Video Inpainting Using Seam Carving Method

Matheel E. Abdulmunim and Rana Mohamed H. Zaki

Computer Science Department, University of Technology, Baghdad, Iraq

Received 4 May 2016  Accepted 20 June 2016

Abstract

Video became something important of our daily lives and in many areas such as surveillance cameras and mobile cameras therefore it has attracted considerable attention in multimedia and computer vision researches in recent years. One such research is video inpainting. Video inpainting aims to restore missing or tainted regions or remove objects present in a video sequence from neighboring scenes. The proposed method in video inpainting is object removal technique based on seam carving algorithm, where the static AVI video type and object dynamic are used and working video for different number of frames per-second with resolution (320 pixel by 240 pixel) and (640 pixel by 480 pixel). The implementation in the proposed seam carving algorithm for object removal is obvious that discrete Laplacian edge detection filter gives the highest matching with four videos, which proves that its work is the best in detecting the energy of a video, and the best accuracy for whole system is (86.57516).

Keywords: Video, Inpainting, Seam Carving, Laplacian.
Video Inpainting Using Seam Carving Method

Matheel E. Abdulmunim and Rana Mohamed H. Zaki

طلاء الفيديو باستخدام طريقة نحت التماس

مثيل عماد الدين عبدالمنعم و رنا محمد حسن زكي

قسم علوم الحاسوب - الجامعة التكنولوجية

الخلاصة

أصبح الفيديو شيء مهم في حياتنا اليومية وفي العديد من المجالات مثل كاميرات المراقبة والكاميرات المحمولة لذا فقد استقطب اهتماما كبيرا في مجال البحوث في الوسائط المتعددة والرؤية بالحاسوب في السنوات الأخيرة. وأحد هذه البحوث هو طلاء الفيديو الذي يهدف إلى استعادة المناطق المفقودة أو الملوثة أو إزالة الأجسام الموجودة في تسلسل الفيديو من مشاهد المجاورة. إن الطريقة المقترحة في طلاء الفيديو هو تقنية إزالة الكائن بناءاً على خوارزمية نحت التماس، حيث يكون الفيديو ثابت من نوع AVI والكائن المستخدم متحرك ويعمل الفيديو على عدد مختلف من الاطارات في الثانية 480X680 و 240X320. تجربة خوارزمية نحت التماس المقترحة لإزالة الكائن من الواضح أن مرشح الكشف عن حافة لابلاسين يعطي أعلى مطابقة مع أربعة أشرطة الفيديو، وهو ما يثبت أن عملها هو الأفضل في اكتشاف الطاقة من شريط فيديو. وأفضل دقة للنظام كله (68.56). 

الكلمات المفتاحية: فيديو،طلاء،نحت التماس،لابلاسين.

Introduction

Inpainting historically painters to be removing flaw from photos and paintings manually, it was called (image inpainting). Image inpainting the concept exists very long years back and from the birth of computer vision, and researchers are looking for a way to do this process automatically [1], developed process to remove the certain areas or restoration a damaged area in a video is known (video inpainting). There are many algorithms and applications of image inpainting. It can be used in cinema and photography for “restoration”, to remove effects such as scratches, dust spot from images, removal of superimposed text like dates, publicity, or subtitles, (called deterioration). It can also be used for adding or removing object and elements from image/video. In this paper, seam carving operator was implemented to remove object from each frame by using fivetypes of edge detection filters to detect the energy of object in
Video Inpainting Using Seam Carving Method

Matheel E. Abdulmunim and Rana Mohamed H. Zaki

frame. This paper will be as follow: section 2 give an idea about Object Detection and Removal Principle, Section 3 present seam carving method, and Section 4 gives proposed algorithms implementation. Section 5 gives an experimental tests and results. Section 6 evaluated and measuring performance using confusion matrix. Finally section 7 presents conclusions.

Object Detection and Removal Principle
Track and object detect plays a main role in a lot of PC vision and pattern recognition presentations as vehicle navigation, surveillance and autonomous robot navigation. Object detection includes detecting objects and distinguishing patterns in the frame of video series. It needs a mechanism for object detection in each tracking technique either in every frame or while the object appears first in the video. The information used in unique frame is the greatest common technique for object detection. Some techniques of detect the object use the progressive information calculated from analyze series of frames with a view to decrease the number of true detections and raise accuracy rate [1]. Remove an object from the image should require detection of the object first based on detection techniques [2]. Digital image processing and remove strangers about the photographic image and logos from images/videos are at the time referred to as the object removal. Many algorithms were used to object removal from digital photographs, one of these algorithms is called (object removal by exemplar based inpainting). The exemplar based methods implement fine for level textures and extended linear image structures called (for propagating) [3] [4]. One the most famous and best algorithms to remove the object called seam carving [5].

Seam carving method
Seam carving was implemented according to algorithm proposed by “Shai Avidan and Ariel Shamir in 2007” [5]. The basic algorithm is elegant and quite simple. Such as includes many applications (enlarging, shrink, cropping, and object removal, etc.). Seam carving technique is extended to video retargeting using forward energy seam carving by remove the seam from an image and enters new pixel neighbors to the image [6] [7].

A seam is defined as an optimum 8-connected path of pixels on one image which extends from either left to right or top to bottom. Optimality is known by image energy function and try to label pixels based on amount of content owned. Through inserting seams or carving out in both
directions called (change size image). Seam carving may be used to image object removal and image content enhancement. Seam carving is used to remove the pixels with “low energy” and avoid removing pixels with “high energy” [5].

1. Finding energy using edge detection
The seam carving algorithm supports the least amount of information in an image that can be computed in the use of some kinds of energy functions like:
1- Edgeness gradient magnitude.
2- Entropy.
3- Histogram of Gradient (HOG).
4- Saliency.

Edge detection is the process lack of continuity or sudden change in some visual properties (such as the color, intensity of lighting, composition) and treatments these processes is very important to understand the images and the analysis can be used in the process of distinction objects in digital images [8,9]. Edge detection is traditionally performs by convolving the signal in computer vision with some kind of linear filter usually a filter that approximation a first or second derivative operator [10]. There are two methods to obtained edges using gradient G of intensity change [I(x,y)]. The first order gradient operators are Sobel operator, Prewitt operator, Frei-Chen masks, and First order derivative of Gaussian. The second order gradient operator is Discrete Laplacian. When convolving these filters with an image, the obtained result is gradient along x and y direction [1]. Once getting those gradients, magnitude is computed using the following formula:

$$e_1 (I) = \left| \frac{\partial}{\partial x} I \right| + \left| \frac{\partial}{\partial y} I \right|$$

2. Select the seam
The user must choose whether he wants carve horizontal seams, vertical seams, or a combination of both horizontal and vertical seams.
Let L be an n × m image.
A vertical seam is computed as:

$$s^k = \{ s_i^k \}_{i=1}^{n} = \{ (x(i), i) \}_{i=1}^{n}, \text{ s.t. } \forall i, |x(i) - x(i-1)| \leq 1,$$

Where x is a mapping x: [1, . . . , n] [1, . . . , m]

a vertical seam

Is an 8-connected path of pixels in the image from top to bottom.

Containing one, and only one, pixel in each row of the image [11] [12].

The pixels of the vertical seam \( \{ S_i \} \) thus be

$$L_S = \{ L(S_i) \}_i^n = 1 \ \{ L(x(i), i) \}_i^n = 1$$

### 3. Computed Seam Cost

The cumulative minimum energy was computed for all possible connected seams at each entry \((i,j)\) using dynamic programming to find lowest energy approach according to the following:

**Forward Pass (top to bottom for finding seam)**

Forward formula computes the least energy to be removed. Forward formula is:

$$C_L(i, j) = |I(i, j + 1) - I(i, j - 1)| + |I(i - 1, j) - I(i, j - 1)|$$

$$C_V(i, j) = |I(i, j + 1) - I(i, j - 1)|$$

$$C_R(i, j) = |I(i, j + 1) - I(i, j - 1)| + |I(i - 1, j) - I(i, j + 1)|$$

$$M(i, j) = \min \left( \begin{array}{c} M(i - 1, j - 1) + C_L(i, j) \\ M(i - 1, j) + C_V(i, j) \\ M(i - 1, j + 1) + C_R(i, j) \end{array} \right)$$

**Backward Pass (bottom to top for finding seam)**

Backward ward formula computes the cost of removing seams that insert the least energy to an image[5]. Backward formula is

$$M(i, j) = \text{Energy}(i, j) + \min(M(i-1, j-1), M(i-1, j), M(i-1, j+1))$$

**Proposed algorithm implementation**

The idea of the proposed algorithm depends on the monitoring of any building by using a static camera. The system plays the AVI video file format that is captured by using a
camera. Digital video is a kind of video record system that working by use each frame extraction passed through preprocessing. There are several preprocessing techniques used in proposed system they are: motion detection to detect the object to remove it, and reconstruction as illustrated in Figure (1). In this paper, a color AVI video type is used for object removal using seam carving.

![Figure (1): The main block diagram of inpainting video using seam carving algorithm.](image)

1. Main Phases of proposed work

In this section a video inpainting based on object removal will be given as shown in Figure (1) which consists of different stages.

1.1 Video Samples

1. Offline Mode

In offline mode one, can use a video record or data depending on universities, the input video format is AVI. An AVI file store video data. Usually the video data is stored in AVI files in compressed format with various parameters and codes.
2. Video structure

Video structure plays very important role in the understanding of video, it is well known about structure contains measure of frame size, if it has a width of W pixels and a height of H pixels at Color Depth (CD) of 24 bits, and the rate at which frames are displayed in Frames Per Second (FPS).

The most important video properties are video size and bit rate. The formulas related to these two with all other properties are:

\[
\text{Bit rate (BR)} = W \times H \times CD \times \text{FPS} \quad (9)
\]

\[
\text{Video size (VS)} = BR \times \text{Time} \quad (10)
\]

3. Video to Frame Extraction

After obtaining the multimedia, which includes video, AVI file is read and it is decomposed into frames and store in one dimension matrix.

2. Preprocessing of each frame

The preprocessing of each frame that includes many steps are illustrated in Figure (2).

---

**Figure (2):** Block diagram for proposed preprocessing algorithm.
a) Motion Detection: This section focuses on the detection of moving objects in a series of captured frames of a static camera. The most common algorithms are to compare the current frame with the previous frame. It’s useful to estimate changes. This algorithm presents an image with (RGB) on the place where the current frame is different from the previous frame. So it is possible to count the pixels and different frames calculated the difference between two frames. After that, the absolute difference is stored. If the result of the difference between frames equals zero means there is no lack of movement object, then working to exclude frame is done. But if the result is not equal zero, this frame goes to treatment. Let’s take an example; if one takes a sequence of frames, the current frame and the previous frame are considered in every calculation. After this calculation, the current frame becomes previous and the frame that comes in the sequence becomes the current frame in a loop. The end produced about a motion event.

b) Conversion to Grayscale

Each pixel in different frame contains (RGB) must be converted to gray scale image by using luminance method.

c) Convert to binary (use threshold)

After the frame differencing process, each pixel in different frame contains gray scale image must be convert to binary image by using threshold value and then the moving object has been identified. When the pixel corresponding to the moving object is set to 1 and rest is dealt with as background which sets to 0.

d) Connected Component Label (CCL)

After object detect the moving, it is labeled. The primary aim of the labeled tool is to decrease the manual work that is required in detection, this means frame that contains label that indicates an object therefore it will continue to work in the processing on the convex hull for object and frame that do not contain label.

e) Object Bounding Method

After determine the label to see the presence of the object it became possible to identify the object by more accuracy and that is through the use of convex hull and bounding box method.

1- Convex hull method
Convex hull is work on binary image and returns a two dimension matrix which sets the minimum convex polygon that can include the region.

2- Bounding box method

Bounding box painted on the convex hull of an object is a rectangular box requires for parameters(x, y, width and high) to select the object. The bounding box is performed around the moving object in every frame of the video. After knowing the dimensions and location of the object for each binary frame it is now possible to remove the object using the algorithm of seam carving.

3. Seam Carving Algorithm

To remove the object, Seam Carving (SC) algorithm is used. SC is an active method to contented aware image retargeting. At determine the beginning and end of the bounding box by rounding the object and make the value of the bounding box equal to zero and with less energy, so pass the algorithm of seam carving to remove the object.

The SC algorithm is passing through several stages in order to get the best seam removable:

1- Energy Function

The first stage is the energy function that computed by many methods in the energy gradient that was selected because it is fast to find the least noticeable pixels and it was removed. The edge detection filters that were used to find the energy are: Sobel, Prewitt, Discrete Laplacian, Frei – Chens and First order derivative of Gaussian. When convolving these filters with an image, the obtained result is gradient along x and y direction. Once getting those gradients, magnitude is computed energy, as shown in algorithm (1).

<table>
<thead>
<tr>
<th>Algorithm (1): Compute magnitude algorithm.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Input:</strong> Image ,Sobel mask horizontal and vertical operators // im, Hₓ, Hᵧ</td>
</tr>
<tr>
<td><strong>Output:</strong> Magnitude image // mag.</td>
</tr>
<tr>
<td><strong>Step1:</strong> Start</td>
</tr>
<tr>
<td><strong>Step2:</strong> Read image matrix and convolution with Hₓ and returns the centric portion of the convolution of the similar size as image to find image gradient = gₓ</td>
</tr>
<tr>
<td><strong>Step3:</strong> Read image matrix and convolution with Hᵧ and returns the centric portion of the convolution of the similar size as image to find image gradient = gᵧ</td>
</tr>
<tr>
<td><strong>Step4:</strong> mag = sqrt ( gₓ * gₓ + gᵧ * gᵧ )</td>
</tr>
<tr>
<td><strong>Step5:</strong> Return mag ;</td>
</tr>
<tr>
<td><strong>Step6:</strong> End.</td>
</tr>
</tbody>
</table>
Compute Seam Costs

The second stage is to pass the frame from the second row to the last row the cumulative minimum energy was computed for all thinkable related seams at all entry (i,j) using dynamic programming approach according to the following forward. After computing M seams keep track the minimal values, Algorithm (2) illustrated steps to finding forward energy.

2- Optimal Seam Path and remove

The last stage is to select the optimal seam path from the smallest value in the last row that refer to finish of the minimum of linked vertical seam, and remove the whole column with lowest energy in each row. Backtracking about that minimum energy was posted to find the optimal path seam for deleted.

To shrink an image by removing K vertical seams, in the first iteration, find a path along the cumulative minimum energy and remove that path, then shift all the pixels in the image to the
The output of one iteration will be the input of the next iteration. Algorithm (3) illustrates the steps to find the optimal seam path.

<table>
<thead>
<tr>
<th>Algorithm (3): optimal seam path algorithm.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Input</strong>: forward energy matrix for each frame // [seam]</td>
</tr>
<tr>
<td><strong>Output</strong>: Optimal seam path // k and store the index of minimum column //y is track</td>
</tr>
</tbody>
</table>

**Step 1:** Start

**Step 2:** Read matrix for forward energy and return number of rows, number of columns // [m, n]=size(mag)

**Step 3:** Create matrix contains the one column and many rows equal zeros // y= zeros(1,m)

**Step 4:** Find the minimum cost from the last row of cost array because the cost calculation is from top to down (forward) using dynamic programming

**Step 5:** Compute minimum of (the north, north east, north west) for each pixel inorder to store the optimal connected minimum path along the image.

For i=m-1 to 1 step -1

**Step 6:** Index the current column (i=m-1 from bottom to up calculation to find optimal seam path) // j=y(i+1)

**Step 7:** Check if the col=1, take the south and south east of the pixels in column 1 // (j==1)

if (seam_cost(i,j) <= seam_cost(i,j+1))

y(i)=j;
else
y(i)=j+1;
end

**Step 8:** Check if the col=n, take the south and south west of the pixels in column n // (y(i+1)==n)

if (seam_cost(i,j) <= seam_cost(i,j-1))

y(i)=j;
else
y(i)=j-1;
end

**Step 9:** else

**Step 10:** if (seam_cost(i,j) <= seam_cost(i,j-1))

if (seam_cost(i,j) <= seam_cost(i,j+1))

y(i)=j;
else
y(i)=j+1;
end

**Step 11:** else

if (seam_cost(i,j-1) <= seam_cost(i,j+1))

y(i)=j-1;
else
y(i)=j+1;
**Step 12:** return to step 5 to find the optimal seam row

**Step 13:** End.

After selecting the best seam deleting the seam was done, which is removing the pixels with lowest energy to three channels (R, G and B). This idea is shown in algorithm (4).
Algorithm (4): Remove seam algorithm.

<table>
<thead>
<tr>
<th>Input:</th>
<th>Optimal seam path // k and store the index of minimum column // y is track</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output:</td>
<td>Remove the optimal pixels with lowest energy // k</td>
</tr>
</tbody>
</table>

Step 1: After saving the index of the current column, (i=m-1 from bottom to up calculation to find optimal seam path) // j=y(i+1)

Step 2: for i = 1 to last row

Step 3: If the pixel to be removed is in the first column skip the first column and copy from the second column

Step 4: draw the seam store black pixel // rgb_img(i,1,:) = 0;

Step 5: Else

Step 6: If the pixel to be removed is in the last column skip the last column and copy until last column.

Step 7: draw the seam store black pixel // rgb_img(i,n,:) = 0;

Step 8: Else

Step 9: copy until reaching the specified column of the pixel to be removed and skip then restart the copy from the next column

Step 10: draw the seam store black pixel // rgb_img(i,y(i),:) = 0;

Step 11: return to step 2 until to remove all optimal seam path.

Step 12: End.

4. Resizing

When object remove from each frame, the result has a different size of the similar frame and characteristic vector to be various length. Therefore, there is a necessity to reform both the data for all the vectors of images to be similar size as is the in equation (10). The idea is show in figure (3). And Algorithm (5) is shown after the size of frame is resized.

\[
\text{resised}(m, n) = \sum_{y=1}^{n} \sum_{x=1}^{m} \text{current frame } (x,y)
\]  

(11)

Where

original frame in high = x

original frame in width = y

Target frame in high

Target frame in width

After that,

\[
\text{ratio of source} = \frac{\text{original frame in width}}{\text{original frame in high}}
\]  

(12)

\[
\text{ratio of target} = \frac{\text{Target frame in width}}{\text{Target frame in high}}
\]  

(13)
Video Inpainting Using Seam Carving Method

Matheel E. Abdulmunim and Rana Mohamed H. Zaki

level of width = \[
\frac{\text{original frame in width}}{\text{Target frame in width}}
\] (14)

level of high = \[
\frac{\text{original frame in high}}{\text{Target frame in high}}
\] (15)

width resized = \[
\frac{\text{original frame in width}}{\text{level of width}}
\] (16)

heigh resized = \[
\frac{\text{original frame in high}}{\text{level of high}}
\] (17)

Figure (3): A block diagram of object removal for each frame and resizing.

<table>
<thead>
<tr>
<th>Algorithm (5): frame Resizing.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Input:</strong> Frames // array of frames, FW // Frame of Width, FH // Frame of High.</td>
</tr>
<tr>
<td><strong>Output:</strong> Resizing frame // RES</td>
</tr>
<tr>
<td>Step1: Start</td>
</tr>
<tr>
<td>Step2: Become high and width of original frame, ho=height, wo=width</td>
</tr>
<tr>
<td>Step3: Compute the ratio of original, (ro = \frac{ho}{wo})</td>
</tr>
<tr>
<td>Step4: Compute the ratio of target, (rt = \frac{50}{50})</td>
</tr>
<tr>
<td>Step5: Find level of width, (lw = \frac{\text{original frame in width}}{\text{Target frame in width}})</td>
</tr>
<tr>
<td>Step6: Find level of high, (lh = \frac{\text{original frame in high}}{\text{Target frame in high}})</td>
</tr>
<tr>
<td>Step7: Compute width resized, (wr = \frac{\text{original frame in width}}{\text{level of width}})</td>
</tr>
<tr>
<td>Step8: Compute high resized, (hr = \frac{\text{original frame in high}}{\text{level of high}})</td>
</tr>
<tr>
<td>Step9: Return RES, wr and hr.</td>
</tr>
<tr>
<td>Step9: End.</td>
</tr>
</tbody>
</table>
5. Reconstruction Process

The implementation of a reconstruction is come after applying the object removal by seam carving method and resizing for each frame in order to get the cell array of frames to display video. The algorithm of reconstruction video is illustrated in algorithm (6).

<table>
<thead>
<tr>
<th>Algorithm (6): Reconstruction frames.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Input:</strong> Video read /vid=video structure, file name=name of video file and path= location of video file.</td>
</tr>
<tr>
<td><strong>Output:</strong> Display video after implementing a removal object and resizing process</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Step1: Start</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step2: Frm = array of frames</td>
</tr>
<tr>
<td>Step3: N of Frm = number of frames</td>
</tr>
<tr>
<td>Step4: Frm rate =15 // Determine the frame rate in the per-second</td>
</tr>
<tr>
<td>Step5: Video write (video name, path) For i = 1 to N of Frm</td>
</tr>
<tr>
<td>Step6: Write video (video, N of Frm (i), Frm rate) //Start writing the video with name of video, number of frame and frame rate</td>
</tr>
<tr>
<td>Step7: Return (i)</td>
</tr>
<tr>
<td>Step9: End.</td>
</tr>
</tbody>
</table>

**Experimental Tests and Results**

1- Using image preprocessing includes (Motion detection, Convert to binary, Connected component label, Object bounding method and Object removal) and last which can be used in the system, as show in Table(1).

<table>
<thead>
<tr>
<th>Table (1): Sample of AVI video.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample of AVI video</td>
</tr>
<tr>
<td>---------------------</td>
</tr>
<tr>
<td><img src="image1" alt="Sample of AVI video" /></td>
</tr>
</tbody>
</table>
Evaluated and measuring performance using confusion matrix

The definite table planning that allows conception of the presentation or the accuracy of an algorithm is called Confusionmatrix. All column of the matrix represents the cases in a predicted class, though all rows denote the cases in a real class. Confusion matrix is a table that has four elements that determination the number of False Positives (FP), False Negatives (FN), True Positives (TP), and True Negatives (TN). Many further calculations can be taken from these elements. So,

\[
\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}.
\]

In this paper confusion matrix is used to evaluate the energy detected by the five energy detection filters, as shown in Table (2).

<table>
<thead>
<tr>
<th>Edge Detection Filter</th>
<th>TP</th>
<th>TN</th>
<th>TF</th>
<th>FN</th>
<th>Accuracy %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sobel</td>
<td>46230.4761</td>
<td>201021.190</td>
<td>41168.5238</td>
<td>18779.8095</td>
<td>80.48557</td>
</tr>
<tr>
<td>Prewitt</td>
<td>46235.5238</td>
<td>201025.238</td>
<td>41164.4761</td>
<td>18774.7619</td>
<td>80.48853</td>
</tr>
<tr>
<td>Discrete Laplacian</td>
<td>46239.0952</td>
<td>201183.809</td>
<td>41005.9047</td>
<td>18771.1904</td>
<td>80.54130</td>
</tr>
<tr>
<td>Frei - Chens</td>
<td>46230.4761</td>
<td>201021.190</td>
<td>41168.5238</td>
<td>18779.8095</td>
<td>80.48557</td>
</tr>
<tr>
<td>First order derivative of</td>
<td>46241.2857</td>
<td>201112</td>
<td>41077.7142</td>
<td>18769</td>
<td>80.51865</td>
</tr>
</tbody>
</table>

From the tables shown, it is obvious that discrete laplacian edge detection filter gives the highest matching, which proves that its work is the best in detecting the energy of the video.

Conclusions

The proposed method can process the pre-processing step to moving object detection, bounding object and removal object using seam carving algorithm automatically that will implemented successfully in all frames of videos and when using a variable size of video with different resolution in addition to the different number of frames for each video, it is obvious that Discrete Laplacian edge detection filter gives the highest matching with the video, which proves that its work is the best in detecting the energy of a frame. The accuracy of Discrete Laplacian detection filter is (80.54).
Video Inpainting Using Seam Carving Method

Matheel E. Abdulmunim  and  Rana Mohamed H. Zaki

References


