Novel View Synthesis from Two Cartoon Face Drawings

Hojin Choi
Kyunghee University
Seoul, Korea
hojinchoi@khu.ac.kr

Seungkyu Lee
Kyunghee University
Seoul, Korea
seungkyu@khu.ac.kr

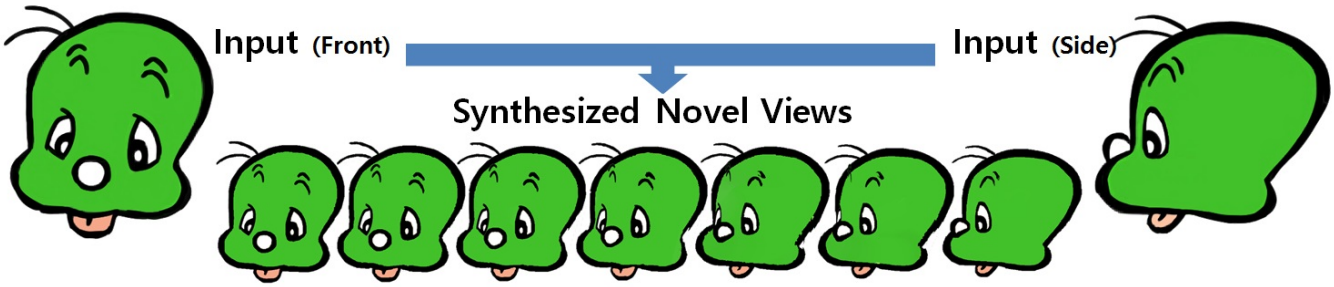

Figure 1: Cartoon faces synthesized by our framework. The leftmost and rightmost images are two input images from the user. Images in the middle are examples of automatically generated results. (Input images courtesy of Doolynara, Inc.)

ABSTRACT

We propose a framework for novel view synthesis for a cartoon face given two images of key views such as front and side views. Due to the structure inconsistency of cartoon faces in manually drawn images, matching region detection between different views is a challenging task. We propose a local complexity based segmentation that grabs facial components such as eyes, eyebrows and mouth that are not clearly separated by contour in cartoon faces. We automatically find the matching region in the two input key images and morph them to create intermediate views.

KEYWORDS

Cartoon synthesis, Segmentation, Registration, Image morphing

ACM Reference format:
Hojin Choi and Seungkyu Lee. 2017. Novel View Synthesis from Two Cartoon Face Drawings. In *Proceedings of SIGGRAPH '17 Posters, Los Angeles, CA, USA, July 30 - August 03, 2017,* 2 pages.
https://doi.org/10.1145/3102163.3102229

1 INTRODUCTION

In the traditional cartoon creation framework, key frames are drawn manually to create the illusion of motion. There have been some previous works which try to facilitate this process by generating intermediate images from two or more key frames (e.g., [Rivers et al. 2010]), although they require some form of human input to register and deform character parts. Our aim is to automatically generate new views of a cartoon character's face when given two different views of the face by the user. Cartoon objects usually contain structural inconsistency between views, which makes 3D models not directly applicable. This also results in facial regions often not having a clear boundary, making the detection and matching problem nontrivial and challenging. We deal with this problem by using a segmentation method based on local complexity hierarchy. We detect facial components such as eyes and mouth with entropy scale space, and segment them into smaller parts based on color information. We find the best matching of these parts between the two views and generate intermediate shapes by applying image morphing.

SIGGRAPH '17 Posters, July 30 - August 03, 2017, Los Angeles, CA, USA

ACM ISBN 978-1-4503-5015-0/17/07…$15.00
https://doi.org/10.1145/3102163.3102229

2 REGISTERING FACIAL COMPONENTS

In order to create an intermediate view, we first register the two input images by finding the correspondence between facial structures.

2.1 Facial Component Detection and Matching

To preserve the texture details in the face region as much as possible, images for important facial components such as eyes or mouth should be generated separately from the rest of the facial region. Therefore, we first detect these component regions and then match them between the two images. To do this, we create a scale space of entropy information, where each entropy map is affected by the map of the previous upper level. The final entropy map is processed to create a binary image that contains the detected regions. These regions are then matched between two images based on their relative position and color information. It is important to note that a component might not have any matching counterpart in the other image, because some parts visible in one view can be occluded in the other and vice versa. These components are excluded from the registration process that follows, and are handled separately during the image synthesis stage. As fig. 2 shows, our algorithm can detect

components that don't have a well defined outline, such as brows composed of several unconnected line segments.

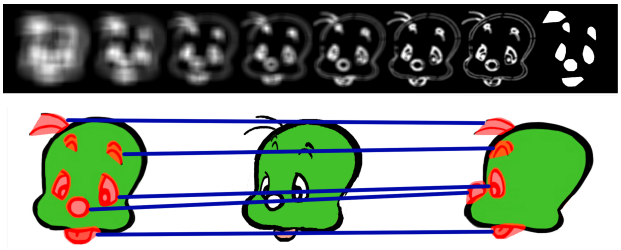

Figure 2: Facial component matching process. Entropy scale space is shown in the top row, with the resulting face regions displayed on the rightmost side. Bottom row shows the matching of facial components between two images, with the intermediate synthesized image in the middle.

2.2 Matching Local Regions within Facial Components

For each matched facial component pair, we create a mapping for the regions inside the components. We use color based segmentation to separate the components into distinct parts. For example, the eye component would be segmented into parts such as sclera, iris, pupil and highlight. We first apply Statistical Region Merging [Nock and Nielsen 2004]. Then we detect segments that are too small or too thin to be significant and merge them with neighboring segments that are most similar in color.

Figure 3: Segmentation result. From left to right: Original image, result of Statistical Region Merging [Nock and Nielsen 2004], and result after merging insignificant segments.

After segmentation, we need to register the segmented parts between the two input images. We form this as an assignment problem where we minimize the cost of assignment between the segments in two images. For each segmented part, we create a feature vector that consists of mean color values, area of the segment relative to the entire region, and the position of the segment relative to all other segments. Given two segments, the matching cost between them is initialized as the magnitude of their feature vector difference. Then we reduce the cost for every adjacent segment they have in common, and increase the cost for all segments that are adjacent to only one of them. Once we have calculated the pairwise matching costs, we use the Hungarian method to find the best assignment. Just like in the case of face components, it is possible that a segment does not have any matching counterpart because of occlusion. To detect these segments, we add dummy segments by inserting predefined maximum values into the cost matrix, and find all segments that match to these dummy segments by the Hungarian algorithm.

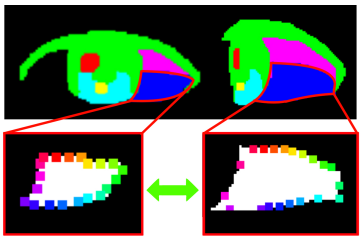

Figure 4: Result of registration. On the top, matched segments are shown in the same color. Segments that don't have any match are not shown. The bottom images show the registered control points in one segment.

3 GENERATING INTERMEDIATE IMAGES BY MORPHING

Once the registration of facial components is complete, we create intermediate images of each component by applying image morph. The control points for morphing are automatically detected as follows: For each facial component pair, we extract points from the boundary of the component and find the best matching between the two views based on the color, position and surrounding pixel shape. Then we generate the morphed image between the two views of the component. The morphed components are combined in order to generate the final morphed facial image. As for those components that do not have matching counterparts in the other input image, we simply translate them toward the direction which other components are moving in, and scale them down along the axis corresponding to that direction. This creates the illusion of said components gradually facing away from the viewer and eventually disappearing due to occlusion.

4 RESULTS AND DISCUSSION

Our current algorithm works relatively well with cartoon faces that are not too complex in shape and have sparse, well-defined facial components. In more difficult examples, however, it can fail to detect certain facial components or incorrectly groups multiple components into one. We plan to add color based information on top of the entropy maps to improve performance. Also, matching between facial components (and between each component pair's segmented parts as well) can be wrong if the shapes in two input images are significantly different from each other. One possible idea is to model the relative position between components as graphs and apply a graph matching method, so that the spatial relationship between components will be preserved while matching.

REFERENCES

Richard Nock and Frand Nielsen. 2004. Statistical Region Merging. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26, 11 (Nov. 2004), 1–7. https://doi.org/10.1109/tpami.2004.110

Alec Rivers, Takeo Igarashi, and Frédo Durand. 2010. 2.5D Cartoon Models. *ACM Transactions on Graphics* 29, 4 (July 2010).