Artistic Paper-Cut of Human Portraits

Meng Meng
Lotus Hill Institute &
Beijing Institute of Technology
mmeng.lhi@gmail.com

Mingtian Zhao
Lotus Hill Institute & UCLA
mtzhao@stat.ucla.edu

Song-Chun Zhu
Lotus Hill Institute & UCLA
sczhu@stat.ucla.edu

ABSTRACT

This paper presents a method to render artistic paper-cut of human portraits. Rendering paper-cut images from photographs can be considered as an inhomogeneous image binarization problem, to which ideal solutions should reproduce vivid image details with sparse cuts. Especially for portrait paper-cut, good artworks should capture impressive facial features. To achieve this goal, our approach integrates bottom-up and top-down cues to better determine the binary values. In the bottom-up phase, facial components are localized on the input photograph, and their draft binary versions are proposed. In the top-down phase, we use pre-collected representative paper-cut templates, with which we synthesize the final paper-cut image by matching them with the bottom-up proposals. Experimental results show that our approach can produce visually satisfactory results.

Categories and Subject Descriptors

I.3.3 [**Computer Graphics**]: Picture/Image Generation; I.4.10 [**Image Processing and Computer Vision**]: Image Representation—*Hierarchical*; J.5. [**Computer Applications**]: Arts and Humanities

General Terms

Algorithms

Keywords

And-Or Graph, Binarization, Paper-Cut, Portrait

1. INTRODUCTION

Artistic paper-cut is a traditional and popular Chinese decorative art which, usually in a very concise two-tone (red foreground and white background) form, has its unique beauty of expressive abstraction.

Researchers in the computer graphics community have been studying algorithms for rendering paper-cut images

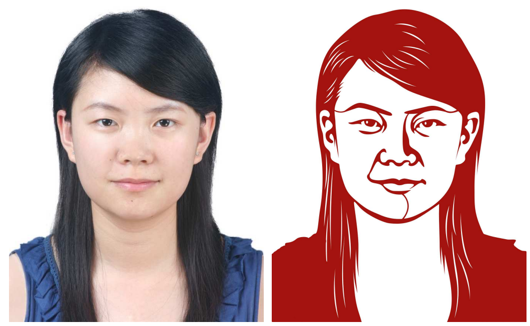

Figure 1: An example portrait paper-cut rendered using our method.

from photographs, which is usually considered as an inhomogeneous image binarization problem. In this direction, popular past methods include thresholding [5, 7, 9, 11], artistic energy optimization [10], etc.

However, it is noticed that good paper-cut artworks reflect not only the pixel intensities of the original photographs, but also plenty of hallucinated information by the artists according to certain prior knowledge. For example, in a portrait paper-cut, lots of facial details, especially around areas of eyes and nose, as well as some decorative curves are usually not observed in its corresponding photograph. In fact, artists often make up such features to emphasize certain characteristics of appearance or personality. To address this problem, we present a method to render artistic paper-cut images of human portraits by integrating bottom-up (data-originated) and top-down (prior) cues for better binarization. Fig.1 displays an example of our rendering result.

In our approach, a bottom-up phase is first executed to extract information from the input photograph. In this phase, facial components such as nose, eyes, etc., are localized using the active appearance model (AAM) [4], and draft binary versions of these components are proposed using a dynamic thresholding method. Then a second top-down phase is used for matching the above proposals with pre-collected representative paper-cut templates created by artists, which are organized using a hierarchical and compositional representation named And-Or graph (AOG) [6, 12]. This is for bringing in prior information to achieve fine details unavailable from the photograph. In addition, a post-processing phase is performed for rendering hair and clothes, and enforcing the connectivity of the foreground which is an important characteristic of artistic paper-cut.

2. REPRESENTATION

2.1 Paper-Cut Templates

We have asked professional artists to create red-and-white images for portrait photographs using common interactive image processing software, mimicking the paper-cut effects, then manually decompose them into facial components (see Fig.2 for a few examples). Keypoints describing the shapes of these components are also manually annotated on the images. We then construct a dictionary of paper-cut templates, including the above paper-cut portrait components and keypoints organized in a structural representation named And-Or graph (AOG). The detailed usage of this dictionary will be explained in Section 3.2.

2.2 And-Or Graph of Human Portraits

AOG is a hierarchical (tree-like) and compositional (part-based) model for structural information representation, consisting of *alternate* layers of And-nodes and Or-nodes downwards the tree from root node to leaf nodes. Each And-node represents the *decomposition* of a structural pattern (e.g., object) into its constituent sub-structures (e.g., parts), while each Or-node represents a virtual structural concept which *switches* among multichotomous instances. Readers may refer to the literature [6, 12] for detailed formulations of the AOG model.

We use a 3-layer AOG to organize human portrait photographs, as well as the dictionary of paper-cut templates. In this AOG, the root node is an And-node corresponding to the generic face, which is decomposed into facial components such as eyebrows, eyes, nose, mouth, etc.; each of them is an Or-node switching among a number of leaf nodes in the next level, corresponding to different instantial versions. For example, the "nose" Or-node has children nodes corresponding to instances of the nose region cropped from different source images. We consider the leaf nodes as atomic components which are usually perceived as single elements.

2.3 Parse Graph

While the whole AOG corresponds to a large number of configurations due to the existence of Or-nodes, an instance of the AOG, named *parse graph*, represents one specific version by choosing one branch for each Or-node. In our case, by selecting one specific mouth, one specific nose, etc., from the template dictionary, we can put together a set of facial components to form an entire face. Fig.3 visualizes the parse graph of a face image and its corresponding paper-cut.

Through different combinations of sub-structures, the number of possible parse graphs which can be instantiated from an AOG is much larger than the number of original examples used for constructing it. For example, in our case, we can combine a left eyebrow, a mouth from, etc., from two or more different original paper-cut templates into a new instance. In this way, AOG's great power of representation can help us render various portrait paper-cuts once we have a few representative templates.

3. COMPUTATION

3.1 The Bottom-Up Phase

We first localize the facial components in terms of their corresponding keypoints (as shown in Fig.4(a)) using the active appearance model (AAM) [4]. With the coordinates of these keypoints, we are able to extract image regions of the nose, eyes, eyebrows, etc., and we pass these image regions to the next binary proposal step.

Figure 2: Example paper-cut facial components in our template dictionary.

We use a dynamic thresholding method to compute binary proposals for the facial components. For each pixel, its binarization threshold is computed using Otsu's method [7] inside its neighborhood window, thereby different pixels may have different thresholds. By doing such binarization with different window sizes, a binary sequence for each facial component can be computed (see Fig.4(b)). It is noticed that small window sizes tend to keep more details due to the variation of local statistics over the spatial domain, on the contrary, large window sizes tend to produce smooth results by sharing similar global statistics among pixels close to each other (since the overlapping area between neighboring windows covers a large proportion of the window size).

For each facial component, we further select from the above binary sequence one version that best matches the original image in terms of total error. Let $\mathbf{I}$ be the original image region and $\mathbf{I}'_k$ be the kth binary proposal in the sequence, then we select

$$k = \arg\min_k |V(\mathbf{I}) - V(\mathbf{I}'_k)| \quad (1)$$

where V returns the sum of pixel intensities (8-bit gray-level or $(0, 255)$-binary) in the image region. The selection of k can be different for different facial components. Fig.4(c) shows the composition of selected binary components using the above method into a whole binary face image.

3.2 The Top-Down Phase

The top-down phase introduces prior information of artistic paper-cut to improve the bottom-up proposal. With the representation introduced in Section 2, the problem of generating a paper-cut by combining bottom-up and top-down cues actually becomes to instantiate a parse graph from the AOG of paper-cut template dictionary that best matches the bottom-up proposal, to ensure that the paper-cut looks like the original subject in the photograph.

We do this selection in the following way. Let

$$\mathcal{T}_{i,j,\cdots} = \{\text{nose}_i, \text{mouth}_j, \cdots\} \quad (2)$$

be a set of leaf nodes in the above AOG that form a whole face. Similarly, let

$$\mathcal{T}' = \{\text{nose}', \text{mouth}', \cdots\} \quad (3)$$

be the bottom-up proposal including both the binary image

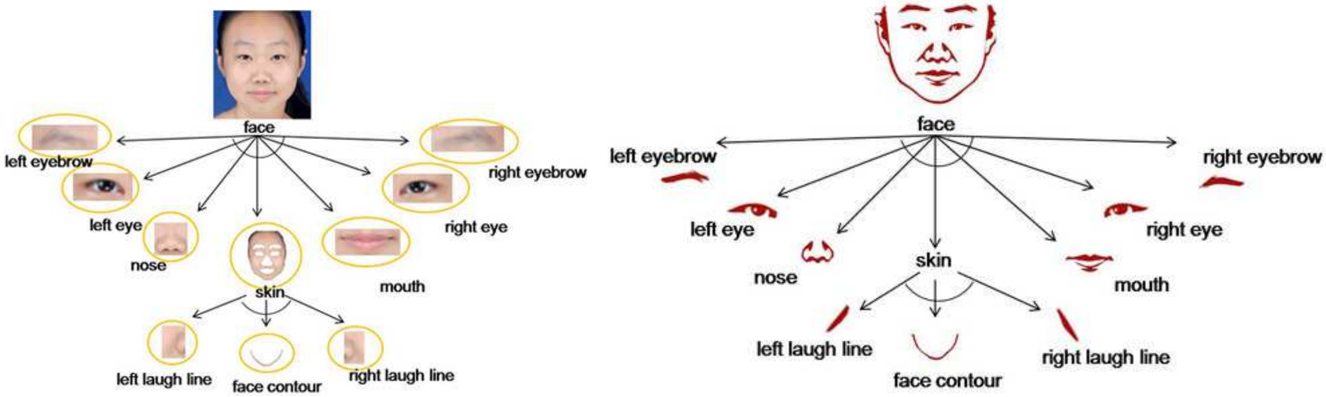

Figure 3: Parse graph of a face image and its corresponding paper-cut.

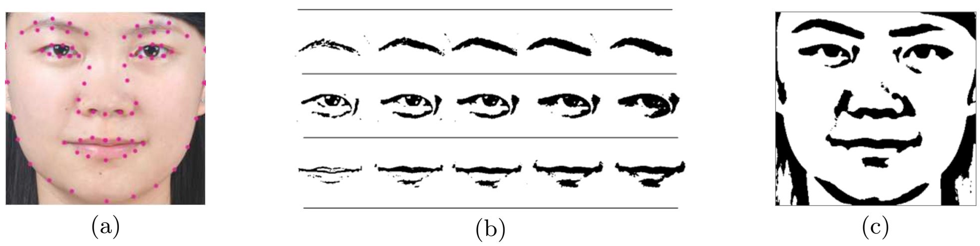

Figure 4: The bottom-up phase corresponding to the example in Fig.1. (a) shows the AAM keypoints. (b) shows the binary sequence of left eyebrow, left eye and mouth. (c) is the draft binary proposal.

and keypoints. We search for

$$(i, j, \cdots)^* = \arg\min_{i,j,\cdots} d(\mathcal{T}_{i,j,\cdots}, \mathcal{T}') + \lambda c(\mathcal{T}_{i,j,\cdots}) \qquad (4)$$

in which d is the distance between the template and the proposal, c is the count of original instances that $\mathcal{T}_{i,j,\cdots}$ covers (i.e., the number of different original paper-cut templates that the selected components are taken from), and λ is a tuning parameter. Intuitively, small λ leads to better matched facial components, while large λ leads to more compatible components in the result (since more components tend to come from the same original paper-cut) thus more consistent global styles, at the cost that the result paper-cut may be less similar to the same person in the photograph. In Eq.(4), the distance function d is defined as the sum of Euclidean distances between images of facial components in $\mathcal{T}_{i,j,\cdots}$ and $\mathcal{T}'$ after warping the former using thin plate spline (TPS) [1] to match the keypoints of the latter.

In implementation of the above search, we use a greedy approximate algorithm:
(1) Compute the distances between all pairs of corresponding template and proposal components;
(2) Construct an initial solution with the template components in which each best matches a proposal component;
(3) Substitute the worst matched component in the current solution with an alternative one that decreases the total cost defined in Eq.(4) most. Iterate this step until there is no better solution with lower cost.
Once $(i, j, \cdots)^*$ is available, a portrait paper-cut is obtained excluding hair and clothes which are handled in the post-processing step.

4. POST-PROCESSING

4.1 Hair and Clothes

The post-processing of hair and clothes starts with the detection of these two regions. Since the AAM includes the position of the face, we use a simple spatial prior to first detect the rough positions of the two regions (i.e., hair is above the face, and clothes is below the face). We then use the Graph-Cut algorithm [3] to refine the segmentation of these two regions, with automatically generated scribbles according to the rough positions.

During the creation of examples for our paper-cut template dictionary, the artists have created many hair templates. To select a hair example in addition to the face paper-cut obtained above, we select the one with the most similar contour shape to the input photograph's hair region, in terms of the shape context distance metric [2]. As for the clothes, since this region does not include many features affecting our perception of the paper-cut, we simply use a boundary-smoothed [8] version of the binary proposal obtained in the bottom-up phase as the final result, with edges added at its boundary to the background.

4.2 Connectivity

Based on directions from artists, we pre-defined a few possible curves for enforcing the connectivity (i.e., all foreground pixels are connected), which is an important characteristic of traditional paper-cut. Typical positions of the curves include: between eyebrows and facial contour, between nose and mouth, etc. Starting with all these curves turned off, we randomly turn on a few of them (but at most one for each pair of components) to achieve the connectivity.

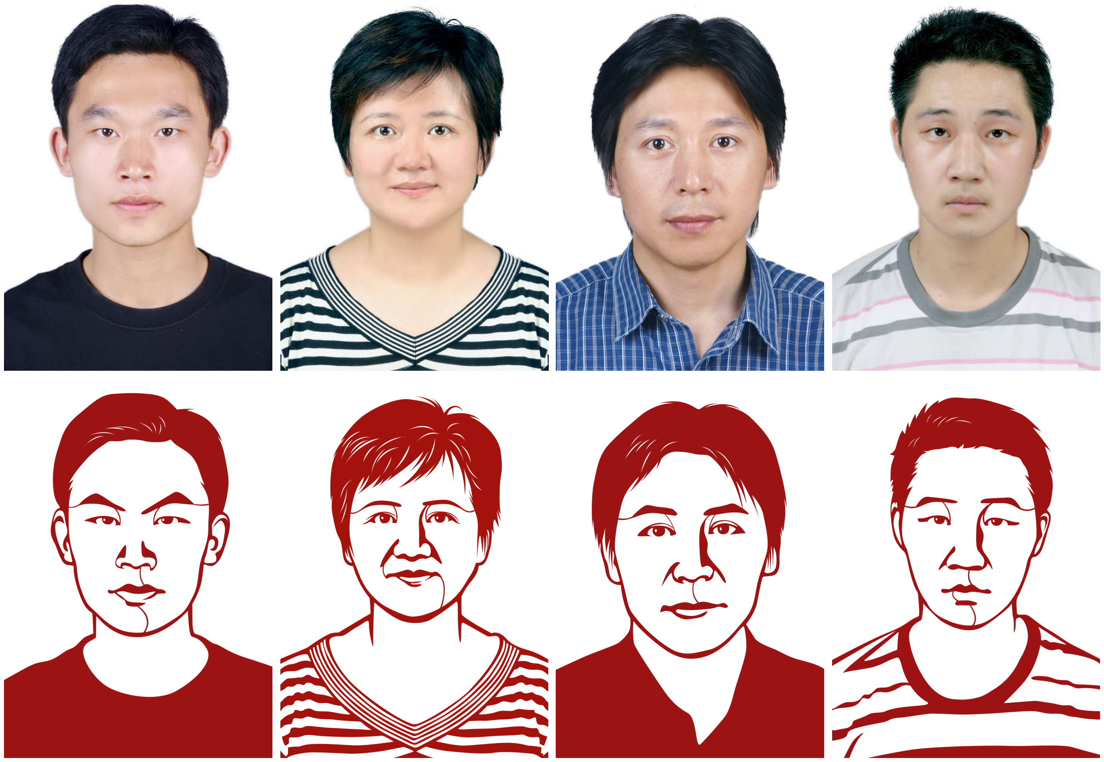

Figure 5: More results generated by our paper-cut system.

5. EXPERIMENTAL RESULTS

Figs.1 and 5 demonstrate typical rendering results of our system. We can see some features appeared in these paper-cuts are not from the original photographs (e.g., hollows in hairs, nose bridges), which are difficult to achieve with purely bottom-up methods but important for artistic effects.

6. CONCLUSION

In this paper, we have briefly introduced our method to render artistic paper-cut of human portraits from photographs. Our approach integrates bottom-up and top-down cues, with which features originated from artistic prior in addition to input images are achieved.

In future work, it is necessary to conduct evaluative human studies, to examine the abstract effects as displayed by original paper-cut artworks. Besides, to extend the current method for general photographs beyond portraits, it is necessary to enrich the template dictionary and the AOG with knowledge learned from artists and masterpieces.

7. ACKNOWLEDGMENTS

This work has been supported by three National Natural Science Foundation of China (NSFC) grants 60832004, 60875005 and 60970156, and a Chinese National 863 grant 2009AA01Z331.

8. REFERENCES

[1] I. Barrodale, D. Skea, M. Berkley, R. Kuwahara, and R. Poeckert. Warping digital images using thin plate splines. *Pattern Recog.*, 26(2):375–376, 1993.

[2] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *TPAMI*, 24(4):509–522, 2002.

[3] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *TPAMI*, 26(9):1124–1137, 2004.

[4] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *TPAMI*, 23(6):681–685, 2001.

[5] B. Gooch, E. Reinhard, and A. Gooch. Human facial illustrations: Creation and psychophysical evaluation. *ACM Trans. Graph.*, 23(1):27–44, 2004.

[6] F. Min, J. Suo, S. Zhu, and N. Sang. An automatic portrait system based on and-or graph representation. In *Proc. EMMCVPR '07*, pages 184–197, 2007.

[7] N. Otsu. A threshold selection method from gray-level histograms. *TSMC*, 9(1):62–66, 1979.

[8] Potrace. *http://potrace.sourceforge.net.*

[9] L. Wang and J. Bai. Threshold selection by clustering gray levels of boundary. *Pattern Recog.*, 24(12):1983–1999, 2003.

[10] J. Xu and C. S. Kaplan. Artistic thresholding. In *Proc. NPAR '08*, pages 39–47, 2008.

[11] J. Xu, C. S. Kaplan, and X. Mi. Computer-generated papercutting. In *Proc. PG '07*, pages 343–350, 2007.

[12] Z. Xu, H. Chen, S.-C. Zhu, and J. Luo. A hierarchical compositional model for face representation and sketching. *TPAMI*, 30(6):955–969, 2008.