Implementing Discontinuous Seam Carving for Video Retargeting

https://github.com/dillonmchenry/discontinuous_seam_carving

David Lazauskas
dal346@drexel.edu

Dillon McHenry
dsm98@drexel.edu

Davis Ranney
dr847@drexel.edu

Abstract

We've implemented a method of seam-carving for video resizing originally published by Matthias Grundmann et al. at the Georgia Institute of Technology which calculates seams in an image or video that are both spatially and temporally discontinuous from one another. Their process goes against the previous video resizing convention to maintain the smoothness of seams across the video volume by utilizing 3D volume surfaces for seams. The algorithm uses appearance-based temporal coherence to compare a frame with its optimal temporally coherent version. Spatial coherence is then calculated to minimize the gradients that occur per pixel when retargeting. Lastly, rather than applying the paper's automatic spatial-temporal saliency, we calculate a saliency map using a forward energy function. These three calculations become a weighted combination that informs the path of a discontinuous seam. In addition to implementing this algorithm, we experiment with varying pixel-jump limits for seams as well as the window size considered when computing spatial coherence per pixel.

1. Introduction

Video retargeting has become more important recently with the multitude of devices used to playback video. Initial attempts to adapt video sizes used either direct scaling or cropping of selected regions which often removed or deformed important content from each frame, or added black bars to the edges of content to fill out the screen, which doesn't maximize screen real estate. Initial seam-carving techniques made informed decisions on which content "seams" to remove based on an image's energy loss [4]. While this method worked well for images, it was computationally expensive and created artifacted results when applied to video. Common issues were warping of the cut regions (poor coherence), excessive seam jumping between frames, and poor spatial awareness of the source.

Later research in seam-carving for video resizing found that ensuring geometric smoothness in the placement of seams across the video frames improved the previously artifacted results [2]. Grundmann's paper poses that geometric smoothness remains sufficient but should not be overly constrained for coherent retargeting of videos. Instead, the authors tested an appearance-based temporal coherence measure on seams. When not prioritizing smoothness, spatial seams can be applied to a frame that varies by more than one pixel from row to row and large salient regions of the image can be circumvented by seam-hopping. In addition, the authors offer a method to improve spatial detail over seams by using a coherence measure that considers the variation in gradients among frames, as opposed to the gradient values themselves. This preserves more detail than

previous methods of minimizing color difference alone [1].

The greatest advantage of the Grundmann's seam carving method lies in its efficiency to resize on a frame by frame basis which increases time efficiency four-fold. Granted, we implemented a forward energy function in place of the authors' proposed automatic saliency measure which stunted the potential time optimization.

2. Related Works

Avidan's original paper on image seam-carving proposed their method as a way to consider image-content while resizing as opposed to raw-scaling [3]. Rows and columns would be removed from or added to images through continuous chains of pixels decided by minimizing energy function values.

Earlier contributions to seam-carving as a form of video retargeting modeled the removal of 2-D seams from a 3-D space-time volume (video) [2]. A 2-D seam was required to be monotonic and connected throughout the 3-D volume to be considered valid. Rather than seeking to remove seams with the least energy value, the operator now looks forward in time to calculate the seam that will *introduce* the least amount of energy into the retargeted frame [1].

As mentioned, the method we implemented removes the need for connection between 2-D seams both within frames and between frames. Instead, seams are chosen through minimizing the weighted combination of temporal coherence, spatial coherence, and gradient-based saliency. By utilizing temporally and spatially piece-wise seams, our implementation of "seam-hopping" is mitigated by the variable window involved in calculating spatial coherence.

3. Seam-Removal Method

In order to downsize a video by M x N pixels, we first iterate N times of calculating the least energy introducing seam and removing it from the current frame. To account for the horizontal down-sizing, the image is transposed and the same seam-removal process is applied M times.

As mentioned, the previous video retargeting method proposed by Rubenstein et al. transferred the image resizing approach of [1] into a video by representing the seams as a surface in the video volume. This ensures temporal coherence of seams over frames, with the obvious downside being that seams could only shift one pixel between frames. If a person were to move rapidly across the screen, a seam in an eventual frame would be doomed to intersect with the person of interest.

The algorithm in our paper of interest considers temporal coherence when deciding on pixels for seam paths, but not *absolutely.* If we only optimized seams on temporal coherence, the same seam would be chosen in every frame to maintain the alignment of the original video. The authors achieve temporally discontinuous seams by instead valuing the appearance of the resized frame when compared the same frame with the optimal temporally coherent seam applied. Implementation wise, this involves applying the previous seam's calculated seam to the current frame and calculating per pixel temporal coherence from there.

Seams being discontinuous also allows all seams to be calculated sequentially by frame. For each pixel in each frame, the spatial (Sc) and temporal (Tc) coherence costs are calculated before being placed in a weighted ratio with the saliency (S) map to create the measure *M*. The standard M weight ratio is Sc:Tc:S → 5:1:2. For highly dynamic video the ratio is 5:0.2:2. The M values per pixel inform the decision for choosing the minimum cost adding a seam. We experimented with the ratios of *M* below.

3.1 Temporal Coherence

In order to accomplish an appearance-based temporal coherence, we want to remove a seam from the current frame such that the retargeted result is visually similar to the previous seam removed from that frame (temporally optimal).

To calculate per pixel (x,y), the authors offer the following equation where *Rc* represents the temporally optimal that is compared to *Ri*: the retargeted version of the original frame *Fi* [1]

$$T_c = \sum_{k=0}^{x-1} ||F^i_{k,y} - R^c_{k,y}||^2 + \sum_{k=x+1}^{m-1} ||F^i_{k,y} - R^c_{k-1,y}||^2. \quad (1)$$

These per pixel costs can be computed before any seams which allow temporal and spatial coherence to be combined.

3.2 Spatial Coherence

The proposed method for spatial coherence is similar to the forward energy proposed by Rubenstein in [2], but it is not based on intensity variation, rather a piecewise model of *variation in the gradient of intensity.* The spatial coherence is measured by $Sc = Sh + Sv$ which is the sum of spatial error induced in the horizontal and vertical directions. The following diagram from our article of interest models three pixels with no horizontal error, complete horizontal error, and minor horizontal error respectively [1].

(a) No detail lost (b) Detail lost (c) Linearity preserved

Svdepends both on the neighboring pixels and the best seam neighbor in the row above. Since we are creating discontinuous spatial seams, we can consider more neighbors than just the three above the pixel in question. Each pixel (xb, y) has a summed spatial transition cost for a pixel $(xa, y - 1)$ in the row above [1].

$$S'_v(x_b, x_a, y) = \sum_{k=x_a}^{x_b-1} |G^v_{k,y} - G^d_{k,y}| + \sum_{k=x_a+1}^{x_b} |G^v_{k,y} - G^d_{k-1,y}|$$

This summed transition cost is calculated on each above pixel that lies within the window of consideration. All of those sums comprise the spatial coherence of one pixel. We experiment with the window size for a pixel's spatial coherence sum in the following section.

Figure 1 *Nintendo inc.*

3.3 Saliency Mapping

To accentuate salient objects in images for seam-carving, a saliency map must be applied to the image. Our implementation of forward energy saliency applies a per-frame gradient-based saliency map (*figure 1*). While our method is serviceable for some video frames, it falls short for tasks involving abrupt movements and lighting changes between frames. In those cases, the proposed automatic spatio-temporal saliency [1] would be more appropriate considering saliency is averaged over temporal regions rather than per pixel. The result is a smooth variation in focused regions across frames. This method proves most useful for tasks such as face detection where tracking bounding boxes of heads between frames is necessary.

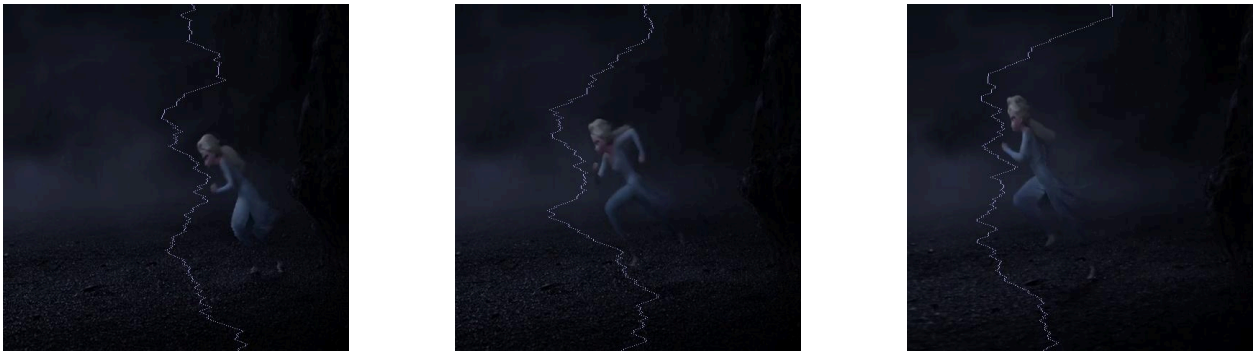

Figure 2 (Disney Productions)

4. Results and Limitations

Our retargeted videos met the expectation set by our referenced paper, but we could not meet the same quality of retargeting as [1] due to replacing auto spatial temporal saliency with a simple gradient-based saliency.

As evidenced in *figure 2*, our implementation still prevents rapidly moving objects from distorting through seam removal. This is due to the discontinuous nature of the algorithm that allows seams to jump and keep up with salient object movement.

Even when dealing with videos whose objects of interest were nearly identical in base color to the backdrop, proper seams were carved to dodge object distortion (*see figure 4*).

When experimenting with the window size allowed for seam jumps and spatial coherence calculation, we first experimented with a size of 15 pixels (as recommended by [1]). We experienced greater distortion in retargeted videos with this window size likely due to window taking up around 10% of the frame width. When reducing the window size to 10px or even 5px, we received more stable results because seam pixels lacked ability to leap across the screen.

We additionally experimented with the ratios of spatial coherence, temporal coherence and saliency to form the cost *M* per pixel. 5:0.2:2 was proposed as a respective ratio to process more dynamic video content. When retargeting the chaotic video game footage referenced in *figure 3*, using a ratio that lowered prioritization of temporal coherence resulted in less distortion overall. The results make sense because a video with constantly altering background should not considerably weigh the seam of the previous frame.

Our decision to not implement the auto temporal saliency posed by [1] limited the computational efficiency of our implementation as well as its ability to deal with highly dynamic video content. We did not come close to the paper's impressive 2 frames per second when shrinking a retargeted video. Videos with several moving components and rapidly changing backgrounds display heavy distortion in their retargeted size as evidenced by *figure 3*.

Figure 3
(Nintendo Inc)

Figure 4 (*LucasArts Entertainment*)

References

1. Grundmann, Matthias, et al. "Discontinuous Seam-Carving for Video
2. Retargeting." *Georgia Institute of Technology*, 2010.
3. Rubinstein, M., Shamir, A., Avidan, S. 2008. Improved Seam Carving for Video Retargeting. *ACM Trans. Graph. 27*, 3, Article 16 (August 2008), 9 pages. DOI = 10.1145/1360612.1360615 http://doi.acm.org/10.1145/1360612.1360615.
4. S. Avidan and A. Shamir. Seam carving for content-aware image resizing. ACM SIGGRAPH, 2007.