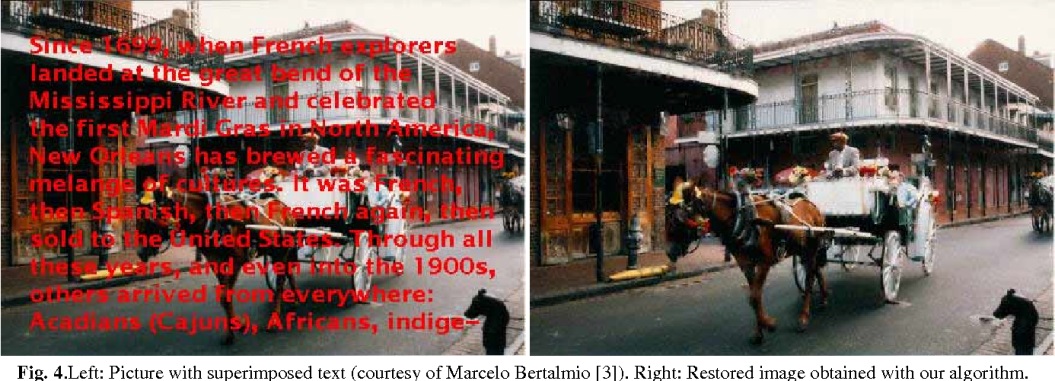
1. Mở đầu

Ngày nay, video trở thành phương tiện truyền thông quan trọng trong thế giới thực chính vì vậy các kỹ thuật liên quan đến xử lý video cũng được rất nhiều các nhà nghiên cứu quan tâm. Một trong những kỹ thuật được sử dụng nhiều nhất trong xử lý video là kỹ thuật inpainting, hay còn gọi là phục hồi video. Kỹ thuật này không chỉ đơn giản là việc phục hồi lại một vùng thông tin trong video sao cho người xem không thể phát hiện ra được, mà nó còn có thể ứng dụng để tạo ra nhiều hiệu ứng khác nhau: xóa đối tượng trong video (xóa phụ đề, logo, …); phục hồi video (xóa các vết xước, cái vết ố, ...); hỗ trợ mã hóa video, …

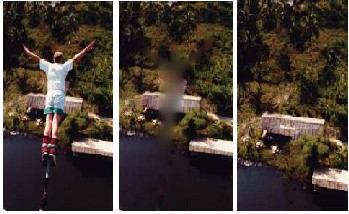
Về cơ bản, inpainting là kỹ thuật để loại bỏ các đối tượng không mong muốn hoặc khôi phục lại một phần của ảnh hoặc video dựa trên các thông tin trong phần còn lại của video sao cho người xem không thể phát hiện được khu vực phục hồi đó. Video inpainting là một ví dụ tiêu biểu với nhiều ứng dụng từ việc xóa các đối tượng không mong muốn, tạo hiệu ứng trực quan, khắc phục các vùng bị thiếu hay các vùng bị hư hỏng trong chuỗi video bằng cách sử dụng thông tin không gian và thời gian. Mục tiêu chính của kỹ thuật này là tạo ra một vùng khôi phục sao cho vùng chỉnh sửa và không chỉnh sửa được vá một cách liền mạch, không thể phát hiện với người xem. Quá trình này được thực hiện rất hiệu quả dẫn đến sự kết hợp trực quan được duy trì mọi nơi và mắt người không thể phát hiện ra sự biến dạng trong khu vực được chọn khi video được phát [1].

Ngày nay, có rất nhiều cách tiếp cận khác nhau được đề xuất. Hầu hết, các cách tiếp cận đều được chia thành 3 nhóm chính: phục hồi ảnh dựa trên phép nội suy; phục hồi ảnh dựa trên các bản vá và tổng hợp hai cách tiếp cận trên.

Các kỹ thuật phục hồi ảnh dựa trên các phép nội suy về bản chất sẽ tổng hợp thông tin của các vùng đã biết sau đó dựa trên các phép toán nội suy để tính toán các giá trị cần thiết cho các điểm ảnh trong vùng cần phục hồi. Bài báo đầu tiên thuộc nhóm này được đề xuất bởi Bertalmio và đồng nghiệp. Trong [2], họ đã giới thiệu một giải pháp cho việc xóa đối tượng trong ảnh dựa trên việc lan truyền thông tin từ vùng biên của đối tượng cần xóa. Cách tiếp cận này được xây dựng dựa trên việc mô phỏng quá trình phục hồi tranh thực tế bằng cách mô hình hóa bài toán bởi các phương trình sai phân từng phần (Partial Differential Equations - PDE); vùng phục hồi sẽ được thực hiện dựa trên việc giải các phương trình này. Kỹ thuật này sau đó được cải tiến bằng việc thay thế các phương trình sai phân bởi phương trình động lực học chất lỏng nhằm cải thiện độ mịn của vùng phục hồi. Một kết quả ví dụ cho đề xuất này được minh họa trong Hình 1. Bornemann and Marz [3] đã đề xuất một cải tiến cho cách tiếp cận của Bertalmio bằng việc thay thế các hàm trọng số và hướng phục hồi theo cạnh bằng hướng kết hợp. Với ý tưởng tương tự, Chan and Shen [4] đề xuất một mô hình phục hồi ảnh dựa trên một framework biến thiên tổng thể, trong mô hình này, một phương trình sai phân từng phần mới được giới thiệu dựa trên việc khuếch tán hướng cong thay cho phương trình sai phân cổ điển trước đây. Trên thực tế, từ các kết quả thu được, người ta nhận thấy các kỹ thuật này chỉ phù hợp khi các miền phục hồi là nhỏ, hẹp vì các kết quả phục hồi thường bị mờ do tính chất tự nhiên của các phép toán nội suy (Hình 2).



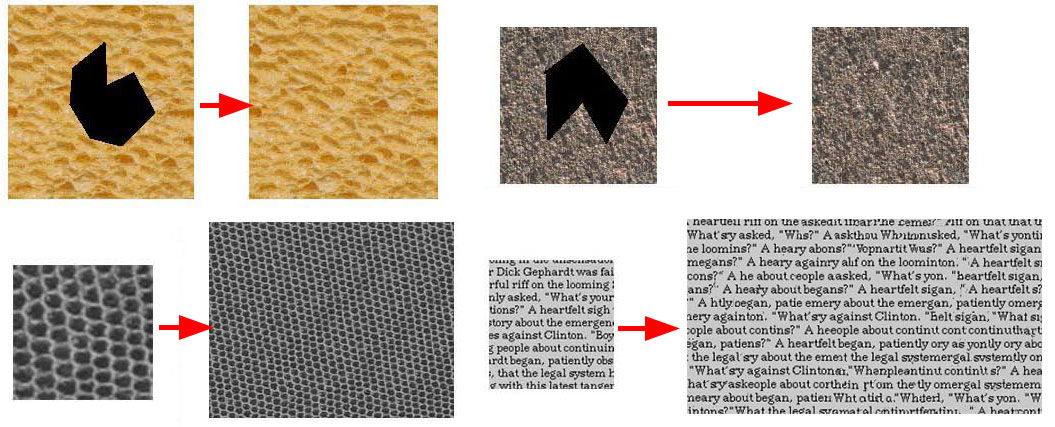
Hình 1. Phục hồi ảnh dựa trên đề xuất của Bertalmio



Hình 2. Thuật toán không hiệu quả khi vùng phục hồi lớn.

Các cách tiếp cận thuộc nhóm thứ hai là dựa trên việc xử lý các bản vá (cửa sổ). Cách tiếp cận này về cơ bản sẽ khắc phục được nhược điểm của các cách tiếp cận trước đây là do nó xử lý từng cửa sổ ảnh thay vì xử lý từng điểm ảnh như các cách tiếp cận trong nhóm 1. Kỹ thuật này bắt nguồn từ kỹ thuật tổng hợp vân ảnh được giới thiệu bởi Efros và đồng nghiệp [6] sau đó được cải tiến bởi Ashikhmin [5] với mục đích là giảm thời gian tính toán dựa trên độ đo liên kết của các cửa sổ ảnh. Một số ví dụ minh họa về việc tổng hợp vân ảnh được trình bày trong Hình 3. Trên thực tế, các ảnh tự nhiên thường là sự tổng hợp của cả vân ảnh và cấu trúc ảnh (là các cạnh hoặc đường nét trong ảnh). Do đó, kỹ thuật tổng hợp vân ảnh không thể áp dụng cho việc phục hồi các ảnh tự nhiên.

Để khắc phục vấn đề này, các tác giả trong [7] đã nỗ lực tách ảnh gốc ban đầu thành hai ảnh thành phần: ảnh cấu trúc và ảnh vân. Sau đó, với mỗi ảnh thành phần sẽ áp dụng độc lập một kỹ thuật phục hồi ảnh. Các ảnh kết quả sẽ được tổng hợp lại để tạo ra ảnh phục hồi cuối cùng. Cách tiếp cận này về cơ bản là hợp lý nhưng thực tế là khá phức tạp cũng như không hiệu quả trong việc phục hồi các vùng lớn.



Hình 3. Tổng hợp vân ảnh

Về cơ bản, việc tổng hợp vân ảnh có khả năng phục hồi các vùng ảnh lớn mà vùng phục hồi không bị mờ. Tuy nhiên, việc tổng hợp vân ảnh chỉ phù hợp với ảnh không có cấu trúc (ảnh chỉ có vân ảnh) nên không phù hợp với các ảnh tự nhiên. Dựa trên quan sát và nguyên lý làm việc của kỹ thuật tổng hợp vân ảnh, *Criminisi* và đồng nghiệp [8] đã giới thiệu một giải pháp khá hay cho việc phục hồi ảnh tự nhiên bằng cách đưa thêm độ ưu tiên trong quá trình phục hồi ảnh. Giải pháp này cho ra một loạt các kết quả khá ấn tượng và hiệu quả. Một loạt các cải tiến thêm được giới thiệu cho cách tiếp cận này như: cải tiến độ ưu tiên dựa trên các isophote chéo nhau, cải tiến độ ưu tiên dựa trên sự phân bố màu, cải tiến sự tổng hợp cửa sổ ảnh, …

Một cách tự nhiên, việc phục hồi video có thể được phát triển dựa trên việc phục hồi ảnh. Vì thực tế, video là một chuỗi các ảnh liên tiếp nhau. Dựa theo nguyên lý này, Wexler [9] và đồng nghiệp đã đề xuất việc phục hồi video dựa trên các thông tin trong cả chiều không gian (trong ảnh) và thời gian (trong các ảnh khác liền kề). Tuy nhiên, do không xem xét tới vấn đề không gian tìm kiếm nên đề xuất này thường khá tốn thời gian để thực hiện. Patwardhan [10] và đồng nghiệp đề xuất một giải pháp đơn giản hơn cho việc phục hồi video dựa trên một quan sát gọi là “hình nền ổn định”. Một lược đồ đơn giản kết hợp các thông tin về không gian – thời gian được đề xuất để sao chép các cửa sổ ảnh từ các khung hình không bị mất thông tin vào các khung hình cần phục hồi. Zhang và đồng nghiệp [11], đã đề xuất một giải pháp dựa trên các lớp di chuyển trong video bằng việc cắt các chuỗi khung hình trong video thành nhiều lớp, sau đó, các lớp được tổng hợp dựa trên đề xuất của Criminisi cùng các thông tin phụ khác. Tất cả các lớp, trừ các lớp không liên quan đến việc phục hồi đối tượng, sẽ được tổng hợp lại để trả về kết quả cuối cùng. Về cơ bản, một trong những vấn đề chính của các giải pháp đề xuất là độ phức tạp tính toán là khá lớn do chưa xem xét đến vùng tìm kiếm và tổng hợp bản vá. Vấn đề thứ hai là việc tổng hợp các bản vá chỉ dựa vào các thông tin hiện có trên khung hình hiện tại mà chưa quan tâm đến các khung hình khác. Nhằm khắc phục các vấn đề này, bài báo tập trung giới thiệu hai cải tiến quan trọng trong quá trình tổng hợp và phục hồi hình ảnh: hạn chế không gian tìm kiếm cho tổng hợp ảnh và tích hợp thông tin của các khung hình khác trong quá trình đối sánh. Phần còn lại của bài báo được tổ chức như sau, Phần 2 mô tả phương pháp đề xuất cho phục hồi video. Các kết quả thực nghiệm được minh họa trong phần 3. Bài báo kết thúc bởi phần kết luận và một số hướng nghiên cứu mở rộng trong tương lai.

2. Phương pháp đề xuất

Một cách tự nhiên, video có thể xem như là một chuỗi các khung hình (ảnh). Mỗi ảnh thường được tổng hợp từ hai thành phần chính: cấu trúc ảnh (structure) và vân ảnh (texture). Việc phục hồi video, về cơ bản có thể được phát triển dựa trên các nguyên lý cơ bản của phục hồi ảnh. Tuy nhiên, sự khác biệt lớn nhất giữa phục hồi video và phục hồi ảnh là việc các ảnh được hiển thị liên tục để tạo lên các hiệu ứng chuyển động. Đây cũng là một trong những vấn đề khiến cho việc phục hồi video trở lên khó khăn hơn so với phục hồi ảnh. Trong bài báo này, chúng tôi đề xuất một giải pháp mở rộng phục hồi ảnh dựa trên đề xuất trong [8]. Tiếp đó, các phần cải tiến sẽ được áp dụng nhằm giảm thiểu các vấn đề hiện có của các giải pháp phục hồi video hiện tại. Lược đồ chung của giải pháp đề xuất được minh họa trong Hình 4.

Video đầu vào

Cắt khung hình IN (N=0)

Xác định vùng phục hồi Ω dựa trên mặt nạ

Tìm biên vùng phục hồi dΩ

Xác định điểm phục hồi có độ ưu tiên cao nhất trên biên, p∈dΩ

Xác định cửa sổ cần phục hồi tại điểm p.

Tìm cửa sổ trong vùng ảnh Φ có độ tương tự lớn nhất với

Phục hồi cửa sổ bởi

dΩ = 0

Cập nhật lại các tham số hệ thống

Hết khung hình

Tiếp tục với khung hình khác, N=N+1

Tổng hợp các khung hình đã xử lý

Video đầu ra

No

Yes

No

Yes

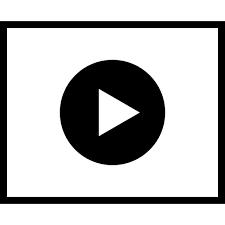
Hình 4. Lược đồ đề xuất phục hồi video

Trong lược đồ này, phần cải tiến được thực hiện tập trung vào bước “Tìm kiếm cửa sổ Wq có độ tương tự cao nhất với cửa sổ Wp”. Thực hiện chi tiết các bước trong lược đồ được mô tả chi tiết trong phần tiếp theo.

2.1 Xóa đối tượng trong video

Theo lược đồ trong Hình 4, video đầu vào sẽ được tách thành từng file ảnh. Tiếp đó, việc phục hồi sẽ được thực hiện trên từng ảnh. Các bước chính trong thuật toán bao gồm:

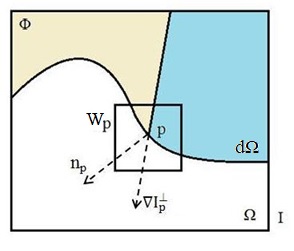
* **Tách ảnh từ video**: Video đầu vào sẽ được tách thành từng ảnh riêng biệt để xử lý. Tiếp đó các ảnh được xử lý một cách độc lập sau đó ghép lại để tạo thành video đầu ra.



Hình 5. Tách ảnh trong video

Bài báo này đề xuất một giải pháp xử lý nhiều ảnh liền kề cùng lúc để làm tăng chất lượng phục hồi ảnh. Trong thực nghiệm, một tập hợp gồm 3 ảnh liền kề được xử lý cùng một lúc. Chi tiết về việc cải tiến được trình bày trong phần sau.

* **Xác định vùng phục hồi**: Ảnh được chia thành 2 vùng chính: vùng ảnh cần phục hồi (Ω) và vùng ảnh còn lại (Φ). Dựa trên ảnh mặt nạ, biên dΩ của vùng phục hồi được xác định.
* **Tính toán độ ưu tiên**: Thuật toán đề xuất thực hiện nhiệm vụ tổng hợp phụ thuộc vào các giá trị ưu tiên được gán cho mỗi cửa sổ trên mặt trước lấp đầy. Việc xác định độ ưu tiên của các cửa sổ phụ thuộc vào: (i) Đang trên sự tiếp nối của các cạnh rõ nét; (ii) Được bao quanh bởi các điểm ảnh có độ tin cậy cao.



Hình 6. Sơ đồ kí hiệu

Độ ưu tiên P(p) của p được xác định như sau:

*P(p) = C(p)D(p)* (1)

Thuật ngữ độ tin cậy C(p) được xem như một phép đo lượng thông tin đáng tin cậy xung quanh điểm ảnh p. Trong quá trình điền, các điểm ảnh trong các lớp bên ngoài của vùng phục hồi thường có giá trị độ tin cậy cao hơn do đó được điền trước. Điểm ảnh ở giữa vùng phục hồi có các giá trị độ tin cậy thấp hơn. Độ tin cậy C(p) được định nghĩa như sau:

*C(p)*= (2)

Trong đó: | | là diện tích của , α là một yếu tố chuẩn hóa, I là toàn bộ hình ảnh.

Trong quá trình khởi tạo, hàm C(p) được xác định: *C(p)* = 0 ∀p ∈ Ω và *C(p)* = 1 ∀p ∈ I - Ω.

Thuật ngữ dữ liệu D(p) là một hàm của các isophotes trong dΩ tại mỗi lần lặp. Thuật ngữ này tăng độ ưu tiên của một cửa sổ mà một isophote “chảy” vào. Đây là một yếu tố quan trọng trong thuật toán vì nó khuyến khích các cấu trúc tuyến tính được tổng hợp đầu tiên và được truyền vào vùng phục hồi. D(p) được định nghĩa như sau:

(3)

Trong đó: np là một đơn vị vectơ trực giao với mặt trước dΩ tại điểm p, ∇ là isophote (hướng và cường độ) tại điểm p và ⊥ biểu thị toán tử trực giao.

* **Phục hồi ảnh:** Khi độ ưu tiên của các điểm ảnh đã được tính toán thì cửa sổ có độ ưu tiên cao nhất được tìm thấy. Sau đó nó được phục hồi với dữ liệu trích xuất từ vùng ảnh còn lại Φ. Thuật toán đề xuất truyền kết cấu hình ảnh bằng cách lấy mẫu trực tiếp từ vùng ảnh còn lại Φ. Cửa sổ tương tự như được tìm kiếm trong vùng Φ.

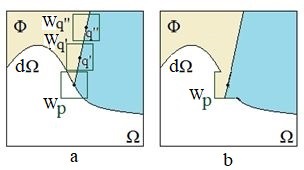
(4)

Trong đó: d(,*Wq*) là khoảng cách giữa hai cửa sổ và được xác định bằng tổng bình phương của các điểm ảnh đã được điền trong hai cửa sổ. Màu sắc điểm ảnh được thể hiện trong không gian màu CIE LAB [12].

Khi tìm được cửa sổ có độ tương tự lớn nhất với ta phục hồi cửa sổ bởi .

Sau khi cửa sổ đã được điền vào các giá trị điểm ảnh mới, độ tin cậy C(p) được cập nhật trong vùng được giới hạn bởi như sau:

(5)

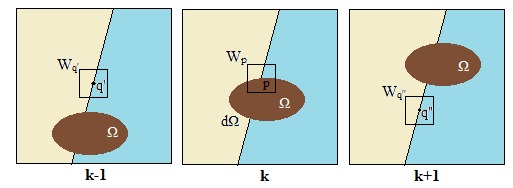


Hình 7. Tái tạo cấu trúc bằng cách tổng hợp cấu trúc dựa trên mẫu

Trong Hình 7: (a) Các ứng viên phù hợp nhất cho Wp nằm dọc theo ranh giới giữa hai kết cấu trong vùng nguồn, ví dụ, Wq' và Wq''. (b) Cửa sổ kết hợp tốt nhất trong tập ứng viên đã được sao chép vào vị trí Wp, do đó điền được một phần của Ω.

2.2 Cải tiến đề xuất

Điểm khác biệt giữa việc phục hồi ảnh và video chính là không gian tìm kiếm đầu vào. Với ảnh thì không gian tìm kiếm chỉ là ảnh đầu vào còn với video thì không gian tìm kiếm đầu vào thông thường người sẽ sử dụng 3 ảnh là ảnh chứa đối tượng cần xóa ban đầu, ảnh phía trước và phía sau nó trong video. Trong thuật toán đề xuất, việc phục hồi video dựa trên các thông tin trong cả chiều không gian (trong ảnh) và thời gian (trong các ảnh khác liền kề). Điều đó giúp hạn chế không gian tìm kiếm cho tổng hợp ảnh và tích hợp thông tin của các khung hình khác trong quá trình đối sánh mà vẫn đảm bảo chất lượng phục hồi cũng như thời gian tính toán là chấp nhận được.



Hình 8. Xóa đối tượng trong video

Trong Hình 8: Giả sử ảnh k là ảnh ban đầu chứa đối tượng cần xóa Ω, ảnh k - 1 (ảnh trước) và ảnh k + 1 (ảnh sau) là hai ảnh liền kề của k trong video. Biên dΩ của vùng cần xóa được xác định, p dΩ. Các ứng viên phù hợp nhất cho Wp sẽ được xác định trong vùng ảnh còn lại Φ của ảnh k và một số vùng lân cận của vùng cần xóa Ω của hai ảnh liền kề là k - 1 và k + 1 ví dụ: Wq' trong ảnh k - 1 hoặc Wq''trong ảnh k + 1.

Việc tìm kiếm cửa sổ cửa sổ có độ tương tự lớn nhất với trong cả hai ảnh liền kề và ảnh gốc giúp tăng chất lượng phục hồi video. Bên cạnh đó, chỉ cần tìm kiếm một số vùng lân cận của các ảnh liền kề giúp giảm không gian tìm kiếm dẫn đến giảm thời gian tính toán của thuật toán mà vẫn đảm bảo chất lượng phục hồi.

3. Một số kết quả

3.1 Kết quả thực nghiệm

Chương trình thực nghiệm của chúng tôi sử dụng ngôn ngữ lập trình C++ trên nền tảng Visual Studio 2017 kết hợp với thư viện mã nguồn mở OpenCV 2.4.13 và được thử nghiệm với nhiều loại dữ liệu video khác nhau được lấy trên mạng Internet: các video có kích thước, độ phân giải khác nhau.

***Bảng 1. Danh sách các video sử dụng thực nghiệm***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **STT** | **Tên video** | **Kích thước (pixels)** | **Định dạng** | **Số ảnh của video** | **Sô ảnh minh họa** |
| 1 | video01 | 600\*458 | MP4 | 8700 | 3 |
| 2 | video02 | 172\*172 | MP4 | 900 | 3 |
| 3 | video03 | 400\*267 | MP4 | 1885 | 3 |
| 4 | video04 | 314\*209 | MP4 | 725 | 3 |
| 5 | video05 | 410\*270 | MP4 | 1125 | 3 |

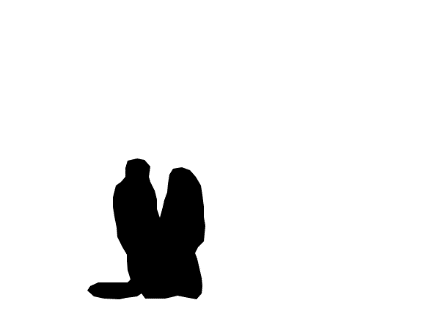
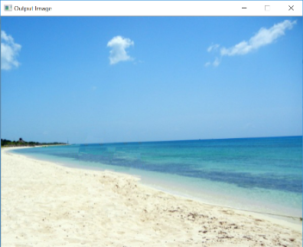
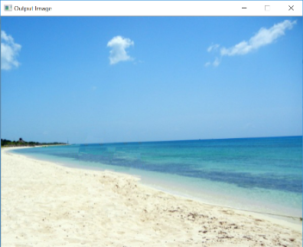
Từ các video trên chúng tôi đã tách thành các ảnh và lấy một số ảnh minh họa dưới đây.

***Bảng 2. Danh sách các hình ảnh sử dụng thực nghiệm***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **STT** | **Ảnh** | **Kích thước** | **Thời gian (s)** | |
| **Trước cải tiến** | **Sau cải tiến** |
| 1 | example\_1 | 600\*458 | 525 | 545 |
| 2 | example\_2 | 172\*172 | 45 | 4 |
| 3 | example\_3 | 400\*267 | 89 | 93 |
| 4 | example\_4 | 314\*209 | 129 | 49 |
| 5 | example\_5 | 410\*270 | 900 | 173 |

****Dưới đây là một số hình ảnh thực nghiệm. Mỗi bộ gồm: 3 ảnh gốc liên tiếp nhau (a, b, c), ảnh Mask (d), ảnh kết quả của các thuật toán chưa cải tiến (e), ảnh kết quả của thuật toán đề xuất (f).

**(a) (b) (c)**

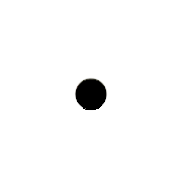


**(d) (e) (f)**

Hình 9. Kết quả kiểm thử với example\_1

**  **

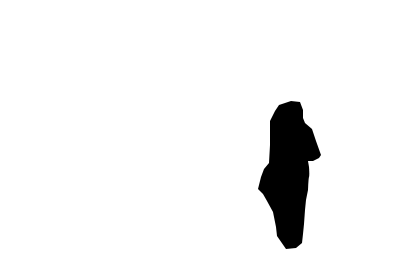
**(a) (b) (c)**

**  **

**(d) (e) (f)**

Hình 10. Kết quả kiểm thử với example\_2

**  **

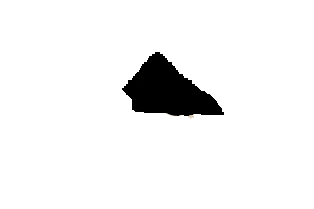
**(a) (b) (c)  **

**(d) (e) (f)**

**Hình 11. Kết quả kiểm thử với example\_3**

****  ** **

**(a) (b) (c)**

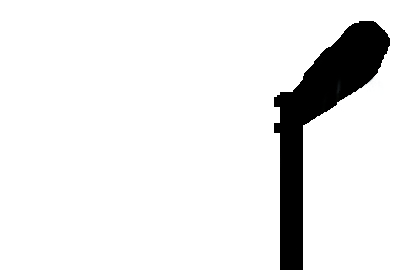
**  **

**(d) (e) (f)**

**Hình 12. Kết quả kiểm thử với example\_4**

**  **

**(a) (b) (c)**

**  **

**(d) (e) (f)**

**Hình 13. Kết quả kiểm thử với example\_5**

3.2. Thảo luận

Thuật toán đề xuất được đánh giá dựa trên hai khía cạnh: chất lượng và tốc độ. Để đánh giá chất lượng của các ảnh kết quả, hiện tại vẫn còn rất ít nghiên cứu đề xuất cách đánh giá chất lượng ảnh sau phục hồi một cách tin cậy và hiệu quả, do đó, trong nghiên cứu này, chúng tôi vẫn sử dụng cách đánh giá truyền thống dựa trên người quan sát [13]. Do hạn chế về điều kiện, nên chúng tôi mới chỉ đánh giá dựa trên số lượng người hạn chế (5 người) có độ tuổi trong khoảng từ 25-40 và dùng chung một máy tính để hiển thị các kết quả đánh giá cho người dùng quan sát. Trong đó, có 2 người có chuyên môn về xử lý ảnh và 3 người còn lại không có chuyên môn. Để đảm bảo tính khách quan, các cặp ảnh kết quả được sắp xếp ngẫu nhiên và người quan sát không biết chính xác ảnh nào được phục hồi bởi thuật toán nào (trước cải tiến hay sau cải tiến). Quá trình đánh giá sẽ được thực hiện lặp đi lặp lại 3 lần cho mỗi người quan sát, các kết quả thu được trong bảng sau:

***Bảng 3. Đánh giá theo quan sát của người dùng***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Ảnh** | **Người 1** | **Người 2** | **Người 3** | **Người 4** | **Người 5** |
| example\_1 | Output 2 | Output 1 | Output 2 | Output 2 | Output 2 |
| example\_2 | Output 1 | Output 1 | Output 1 | Output 1 | Output 1 |
| example\_3 | Output 1 | Output 2 | Output 1 | Output 1 | Output 1 |
| example\_4 | Output 2 | Output 2 | Output 2 | Output 2 | Output 2 |
| example\_5 | Output 1 | Output 1 | Output 1 | Output 1 | Output 1 |

Trong đó, output là các ảnh kết quả đầu ra được xem là tốt hơn sau khi xử lý. Với mỗi bộ ảnh đầu vào sẽ có hai ảnh đầu ra (output 1 và output 2) tương ứng với thuật toán trước cải tiến và sau cải tiến. Mỗi người dùng sẽ lựa chọn ảnh output nào là tốt hơn khi quan sát hai ảnh kết quả. Việc output nào là kết quả của thuật toán nào (trước cải tiến hay sau cải tiến) là không được biết trước và được lưu giữ trong bảng sau:

***Bảng 4. Bảng kết quả***

|  |  |  |
| --- | --- | --- |
| **Ảnh** | **Trước cải tiến** | **Sau cải tiến** |
| example\_1 | Output 1 | Output 2 |
| example\_2 | Output 2 | Output 1 |
| example\_3 | Output 2 | Output 1 |
| example\_4 | Output 1 | Output 2 |
| example\_5 | Output 1 | Output 2 |

Dựa trên bảng kết quả, hầu hết đều cho thấy phương pháp đề xuất cho kết quả tốt hơn so với kết quả chưa cải tiến.

Về hiệu suất của thuật toán, thời gian thực hiện của video trong Hình 9 và Hình 11 có chậm hơn khi thực hiện trên một ảnh nhưng cho kết quả khá tốt. Trong ví dụ ở Hình 10, bên cạnh việc thực hiện thời gian ngắn hơn thì kết quả hình ảnh cũng sắc nét hơn, kết quả của thuật toán đề xuất không bị mờ như trong kết quả của thuật toán trước cải tiến. Ở Hình 12, thuật toán đề xuất cũng cho kết quả nhanh hơn so với kết quả thực hiện của thuật toán trước cải tiến. Kết quả của thuật toán đề xuất của ví dụ ở Hình 13 cũng nhanh hơn so với kết quả thực hiện của thuật toán trước cải tiến.

4. Kết luận

Bài báo đã trình bày một framework cho việc phục hồi hoặc loại bỏ một đối tượng trong video dựa trên một mở rộng quá trình phục hồi ảnh. Video sẽ được cắt ra thành các frame ảnh riêng biệt và được xử lý lần lượt từng ảnh. Sau đó, các ảnh xử lý sẽ được ghép trở lại thành một video đầu ra. Trong bài báo này, chúng tôi sử dụng một giải pháp phục hồi ảnh dựa trên bản vá nhằm đảm bảo việc phục hồi các vùng ảnh lớn mà không bị các hiệu ứng mờ như các cách tiếp cận nội suy. Cách tiếp cận này là phù hợp cho việc phục hồi các ảnh tự nhiên, do đó nó khá phù hợp với việc xử lý các video phổ biến hiện nay. Ngoài ra, hai cải tiến mới đã được đề xuất liên quan đến việc tìm kiếm các bản vá sao chép nhằm giảm thiểu thời gian thực thi và việc sử dụng các ảnh liên tiếp thay vì một ảnh độc lập nhằm nâng cao chất lượng phục hồi. Các kết quả thực nghiệm cho thấy đề xuất của chúng tôi thực hiện khá tốt so với các kết quả khác. Tuy nhiên, trong một số trường hợp việc phục hồi các cấu trúc ảnh vẫn chưa được tối ưu do sự hạn chế về dữ liệu trong vùng ảnh nguồn. Đây được xem là một trong những vấn đề khá phức tạp và sẽ được nghiên cứu thêm trong tương lai.

**TÀI LIỆU THAM KHẢO**

1. Vaishali U. Gaikwad, P. V. Kulkarni, 2013. *Exemplar-based Video Inpainting for Occluded Objects*. CSE Department, CSE Department, GECA, Aurangabad, India.
2. M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, 2000. *Image inpainting*. In Proc. SIGGRAPH, tr. 417–424.
3. Folkmar Bornemann and Tom Marz, 2007. Fast image inpainting based on coherence transport. Zentrum Mathematik, Technische Universitat Munchen, Germany.
4. Tony F.Chan and Jianhong Shen, 2001. *Non\_Texture inpainting by Curvature-driven diffusions (CDD).* In Journal of Visual Communication and Image Representation, vol 4. tr. 436-449.
5. M. Ashikhmin, 2001, *Synthesizing natural textures*. The proceedings of 2001  
   ACM Symposium on Interactive 3D Graphics, tr. 217-226.
6. A. Efros and T. Leung, 1999. *Texture synthesis by non-parametric sampling*. In Proc. Int. Conf. Computer Vision, tr. 1033–1038.
7. Alexey Zalesny, Vittorio Ferrari, Geert Caenen, and Luc Van Gool, 2002. *Parallel Composite Texture Synthesis*. Swiss Federal Institute of Technology Zurich, Switzerlan.
8. A. Criminisi, P. Perez, and K. Toyama, Apr. 2004. *Region filling and object removal by exemplar-based image inpainting*. IEEE Transaction of Image Processing, vol. 13 (9), tr. 1200-1212.
9. Y. Wexler, E. Shechtman, and M. Irani, 2004. *Space-time video completion*. Proceedings. IEEE Computer Society Conference on Computer Vision and Pattern Recognition.
10. K.A. Patwardhan, G. Sapiro, and M. Bertalmio, 2005. *Video inpainting of occluding and occluded objects*. in Proc. ICIP. Vol. II, tr. 69-72.
11. Y. Zhang, J. Xiao, and M. Shah, 2005. *Motion layer based object removal in videos.* Workshop on Applications of Computer Vision.
12. Philippe Colantoni and Al, 2004. *Color Space Transformations.*
13. Streijl, Robert C., Stefan Winkler, and David S. Hands, 2016. *Mean opinion score (MOS) revisited: methods and applications, limitations and alternatives*, Multimedia Systems 22.2, tr. 213-227.

**ABSTRACT**

**OBJECT REMOVAL IN VIDEO**

Dang Thanh Trung, Nguyen Thi Ngoc Anh and Vu Dinh Thuan

*1Faculty of Information Technology, Hanoi National University of Education*

*2Faculty of Natural Sciences, Thai Binh Teachers Training College*

*3Scientific Research and International Cooperation Dept, Thai Binh Teachers Training College*

Image/Video inpainting both are considered as a hot research direction in computer vision field. The main goal is aim to automatically restore or remove an object in image or video so that it is undetectable by viewers. The difference between image and video inpainting is the restoration is implemented based on not only spatial information but also temporal information from video streams. The complex of video inpainting approaches is much higher than complex of image inpainting approaches because of background, light direction, motion, etc. In this paper, two main improvements of video inpainting is introduced including: narrow search space (for speed up inpainting) and using temporal information for similarity (for inpainting quality). The experimental results show that our proposals work robust and efficiently. It represents a strong potential for future research.

***Từ khóa:*** Video inpainting, video restoration, Object removal, image inpainting, image restoration.