Artistic Paper-cut of Human Portraits: Rendering and Perceptual Studies

MENG MENG, MINGTIAN ZHAO and SONG-CHUN ZHU, University of California, Los Angeles
and Lotus Hill Institute

The creation of art from photos aims at raising aesthetic perception of human viewers while preserving important contents in the source photos. In this process, two divergent factors are usually manipulated and balanced: likeness/fidelity and aesthetic depictions, since in many situations perfecting one of them compromises the other. Based on this two-factor model, we present a method for rendering artistic paper-cut from human portrait by extracting likeness cues and reproducing them with aesthetic practices. We segment the input photo for the regions of face, hair, and clothes, and extract information from each region: shapes and shadings for facial components, hair flow and shading for hairlines, and shadings and depths for clothes. Then we render the paper-cut by expressing these likeness cues in artistic formats with additive elements, such as hallucinated curves and decorative patterns. Using paper-cut images rendered with different levels of likeness and aesthetic factors, we have also conducted human perceptual experiments to study whether and how these factors affect visual perceptions, and how to manipulate rendering configurations accordingly for desired results. On the aspect of likeness, we observed that human subjects are able to achieve comparable face learning and recognition performances on paper-cut images and their corresponding photos, and their performances improve when more likeness cues are preserved in rendering. On the aspect of aesthetics, subjects usually favor images with moderate levels of decorative curves and patterns which do not submerge the likeness cues. These experimental results verify our likeness-aesthetics model and the rendering algorithm.

1. INTRODUCTION

Artistic paper-cut is a traditional Chinese decorative art with unique beauty of expressive abstraction in a very concise two-tone form (red foreground and white background). Among animals, flowers,

Author's addresses: M. Meng, M. Zhao, and S.-C. Zhu, Center for Vision, Cognition, Learning and Art, University of California, Los Angeles, 8125 Mathematical Sciences Building, Box 951554, Los Angeles, CA 90095-1554; emails: {joycemeng|mtzhao}@ucla.edu, sczhu@stat.ucla.edu.
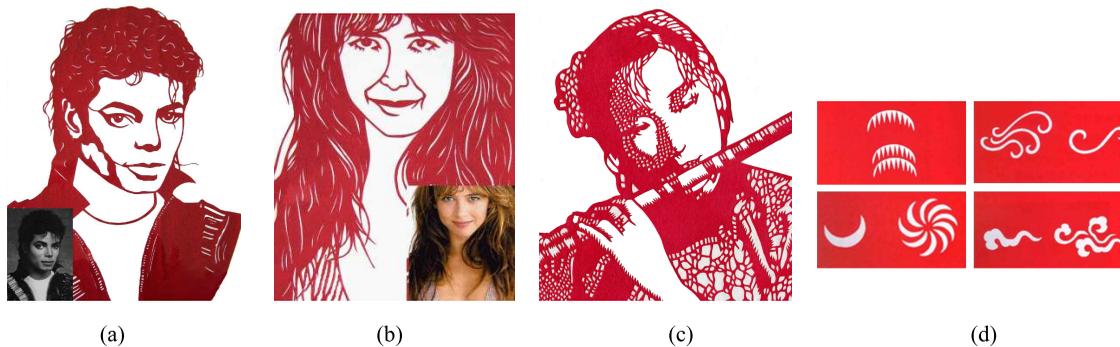
Fig. 1. (a) through (c) include artistic paper-cut of different styles created by folk artists. (d) illustrates some decorative patterns commonly used in Chinese paper-cut.

Fig. 2. Examples of our rendering results. Details on the rendering algorithms are introduced in Section 3.

auspicious patterns, etc., paper-cuts of human portraits are perhaps the most popular. Figs. 1(a) to 1(c) show some paper-cut examples created by folk artists for human portraits. In these images, we can see that a good paper-cut often utilizes characteristic decorative patterns to depict both real and hallucinated contents. For example, facial details, such as the areas around eyes and the nose, are abstracted and connected with smooth curves or delicate grids; hair is flood-filled with hollowed crescent curves; clothes are depicted with hollows of different shapes and sizes to represent shading and depths. Fig. 1(d) displays some symbolic tokens frequently used as graphical elements in Chinese paper-cut.

In general, the artistic quality of a paper-cut can be evaluated by two criteria widely used in measuring the quality of artistic renderings from portrait photos [Zhao and Zhu 2013]:

Likeness. It measures how the rendered result preserves contents from the portrait photo, including identity, shadow, hair flow, surface orientation (normal direction), sharp corners, etc.

Aesthetics. It is often associated with smooth and vigorous curves, textural patterns, cultural symbols, spatial contrasts between red and white areas, etc.

Good artworks usually have appropriate balances between these two aspects, by capturing essential likeness cues and conveying them with aesthetic practices.

The paper-cut of human portrait usually consists of three parts: face, hair, and clothes. The likeness of each part is often determined by the following factors:

Face. It has five components: eyebrows, eyes, nose, mouth, and contour, which should preserve the face identity in two aspects: 1) *Shape.* Both shapes of individual parts and the global configuration

of all parts should resemble those of the source photo. 2) *Shading*. The shading cues reflect the spatial changing of tone, and convey 3D structure of the face [Cavanagh 1991].

Hair. It is usually a flood-filled region with hollowed curves satisfying certain geometric and photometric rules. The likeness of hair is determined by two factors: 1) *Hair flow*. The orientation of each curve should agree with that of the local hair string. 2) *Shading*. The density of the curves and curve groups should follow the shading distribution of the source photo both locally and globally.

Clothes. There are two common styles of clothes paper-cut. The first one is characterized by salient decorative patterns, such as letters, icons, etc. The second style is for clothes with uniform colors. In this article, we focus on the second style, for which the most common practice is to depict the clothes with hollows of different shapes and sizes, which can convey rich information of appearance and geometry. For this style, the likeness is determined by two factors: 1) *Depth*. Depth information is critical for the perception of clothes type and other semantic meanings, such as folds and non-fold areas. The shapes of the hollows should approximate the 3D variations of the local clothes area. 2) *Shading*. In order to achieve similar tone perception, the density of patterns in clothes paper-cut should match that of the source photo.

To synthesize paper-cut with the factors described above, we treat face, hair, and clothes separately using different rendering algorithms. Fig. 2 shows some of our rendering results. Details on rendering will be described in Section 3. For each of the three parts, our algorithm can synthesize paper-cut with varying visual effects by controlling several adjustable parameters corresponding to these factors.

Human perceptual experiments are conducted to study how these factors affect the likeness and aesthetics. For paper-cut of face, we find that 1) the face recognition rate increases with the introduction of more shading into artistic rendering, 2) sometimes paper-cut images are more efficient than photos in facilitating face learning, and 3) participants favor paper-cut with moderate amount of shading. For paper-cut of hair, we find that 1) the likeness perceptions improve with increasing levels of shading, and 2) meanwhile, variances among participants become smaller. Both no shading and too much shading are inferior to moderate levels of shading in aesthetics. For paper-cut of clothes, we find that 1) participants could achieve fairly good 3D interpretations on non-fold areas, but 2) the perceptions on folds are more ambiguous and biased than those on non-fold areas. These observed differences are significant on aesthetic evaluations among 20 subjects. We will elaborate on these points with examples and statistics later in this article.

The rest of the article is organized as follows. Section 2 introduces related work on paper-cut rendering in computer graphics, and some recent literature on two-tone art perception. In Section 3, we describe our rendering algorithm for face, hair, and clothes paper-cuts. For each part, we first introduce the extracting of likeness cues from portrait photos, and then discuss the rendering of artistic paper-cut by incorporating likeness cues and aesthetic additives. In Section 4 we present three sets of perceptual experiments to evaluate the contribution of the selected algorithmic parameters (factors) in human likeness and aesthetic perception. Finally, we conclude our work with discussions in Section 5.

2. RELATED WORK

2.1 In Non-Photorealistic Rendering

From the perspective of non-photorealistic rendering, paper-cut can be viewed as an image binarization problem. The key for artistic binarization is to choose proper thresholds, so that the result satisfies some perceptual criteria. Xu et al. [2007] proposed a method by composing digital paper-cut designs. Each design is either an image rendered via a multi-layer thresholding method, or a procedurally generated shape. The designs are connected using a series of Boolean operators to ensure connectivity of components. Xu and Kaplan [2008] treated image binarization as an energy optimization problem.

They constructed a graph data structure based on the segmentation of a source image, and searched for an optimal black-white assignment of each node that minimizes the energy function. Other thresholding methods include [Otsu 1979; Wang and Bai 2003].

For artistic rendering of portraits, Zhao and Zhu [2013] presented an overview of the latest computerized artistic portrait rendering techniques. They summarized artistic rendering of portraits as an image generating process controlled by two factors, namely likeness and aesthetics, and reviewed the implementations of the two factors in the latest artistic portrait rendering methods [Chen and Zhu 2006; Xu et al. 2008; Meng et al. 2010]. Chen and Zhu [2006] conducted a pioneer work on hair representation and sketching. They proposed a generative model with three layers corresponding to sketch (strokes), vector field, and brightness. The vector field provides the direction for each pixel which is important for hair understanding and stylistic rendering. Xu et al. [2008] proposed a three layer And-Or Graph (AOG) model for face modeling and facial sketch rendering. This generative model integrated the configural relationships among facial parts, including eyebrow, eye, nose, mouth, and contour, and the hierarchical structure of facial parts and details, which are critical in artistic portrait rendering. Based on the AOG representation for face, in an earlier work, we developed an approach for rendering face paper-cut, which incorporates a bottom-up phase for likeness extraction, and a top-down phase for aesthetic rendering [Meng et al. 2010]. This article is an extension to our previous work, with improved rendering algorithms and quantitative perceptual studies.

2.2 In Psychology

Human perception of two-tone (Mooney) images has attracted research interests in psychology in the past half century. Although these impoverished images are left with fewer visual details than photos, psychological studies have demonstrated that human could achieve remarkable perception on two-tone images by means of top-down mechanism in which prior knowledge facilitates recognition [Cavanagh 1991; Moore and Cavanagh 1998]. Kemelmacher-Shlizerman et al. [2008] further verified that a single two-tone image is ambiguous in 3D reconstruction from mathematical computation, and found that the ambiguity can be resolved by exploiting prior knowledge of the structure of at least one face of a different individual, which is consistent with the previous psychological findings.

Many perceptual studies have also been done on evaluating non-photorealistic rendering effects [Gooch et al. 2004; Isenberg et al. 2006; Hertzmann 2010; Mandryk et al. 2011]. There has also been work investigating the ability of two-tone art in expressing visual information. Experiments were designed to compare human performance on two-tone art and corresponding photos or ground truth. For example, Gooch et al. [2004] proposed a method for generating illustrations from portrait photos using a bright perception model, and further evaluated the method using human face learning tasks. They demonstrated that humans could learn faces twice faster with their illustrations than with photos. Cole et al. [2009] investigated the ability of sparse line drawings to depict 3D shape. They found that people could interpret certain shapes almost as well from line drawings as from shaded images, which reveals that line drawing, as a concise two-tone art form, can effectively depict shape.

In our perceptual study, we will investigate how well participants interpret paper-cut, and perceive their aesthetic effects.

3. PAPER-CUT RENDERING

Given a portrait photo, we first use an interactive image segmentation algorithm [Boykov and Kolmogorov 2004] to obtain regions of face, hair, and clothes, and then generate paper-cut for these three parts sequentially.
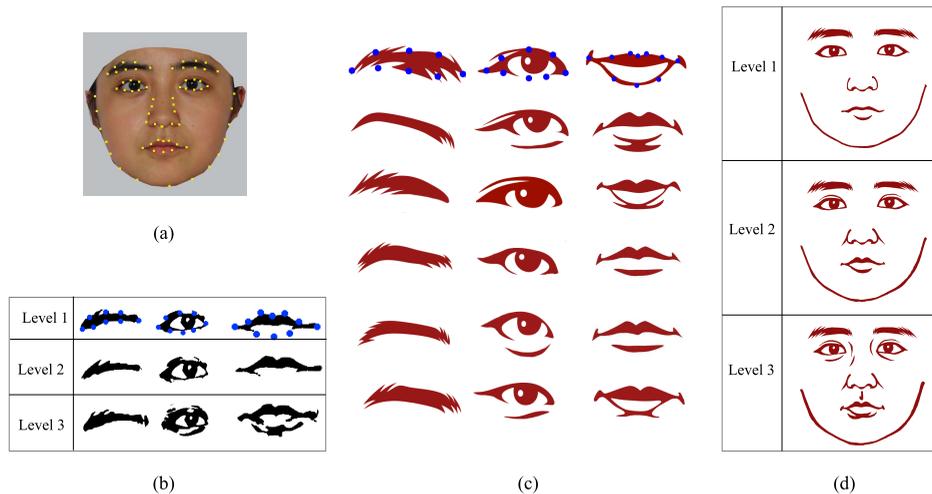
Fig. 3.   Overview of the proposed face paper-cut algorithm. (a) is the input face image with keypoints highlighted in yellow. (b) is the shading images of facial components in (a) with increasing shading from level1 to level3. (c) is the paper-cut template dictionary. Blue dots in (b) and (c) are AAM points for facial components in shading images and paper-cut templates. (d) is the rendered face paper-cut results with shading derived from the three shading images in (b), respectively.

3.1   Face Paper-cut Rendering

3.1.1   *Extracting likeness cues.* We consider two factors, namely 2D shape and shading in the likeness of human face.

2D Shape. The shape factor preserves the configurations and positions of facial parts. The shape of a face is described by positions of 70 landmark points spreading over eyebrows (16 points), eyes (16 points), nose (11 points), mouth (12 points), and outline (15 points). We adopt the Active Appearance Model (AAM) [Cootes et al. 2001] to compute the coordinates of these landmarks, as shown in Fig. 3(a).

Shading. The shading factor reflects 3D facial structures. Shading cues are acquired by applying dynamic thresholding on a portrait photo. For each pixel, its binarization threshold is computed using the method of Otsu [1979] inside its neighborhood window, and thereby different pixels may have different thresholds. By using different window sizes, we obtain shading images with different levels. Fig. 3(b) shows an example derived from Fig. 3(a) at three thresholding levels. With these shading images, we may generate paper-cut results with different shading effects.

3.1.2   *Rendering.* The workflow of our face paper-cut algorithm is shown in Fig. 3. We ask professional artists to draw two-tone images for 100 portrait photos on computers. Then we manually decompose them into facial components, as shown in Fig. 3(c). We call these paper-cuts of facial components *paper-cut templates*, which correspond to different categories of eyebrows, eyes, nose, etc.
Let

$$\Omega = \{\text{left eyebrow}, \text{right eyebrow}, \text{left eye}, \text{right eye}, \text{nose}, \text{mouth}, \text{contour}\} \tag{1}$$

be the set of facial components. For each component $\ell \in \Omega$ we have a set of templates $\Omega_\ell = \{T_i^l, i = 1, 2, ..., m_l\}$. We have collected 100 templates for each facial component. Each facial template has its associated keypoints, as shown in Fig. 3(c), which correspond to the AAM points in the original photo and shading images, as shown in Figs. 3(a) and 3(b). We organize paper-cut templates with their
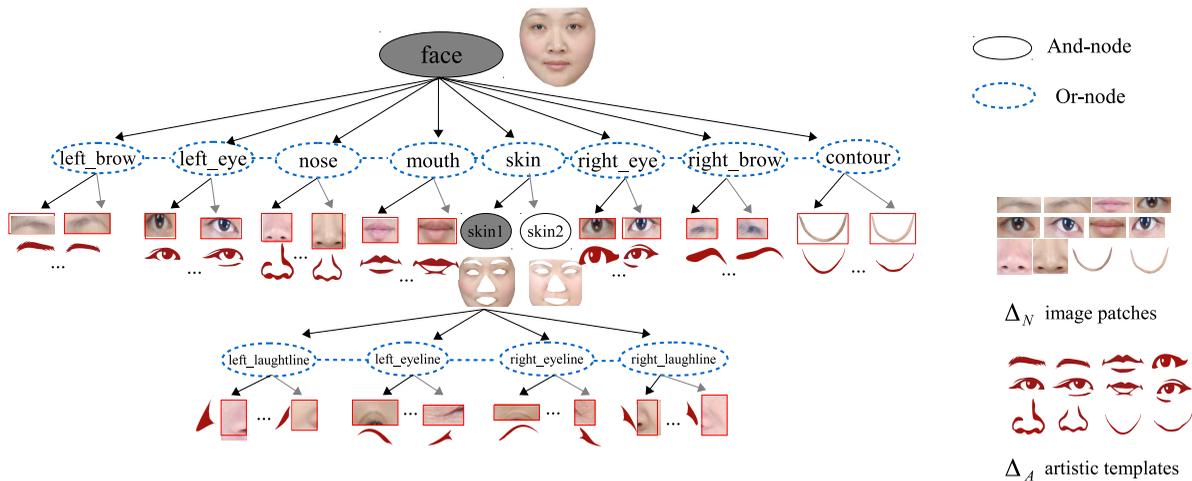
Fig. 4. A three-layer And-Or Tree (AOT) for face representation. The root node is an And-node corresponding to the generic face, which is decomposed into facial components such as eyebrows, eyes, nose, mouth. Each of them is an Or-node switching among a number of leaf nodes in the next level, corresponding to different instantial versions.

(a) parse tree for a face image

(b) parse tree for face paper-cut
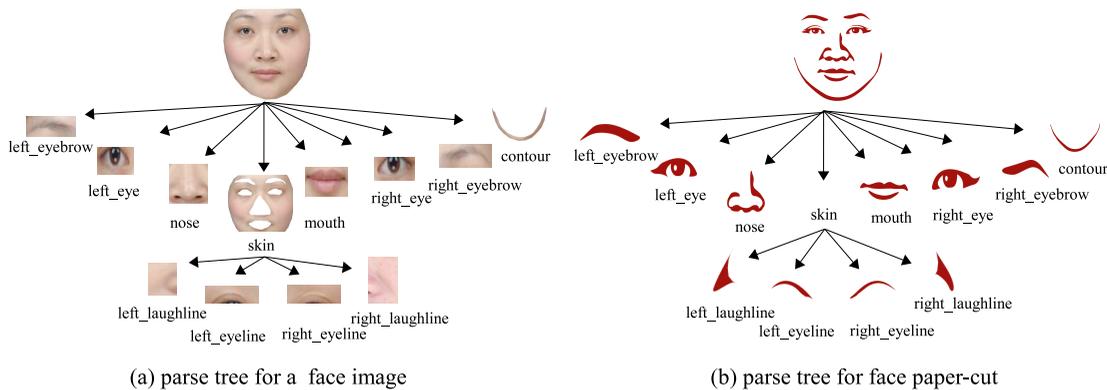
Fig. 5. Parse trees of a face image and its corresponding paper-cut version. They are instances of the AOT shown in Fig. 4.

corresponding keypoints in a structural representation named And-Or Tree (AOT) [Zhu and Mumford 2006; Xu et al. 2008]. As Fig. 4 shows, the root node represents a generic face and is decomposed into facial components, such as eyebrows, eyes, nose, and mouth. Each of them is an Or-node that can be further decomposed into different categories of the facial component. Each category represents one type of the facial component. Therefore, an And-Or Tree corresponds to a large number of possible decompositions of a generic face due to the existence of Or-nodes. By choosing one specific branch for each Or-node, we get a *parse tree* with which we are able to put together a set of parts to form an entire face. Fig. 5 displays the parse tree of a face image and its corresponding paper-cut version. With the concepts of AOT and parse tree, the problem of generating a paper-cut becomes to instantiate a parse tree from the AOT of paper-cut template dictionary that best matches the likeness cues from the original photos, which ensures that the rendered paper-cut looks like the original subject.

To achieve this, we search for the parse tree that minimizes the following function,

$$pt^* = \arg\min_{pt} \sum_{\ell \in \Omega} \left[d(B_\delta^\ell, T_{\phi(\ell),\theta(\ell)}^\ell) + \lambda\theta(\ell) \right] \qquad (2)$$

$B_\delta^\ell, \ell \in \Omega$ is the shading image of facial component ℓ with shading threshold level δ as shown in Fig. 3(b), and $T_{\phi(\ell),\theta(\ell)}^\ell$ is a paper-cut template for component ℓ warped with Thin Plate Spline (TPS) transformation [Barrodale et al. 1993]. $\phi(\ell)$ is its index in the template library, and $\theta(\ell)$ is the TPS warping parameter with $\theta(\ell) = 0$ meaning no warping. d is a difference between the shading image B_δ^ℓ and template $T_{\phi(\ell),\theta(\ell)}^\ell$. λ is a tuning parameter. Intuitively, small λ leads to better alignment of templates, while large λ makes templates less distorted.

Therefore, the computation of pt^* is determined by

$$\{\phi(\ell), \theta(\ell) : \ell \in \Omega\}, \phi(\ell) \in \{1, 2, ..., m_\ell\}, \qquad (3)$$

and we compute a set of facial templates $T_{\phi(\ell),\theta(\ell)}, \ell \in \Omega$ from the template library to form pt^* using a greedy algorithm:

(1) Compute the distances between all pairs of corresponding template and the shading image;
(2) Construct an initial solution with template components best matching the shading image individually;
(3) Substitute the worst matched component template in the current solution with an alternative one that can decrease the total cost defined in Eq.(2) most. Iterate this step until there is no better solution with lower cost.

Once pt^* is obtained, a face portrait can be rendered by putting together the warped facial components template. Fig. 3(d) shows paper-cut images generated from Fig. 3(a) with increasing levels of shading cues.

3.2 Hair Paper-cut Rendering

3.2.1 *Extracting likeness cues.* Hair style in artistic paper-cut is expressed through two factors: hair flow helps guiding the orientation of each curve, and shading reflects the contrast between light and dark regions of the original photo.

Hair flow. Hair flow is computed through the orientation field. We adopt the method of Zeng et al. [2009] used in painterly rendering [Zeng et al. 2009; Zhao and Zhu 2010; 2011] to compute the hair orientation field. The orientation field Θ of an image lattice Λ is defined as the set of orientations at all pixels

$$\Theta = \{\theta(s) : \theta(s) \in [0, \pi), s \in \Lambda\}. \qquad (4)$$

The orientation of each pixel $\theta(s)$ depends on three factors: the same pixel s in the initial orientation field (the sketch graph, shown as yellow lines in Fig. 6(b)), the adjacent pixels ∂s of s in the 4-neighborhood in the orientation filed Θ, and the same pixel s in the prior orientation field which follows a truncated Gaussian distribution. To achieve better results, we allow users to draw interactive lines on the sketch graph (see the blue and yellow lines in Fig. 6(b)), and use the modified graph as the initial orientation field. Fig. 6(c) shows the Line Integral Convolution (LIC) visualization [Cabral and Leedom 1993] of the orientation field derived from the sketch graph in Fig. 6(b).

Shading. The level of shading is controlled by the number of curves. We increase the level of shading by sampling more starting points within the hair contour, from which we trace the curves. All the starting points of curves are sampled using a poison-disk sampler [Deussen et al. 2000; Gamito and Maddock 2009] to approximate the shading properties of the original photo.
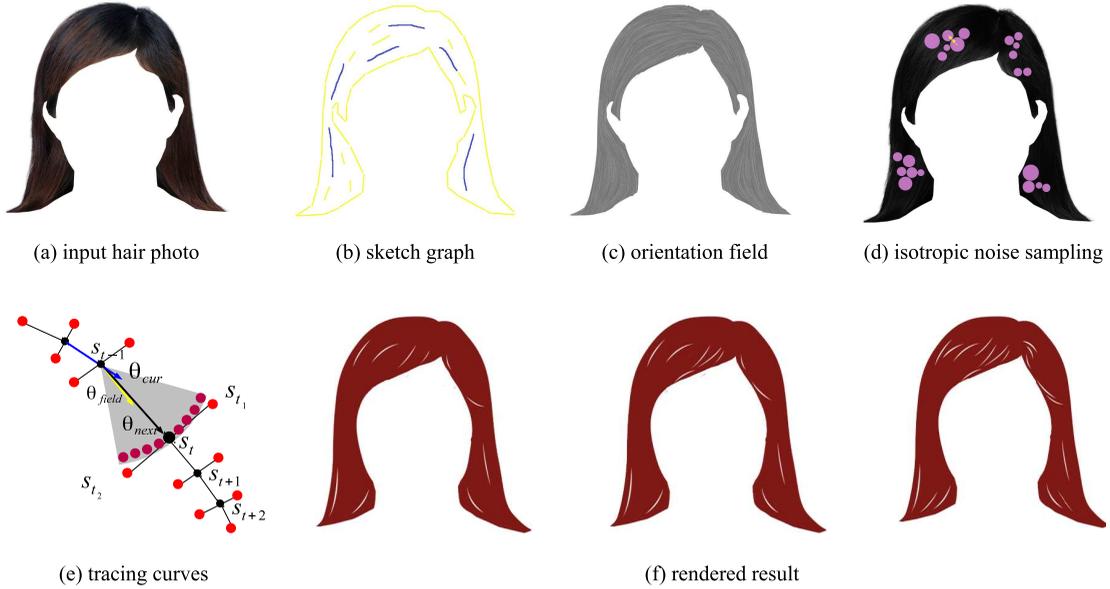
(a) input hair photo   (b) sketch graph   (c) orientation field   (d) isotropic noise sampling

(e) tracing curves   (f) rendered result

Fig. 6. The rendering process of hair paper-cut. (a) is the input hair photo. (b) is the sketch graph for rendering the orientation filed in (c). The yellow lines denote primal sketch computed using the method of Guo et al. [2007]. The blue lines are drawn by users to improve the orientation field. (c) is the LIC visualization of the orientation field. (d) shows sampled points with Poisson-disk properties. The red circle centered at each point specifies the domain of the point that cannot be occupied by other circles. (e) illustrates the sampling process of a curve. Details are described in Section 3.2. (f) shows three versions of paper-cut generated under three levels of curve numbers.

3.2.2 *Rendering.* We consider a curve as generated by a list of skeletal points following a starting point, as shown in Fig. 6(e). Once the starting points is fixed, the sampling of the following skeletal points is affected by the hair flow of each candidate point.

With the orientation field, we trace a number of curves. The position of the starting point for each curve is determined by the probability

$$p(s_0) \propto \exp(\lambda v(s_0))$$
$$\text{subject to } d(s_0, s) < v'(s_0), \forall s \in S \tag{5}$$

in which v returns the pixel intensity at s_0 after applying a 5×5 Gaussian kernel inside its neighborhood window. S is the set of sampled points. v' is the Poisson-disk radius we defined on s_0. See Fig. 6(d) for an illustration. The center of the red dot is the sampled point, and the red circle represents its domain. The radius of each circle is determined by the intensity of the center point. The poison disk property requires no pair of adjacent circles overlaps.

When the starting point is set, we sample the rest of the points of the curve sequentially. As Fig. 6(e) shows, the position of point s_t at time t is sampled within a $60°$ sector around $\theta(s_{t-1})$ in the orientation field. The probability of the candidate point s_t is given by

$$p(s_t|s_{t-1}) \propto \exp\left\{-\left[\frac{\cos^2(\theta(s_t) - \theta(s_{t-1}))}{\gamma} + \frac{\cos^2(\theta(s_t) - \theta_{\text{curve}})}{\delta} - \alpha d(s_{t-1}, s_t) - \beta\ell(s_t)\right]\right\} \tag{6}$$

in which θ_{curve} is the orientation of the segment from s_{t-2} to s_{t-1}. The first two terms penalize the angle between s_{t-1} and s_t deviating from the sampling axis and the current curve direction. γ and δ are their

(a) original portrait photo I (b) shading image S (c) paper-cut P
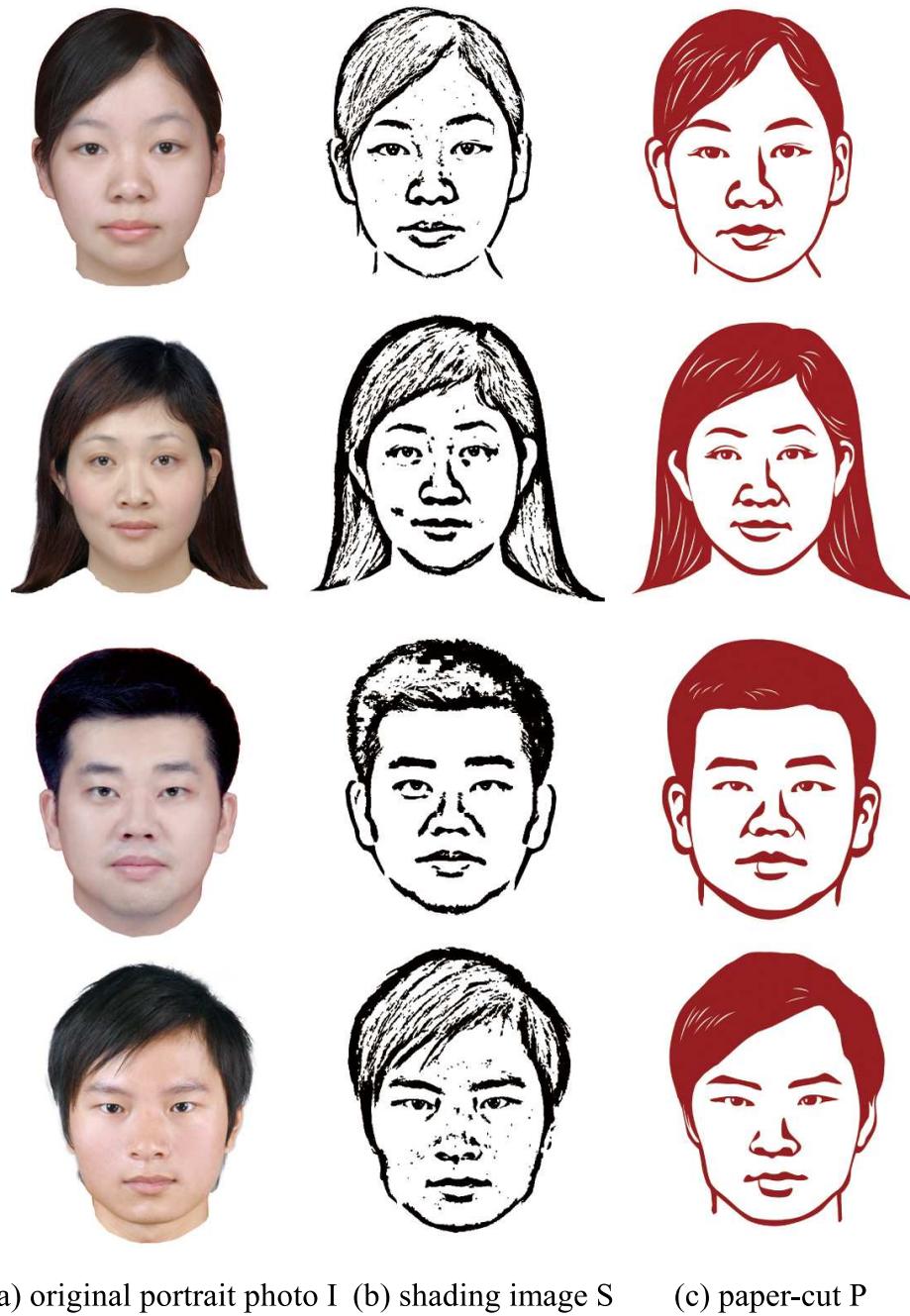
Fig. 7. Four rendering results for face and hair. Left: original photos. Middle: shading images by dynamic thresholding. Right: rendered paper-cut images.

weights. The last two terms restrain the length of the current line segment $s_{t-1}s_t$ and the total length of the curve: d returns the Euclidean distance between two points, and ℓ returns the total length from s_0 to s_t. The sampling process terminates when the probability is smaller than a predefined value for a number of trials. For each skeletal point s_t, we generate a pair of boundary points s_{t_1} and s_{t_2} on the bisector of the angle created by s_t and its two adjacent neighbors. s_{t_1} and s_{t_2} are on different sides of the curve, and their Euclidean distance is inversely proportional to the Euclidean distance between s_t and s_m, which is the median point along the curve. A curve is finally generated by combining all the boundary points along the two sides using Catmull-Rom splines [Catmull and Rom 1974]. Fig. 6(f) shows the rendered hair paper-cut with different levels of curve number. Fig. 7 shows four rendering results for face and hair, from which we can see that both shading images and paper-cuts preserve the identity information from the original face photo. However, unlike the paper-cut images, the aesthetic quality of shading images is reduced by noises and irregular contours.

3.3　Clothes Paper-cut Rendering

Shading and depth are important cues in clothes paper-cut. Shading reflects the light and dark contrast in the original photo, and depth captures the 3D surface.

3.3.1　*Extracting likeness cues*

Shading. For clothes with uniform color, we use hollows with anisotropic poison-disk properties [Li et al. 2010] to depict depth and approximate shading properties of the original photo. Light regions have more hollows of larger sizes, while for dark regions, the numbers and sizes of hollows become smaller.

Depth. Besides number and size, the orientation and aspect ratio of the two axes of an elliptical hollow are set according to the anisotropic poison-disk properties. We model each hollow as a 4-tuple $(T, \theta, P(x_0, y_0), S(a, b))$, in which T specifies its type (pure ellipse or with decorations), θ is the orientation of the major axis, $P(x_0, y_0)$ is its center, and S is its size with a and b being its transverse and conjugate diameters, respectively. P is obtained by applying anisotropic Poisson-disk sampling on the photo. We choose T in a random fashion, and computer θ and S by solving a Jacobian matrix defined on the local 4-neighborhood. During sampling, the attributes of each pixel are determined by the grayscale pixel value $V(x_0, y_0)$ and the eigenvector of a Jacobian matrix

$$g(x_0, y_0) = \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}_{(x_0, y_0)} = \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{7}$$

in which I_x and I_y are the horizontal and vertical gradients at (x_0, y_0). The second and the third term on the right side control the scaling and rotation of the hollow. The ratio a/b should satisfy $a/b = \lambda_1/\lambda_2$. To reduce noise, the photo is preprocessed using a 5×5 Gaussian blur filter. The ellipses in Fig. 8(a) illustrate the anisotropic properties of the sampled points.
We use

$$a(x, y) = \gamma V(x, y)\lambda_1(x, y)$$

$$b(x, y) = \gamma V(x, y)\lambda_2(x, y) \tag{8}$$

in which γ controls the overall size of the pattern, and simultaneously determines the number of patterns, as less patterns could be sampled if each of them occupies a larger area. V returns the grayscale value of the location.

3.3.2　*Rendering.* We apply a method similar to anisotropic dart-throwing [Li et al. 2010] to do the sampling. The samples are generated sequentially. At time t, with the point set of existing samples
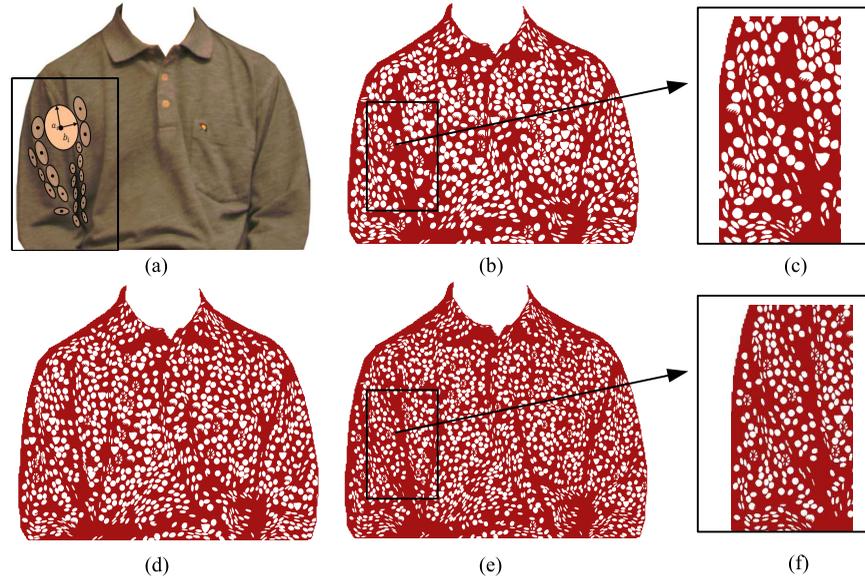
Fig. 8.   Clothes rendering results. (a) is the input clothes photo. The ellipses illustrate sampled points with anisotropic properties. (b)(d)(e) are three paper-cut images derived from (a) with different levels of tile numbers. (c) and (f) are enlarged details of (b) and (e) within the black rectangle.

$S = \{s_1, s_2, ..., s_t\}$, the position of the next sample point s_{t+1} should satisfy: (1) the elliptical domain of s_{t+1} determined by a and b should have no overlap with ellipses of all existing samples in S, and (2) a/b should fall in (1, 3]. The second point is for reducing noises and enhancing the contrast of pattern density between non-fold areas and folds. Finally, we apply affine transform with these attributes on the hollows. Figs. 8(b)–(d) show the clothes paper-cut with different levels of sampled patterns.

As for clothes images with salient structures, such as letters and ties (see Fig. 2(a) for some illustrations), we apply XDOG [Winnemöller 2011] to compute clothes paper-cut.

Now that using the methods described above, we can generate face, hair, and clothes paper-cut with different visual effects by controlling the parameters in rendering, we are interested in 1) whether our paper-cut has comparable likeness perceptions to their corresponding photos, and 2) how different levels of factors used in rendering affect likeness and aesthetic perception. To answer these questions, in the following section, we conduct three sets of psychological experiments with 20 human subjects.

4.   PERCEPTUAL STUDIES ON RENDERED PAPER-CUT

All experiments reported in this paper were conducted with a 24-inch color LCD monitor running at resolution of 1920×1200. Participants include 10 graduate students majoring in computer science, and 10 staff members from IT industry. Their ages range from 23 to 35 (mean: 27). All participants had normal or corrected-to-normal vision.

4.1   Face Paper-cut

In this set of experiments, we investigate how the shading factor influences human perception on likeness and aesthetics. On the aspect of likeness, 20 trials of face learning and recognition tasks are sequentially presented to the participants, in which we pair up original photos and our rendered images to investigate how much likeness cues are preserved in the rendered paper-cut via recognition
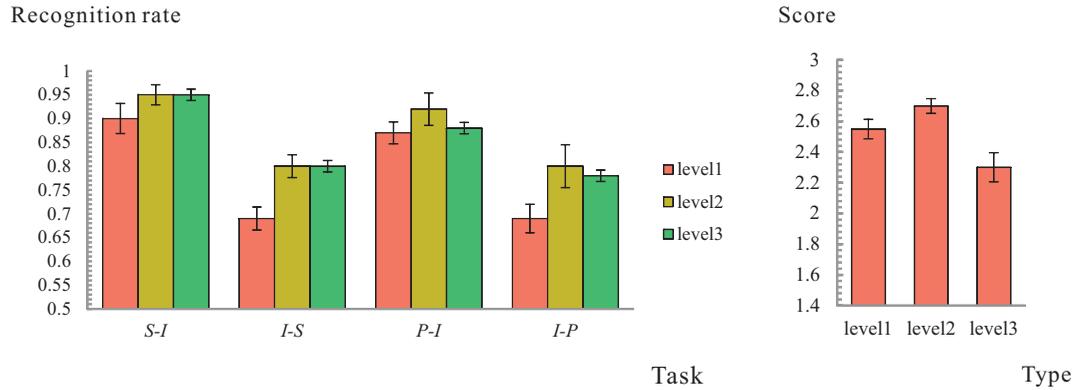
Fig. 9.   An illustration of the face likeness experiment. In this task, a noise mask is shown on the screen first for 2 seconds. Then a paper-cut image displays for 5 seconds. After that, the noise mask is shown again for 2 seconds, and followed by 6 photos. The participant is required to identify the person in the previous paper-cut from the 6 photos (e.g., in this figure, the top-center photo).

Fig. 10.   The experimental designs for the four tasks in the face likeness experiment. The four tasks have the same time sequence. They differ in the type of images displayed. The first two tasks involve photos and shading images, while the last two test photos and paper-cut images. One trial of P-I task is shown in Fig. 9.

rates. On the aesthetics aspect, participants are required to score paper-cut images derived from the same photo with different levels of shading threshold (see   10(b)), based on their aesthetic qualities. Through these experiments, we would like to investigate the following questions:

—How well do shading images and corresponding paper-cut images with different levels of shading convey likeness information?

—Is there a difference in recognition rates between tasks showing photos first and those showing rendered images first?

—Does the variation of shading affect human perception on aesthetics?

We take photos of 20 people different from the 20 experimental subjects. To eliminate the effect of the hair style factor in face recognition, we ask them to wear a shower cap in the photos. We take one photo of each of them with frontal view and neutral expression. The shower cap is later cropped off in the experiment. For each photo, we generate 3 shading images via dynamic thresholding using different window sizes, which are chosen manually to ensure that facial parts are recognizable. We also render their corresponding facial paper-cuts using our rendering algorithm described in the last section. Therefore, we have three types of images as stimuli: original photo (I), shading image (S), and paper-cut image (P), and for each I we have three shading levels Ss and three Ps.

4.1.1   *Face Likeness.* In this experiment, we pair up 20 photos with their rendered shading and paper-cut images. In each trial, the participant will either see a rendered image first and then try

Fig. 11.   Plots for the face experiments. Left: Means +/- SE for face recognition rates for the *S-I* task, *S-I* task, *P-I* task, and *I-P* task regarding shading images and paper-cut with three levels of shading cues. Right: Means +/- SE for aesthetic scores of three levels of paper-cut given by 20 participants.

to recognize the face from 6 photos (*S-I* or *P-I*), or choose the face from 6 rendered images after seeing a photo (*I-S* or *I-P*). Although the first image and the following 6 images are in different styles, participants can make the correct choice if the likeness cues from the first image match those in the later ones. Fig. 9 shows the sequence for a *P-I* trial. The experimental design for the four tasks are shown in Fig. 10.

In order to investigate the differences between learning from photos and rendered images, half of the participants perform the *I-S* and *I-P* tasks, and the rest take the *S-I* and *P-I* tasks. Each participant will finish 20 trials of learning tasks corresponding to the 20 portrait photos. It makes sure that each participant will perform the face learning task only once for one person in the testing images, so that the effect of familiarity could be eliminated. The rendered images are randomly selected.

The left plot in Fig. 11 shows recognition rates for the four face learning tasks, including 100 trials of each. Each task contains approximately 30 trials of rendered images of level 1, 30 trials of rendered images of level 2, and 30 trials of rendered images of level 3. Recall that level 3 has the highest fidelity. From the plot, we can see the following phenomena:

(1) There is an ascending trend in the recognition rates for both shading images and paper-cut from level 1 to level 2. Using ANOVA, the weak effect of image type (paper-cut or shading images) ($F_{1,398} = 0.213, p = 0.645$) demonstrates that our face paper-cut images could trigger likeness perception comparable to shading images, and therefore validates our hierarchical face model in achieving face likeness. Both the *P-I* and *I-P* trials show that the recognition rate for level 3 is lower than level 2. This might be explained by that with the increase of shading cues, more noises are also captured at the same time. At certain levels, the extra noise added into paper-cut may ruin the likeness quality. Therefore a moderate level of shading cues is better for keeping face identity.

(2) It is interesting that the mean recognition rates for *I-S* trials is significantly lower than that of the *S-I* trials ($F_{1,198} = 167.341, p \approx 0$), meaning that participants are either better at learning faces from shading images than from photos, or they are better at identifying learned faces from photos than from shading images. Similarly, the mean recognition rate for *I-P* trials is significantly lower than that of the *P-I* trials ($F_{1,198} = 89.352, p \approx 0$), which further reveals that even without redundant facial information, paper-cut and shading images are capable of facilitating the face learning/memorization task.

4.1.2 *Face Aesthetics.* The three levels of paper-cut for each of the 20 photos are used as stimuli for this experiment. In each trial, the three versions of paper-cut corresponding to the same face photo are shown on the screen at the same time. Participants are given a list of scores ranging from zero to three under each image measuring their aesthetic quality levels and required to select one score for each of the three. They are allowed to give the same score if they think the images are aesthetically equivalent. A total of 20 trials are conducted on each participant.

The right plot in Fig. 11 shows the average aesthetic scores for paper-cut with three different levels of shading. We can see that the mean aesthetic score increase from level 1 to level 2, but decreases significantly in level 3. The decreasing aesthetic score in level 3 may be due to the introduction of excessive amount of hallucinated curves, which dampens the appearance of the face and makes it look older.

In summary, from the two plots in Fig. 11, we can see that both the likeness perception and aesthetic evaluation achieve optimal perception in level 2.

4.2 Hair Paper-cut

In this set of experiments, we aim to identify how human likeness and aesthetics perceptions on hair paper-cut change as the number of hollow curves increase. We will investigate the following questions in our experiments.

—Do these curves contribute to the perception on directions of hair strings?

—Does the number of curves affect the average perceptual errors and variances of all the participants?

—How does the number of curves affect human aesthetic evaluation?

We choose three photos with different hair styles in length (short, medium, and long). For each photo, we render three versions of paper-cut with increasing number of curves. In addition, we render another flood-filled version without any curves in contrast to those with curves. In order to study how textures affect likeness perception, we also render a blurred version of each photo by employing a 5×5 Gaussian filter on every pixel. These stimuli image for the hair experiments are shown in Fig. 12.

4.2.1 *Hair Likeness.* Hair likeness is defined by surface orientation perception. To study how the number of curves affects human perception on likeness, we choose three positions for each of the three hair styles as test points and fix a gauge at each position. As shown in Fig. 12, The gauge is a 2D line segment with one end fixed at the test point and the other end rotated by the user. Participants are asked to rotate the gauge to the perceived orientation of the hair flow. The perceptual direction could be recorded once the participant press the keyboard.

In each trial, the participant need to finish the three gauge-rotation tasks for the three images with short, medium, and long hair styles. So there is a total of 9 tasks for each participant. For each image, one of the five types is randomly selected: blurred photo, flood-fill, and paper-cut of level 1, 2, or 3. This is for eliminating the effect of familiarity, since once the participant sees the blurred version of a hair photo, it is highly possible he/she will perform better on its paper-cut images later. Finally, we use the median vector (i.e., vector of median of each dimension) of all the perceptual directions at each position as the ground truth. The perceptual error of a participant is defined as

$$E = \arccos \frac{<D, Q>}{|D| \cdot |Q|} \tag{9}$$

in which D is the perceptual direction vector of the participant, and Q is the median vector of perceptual direction vectors perceived by all the 20 participants at the same position. The error defined by Eq.(9) is essentially the angle spanned by D and Q.

Fig. 12. Stimuli for the hair experiment. From top to bottom: photo and rendered images for short, medium, and long hair. From left to right: photo, blurred photo, flood-fill, and paper-cut with three levels of curve number. The white line-segment on the left-top is the gauge used for recording perceptual directions in the hair likeness experiment. The circular end is fixed on the test position. The participant could rotate the other end to interpret the hair flow at the test position.
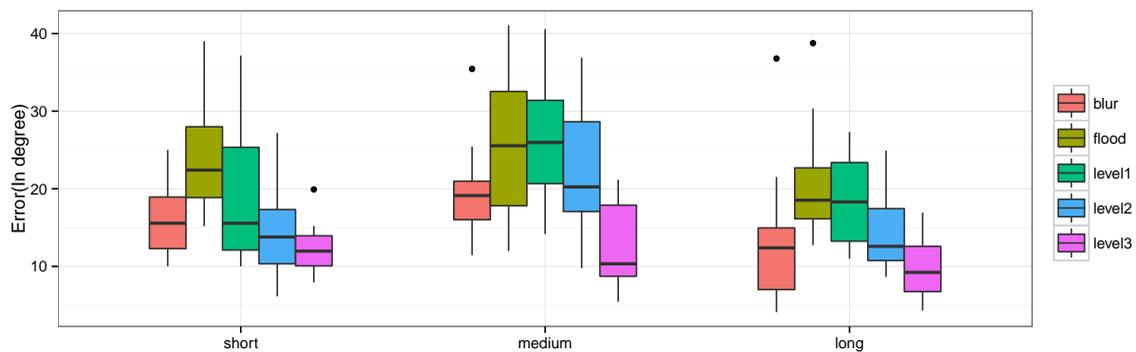

Fig. 13. Perceptual errors of 20 participants for blur photos, flood images, and paper-cut rendered under three levels of curve number.

Fig. 13 shows the box-plots of perceptual errors from ground truth on each of the five image groups derived from source photos of the three different hair styles, respectively. We can see that:

(1) In each hair image group, paper-cut with curves (levels 1, 2, and 3) have much lower perceptual errors than those with no curves (flood-fill). This confirms to our anticipation that curves contribute to the perception of hair directions. With the introduction of curves with density and orientation properties from the photo, paper-cut may convey similar likeness perceptions to that of photos.
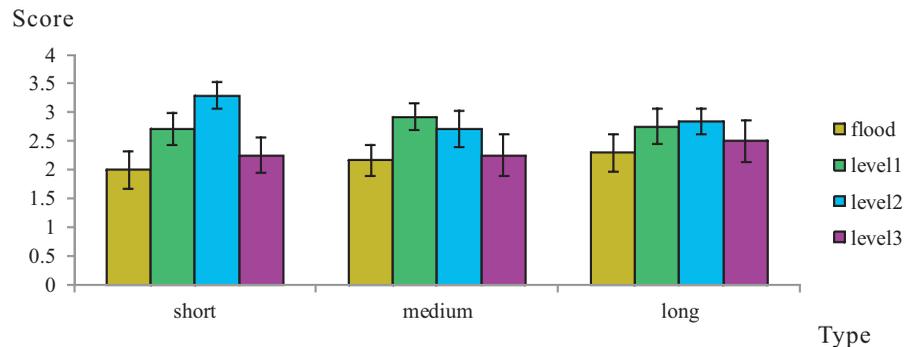
Fig. 14. Means +/- SE for aesthetic scores of flood fill and paper-cut images derived from short, medium, and long hair images with four levels of curve numbers.

(2) The perceptual errors of directions also decrease with the increase of the number of curves, which further validates our curve rendering principle. Moreover, the perceptual directions on blurred photos for all the three hair photos have greater variances than that of paper-cut at level 3. Particularly, for medium and long hair, blurred photos have outliers of more than $30°$. This indicates that with a lack of potential structural information (e.g., gradients), blurred photos are more ambiguous, while curves in paper-cut can provide more stable likeness cues in terms of shading and orientation.

4.2.2 *Hair Aesthetics.* In the hair aesthetics experiment, we use the four rendered images (flood-fill and paper-cuts of levels 1, 2, and 3) as stimuli. In each trial, the four paper-cut images appear on the screen side-by-side. Similar to the face aesthetic experiment, we ask participants to give an aesthetic score to each of them ranging from 0 to 3.

Fig. 14 shows the aesthetic scores. For each of the three hair styles, flood-fill and paper-cut with level 3 are given lower average aesthetic scores than paper-cuts of levels 1 and 2, which indicate that people prefer hair paper-cut with moderate number of curves than those with no or excessive numbers of curves.

4.3 Clothes Paper-cut

Clothes likeness is defined by the perception of depth through surface normal. This set of experiments is designed to study how the number of patterns affects human perception of likeness and aesthetics on paper-cut of the clothes. In the following likeness and aesthetic experiments, we aim to investigate:

—How well do different hollows depict 3D information in human perception in non-fold areas and folds?

—Does a larger number of hollows increase the accuracy of the 3D perception?

—How does the evaluation of aesthetic change with the increasing number of hollows?

As we described in Section 1, hollows in clothes are more complex than curves in hair. They provide various depth cues by altering the length of the two principal axes. For example, for non-fold areas, the two axes of hollows are of similar lengths, while for folds, one axis will be shrinker to approximate the rapid change of the surface normal. In our algorithm, the precision of the 3D is controlled by the sizes of hollows. As we described in Section 3, the sizes of the hollows are determined by γ, which simultaneously determines the number of hollows. As less hollows could be sampled if each of them occupies a larger area. Therefore, it is advisable to use γ to control the precision and the appearance of the paper-cut.
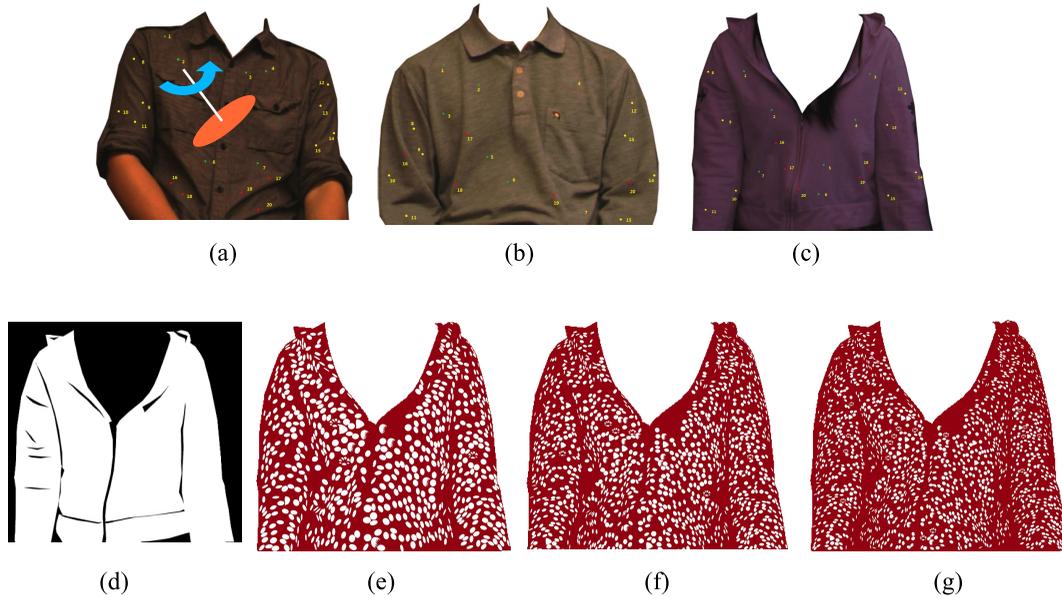
Fig. 15. Stimuli for clothes experiment. (a)-(c) are clothes photos for shirt, polo, jacket we used in this set of experiments. The colorful dots are the 20 test positions we selected for the clothes likeness experiment. The gauge on the top-left is used in the clothes likeness experiment to test human 3D perception. The intersection of the white line-segment and the orange disk is fixed in the test position. Participants could interpret the 3D perception in each test position through rotating the gauge across the surface. (d) is the fold image for (c), and (e)-(g) are three paper-cut derived from (c) with increasing number of hollow numbers.

We choose three clothes photos including a shirt, a polo, and a jacket, and render stimuli for this set of experiment. For each photo, we render three versions of paper-cut by setting three levels of γs, plus a fold image with merely folds. The photos and one set of paper-cut images and fold image are shown in Fig. 15.

4.3.1 *Clothes Likeness.* In the clothes likeness experiment, we adopt the gauge-based technique again. The gauge used in this experiment is 3D, and has been widely used in measuring surface orientations in psychological studies [Cole et al. 2009; Koenderink et al. 1992]. The gauge is a combination of a disk and an axis perpendicular to the disk's surface. This joint unit could spin in 3D with its center fixed at a specific position. For each clothes photo, we manually choose 20 positions for testing, including 15 positions in non-fold areas and 5 positions along the folds. None of the normal directions is perpendicular to the screen. Each participant is asked to perform gauge tasks on the fold image and the three paper-cut images. This is for eliminating the effect of familiarity. The image could be paper-cut rendered under level 1, 2, or 3, or with only folds. The fold images are used for testing the ability of folds in arousing 3D perception. In each trial, 20 trials of gauge tasks are sequentially submitted to a participant respective to 20 test positions. Particularly, if the participant cannot perceive the 3D in a position, they are asked to rotate the gauge perpendicular to the screen surface. The precision with which the participants place a gauge on the paper-cut can be measured by comparing the direction vector on the paper-cut and the median vector of all the participants on the corresponding photo, which is similar to Eq.(9).

Fig. 16 shows the box-plots of perceptual errors on fold images and paper-cut images with three levels of hollow size/density. We can see that:
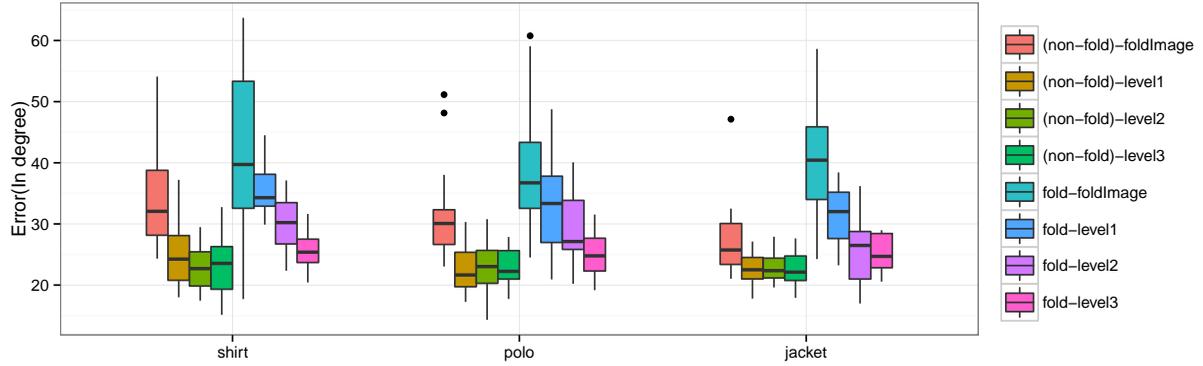
Fig. 16. 3D perceptual errors on fold images and paper-cuts with different hollow number (level1, level2, level3) at different test positions (non-fold areas, folds) for shirt, polo, and jacket.
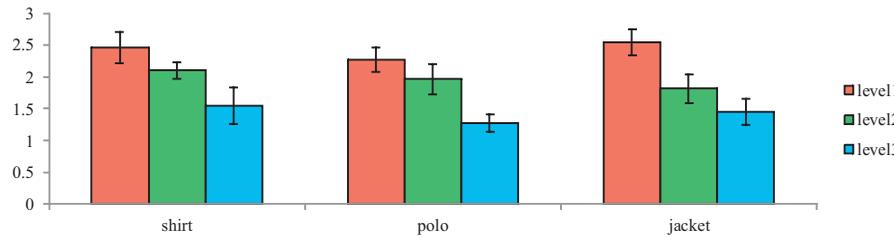

Fig. 17. Means +/- SE for aesthetic scores of clothes paper-cut of three types of clothes with increasing levels of hollow size.

(1) For non-fold areas, there is no significant effect of the number of hollows on any of the three clothes types. The medians of absolute errors for shirt are $24.5°$(level 1), $20.6°$(level 2), $23.6°$(level 3), for polo are $21.7°$(level 1), $23.0°$(level 2), $22.2°$(level 3), and for jacket are $22.4°$(level 1), $22.3°$(level 2), $22.1°$(level 3). While for fold images, the medians are $31.1°$ for shirt, $30.1°$ for polo, and $25.7°$ for jacket. This indicates that in non-fold areas, our clothes paper-cut could facilitate very good 3D perception. We compare the errors of the fold images and paper-cut images, and get $p \approx 0$. This further reveals that the anisotropic hollows are helpful for depth perception.

(2) For folds, the number of hollows is significant on shirt ($F_{2,177} = 17.0, p \approx 0$), polo ($F_{2,177} = 5.7, p = 0.004$), and jacket ($F_{2,177} = 8.9, p \approx 0$). This shows that participants' 3D perceptions improve with the increase of the number of hollows, and the interpretation of folds may need more tiles with smaller sizes. The significance of test position (non-fold areas, folds) for shirt ($F_{1,358} = 36.4, p \approx 0$), polo ($F_{1,358} = 21.3, p \approx 0$), and jacket ($F_{1,358} = 20.5, p \approx 0$) also reveals that participants can perceive the normal directions in non-fold areas much better than in the folds.

4.3.2 *Clothes Aesthetics.* The aesthetic experiment further identifies how human aesthetic perceptions change in terms of the number of hollows used in clothes paper-cut rendering. Similar to the previous aesthetic experiments, in each trial, we show the three levels of paper-cut derived from the same clothes photo to participants, and ask them to give aesthetic scores for each of them ranging from 0 to 3.

Fig. 17 shows the bar plot of aesthetic scores. Again, we analyze the data with ANOVA, and find that there is significant effect of the number of patterns on the three types of clothes. For shirt we have ($F_{2,57} = 3.71, p = 0.003$), for polo we get ($F_{2,57} = 9.37, p \approx 0$), and for jacket we get ($F_{2,57} = 6.75, p = 0.004$). This indicates that aesthetic perceptions on clothes paper-cut are affected by the number of added patterns. Most participants favor clothes paper-cut with less hollows of bigger sizes.

5. CONCLUSION

In this paper, we first present a method for rendering artistic paper-cut from human portrait photos. The visual effects can be controlled by adjusting different parameters in our algorithms. Then we designed three sets of experiments on face, hair, and clothes paper-cuts, and studied their likeness and aesthetic evaluation to guide the parameter selection. We found that human perception of likeness increases with the introduction of more likeness cues into artistic rendering, which verifies our paper-cut models and algorithms. As for aesthetics, moderate levels of decorative patterns and curves could arouse optimal aesthetic perceptions on face and hair paper-cut, and relatively smaller numbers of hollows work better for clothes paper-cut.

In addition to guide renderings, the findings reported in this paper are interesting in two other aspects. First, it proves that the balance between likeness and aesthetics is important for both aspects, and combining moderate levels of likeness cues and aesthetic patterns can produce the best results for visual perception. Second, the two-tone style of paper-cut images can carry very rich information, and by eliminating unrelated details, sometimes paper-cut can make as strong impressions as photos even in challenging recognition tasks.

The methods introduced in this article are still limited in a few aspects. First, the hollows in paper-cut are often related to geometry and appearance simultaneously, and neighboring hollows are often correlated with each other. Therefore, the rendering algorithm could be improved to cover such higher-order statistics. Second, in addition to shading, depth, orientation, etc., the semantic information in paper-cut images also heavily affects depiction. Our method has a AOT model for the face, which could be improved to capture more details and extended to cover more hair styles, clothes types, etc.

REFERENCES

BARRODALE, I., SKEA, D., BERKLEY, M., KUWAHARA, R., AND POECKERT, R. 1993. Warping digital images using thin plate splines. *Pattern Recog. 26,* 2, 375–376.

BOYKOV, Y. AND KOLMOGOROV, V. 2004. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. PAMI 26,* 9, 1124–1137.

CABRAL, B. AND LEEDOM, L. C. 1993. Imaging vector fields using line integral convolution. In *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*. SIGGRAPH '93. ACM, New York, NY, USA, 263–270.

CATMULL, E. AND ROM, R. 1974. A class of local interpolating splines. In *Computer Aided Geometric Design*, R. Barnhill and R. Riesenfeld, Eds. Academic Press, 317–326.

CAVANAGH, P. 1991. *What's up in top-down processing?* Cambridge University Press, Cambridge, UK.

CHEN, H. AND ZHU, S. C. 2006. A generative sketch model for human hair analysis and synthesis. *IEEE Trans. Pattern Anal. Mach. Intell. 28,* 7, 1025–1040.

COLE, F., SANIK, K., DECARLO, D., FINKELSTEIN, A., FUNKHOUSER, T., RUSINKIEWICZ, S., AND SINGH, M. 2009. How well do line drawings depict shape? In *ACM SIGGRAPH 2009 papers*. SIGGRAPH '09. ACM, New York, NY, USA, 28:1–28:9.

COOTES, T. F., EDWARDS, G. J., AND TAYLOR, C. J. 2001. Active appearance models. *IEEE Trans. PAMI 23,* 6, 681–685.

DEUSSEN, O., HILLER, S., VAN OVERVELD, C., AND STROTHOTTE, T. 2000. Floating points: A method for computing stipple drawings. *Computer Graphics Forum 19*, 40–51.

GAMITO, M. N. AND MADDOCK, S. C. 2009. Accurate multidimensional poisson-disk sampling. *ACM Trans. Graph. 29,* 1.

GOOCH, B., REINHARD, E., AND GOOCH, A. 2004. Human facial illustrations: Creation and psychophysical evaluation. *ACM Trans. Graph. 23,* 1, 27–44.

GUO, C., ZHU, S., AND WU, Y. 2007. Primal sketch: Integrating structure and texture. *Computer Vision and Image Understanding*, 5–19.

HERTZMANN, A. 2010. Non-photorealistic rendering and the science of art. In *Proceedings of the 8th International Symposium on Non-Photorealistic Animation and Rendering*. NPAR '10. ACM, New York, NY, USA, 147–157.

ISENBERG, T., NEUMANN, P., CARPENDALE, S., SOUSA, M. C., AND JORGE, J. A. 2006. Non-photorealistic rendering in context: an observational study. In *In Proceedings of the 4th international symposium on Non-photorealist ic animation and rendering, ACM*. 115–126.

KEMELMACHER-SHLIZERMAN, I., BASRI, R., AND NADLER, B. 2008. 3d shape reconstruction of mooney faces.

KOENDERINK, J. J., DOORN, A. J. V., AND KAPPERS, A. M. L. 1992. Surface perception in pictures. *Perception and Psychophysics*, 487–496.

LI, H., WEI, L.-Y., SANDER, P. V., AND FU, C.-W. 2010. Anisotropic blue noise sampling. *ACM Trans. Graph. 29*, 167:1–167:12.

MANDRYK, R. L., MOULD, D., AND LI, H. 2011. Evaluation of emotional response to non-photorealistic images. In *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Non-Photorealistic Animation and Rendering*. NPAR '11. ACM, New York, NY, USA, 7–16.

MENG, M., ZHAO, M., AND ZHU, S.-C. 2010. Artistic paper-cut of human portraits. In *MM '10: Proceedings of ACM Multimedia 2010 International Conference (Short Paper)*. 931–934.

MOORE, C. AND CAVANAGH, P. 1998. Recovery of 3d volume from 2-tone images of novel objects. *Cognition 67*, 45–71.

OTSU, N. 1979. A threshold selection method from gray-level histograms. *IEEE Trans. Sys., Man., Cyber. 9,* 1, 62–66.

WANG, L. AND BAI, J. 2003. Threshold selection by clustering gray levels of boundary. *Pattern Recogn. Lett. 24,* 12, 1983–1999.

WINNEMÖLLER, H. 2011. Xdog: advanced image stylization with extended difference-of-gaussians. In *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Non-Photorealistic Animation and Rendering*. NPAR '11. ACM, New York, NY, USA, 147–156.

XU, J. AND KAPLAN, C. S. 2008. Artistic thresholding. In *Proc. NPAR '08*. 39–47.

XU, J., KAPLAN, C. S., AND MI, X. 2007. Computer-generated papercutting. In *Proc. PG '07*. 343–350.

XU, Z., CHEN, H., ZHU, S.-C., AND LUO, J. 2008. A hierarchical compositional model for face representation and sketching. *IEEE Trans. PAMI 30,* 6, 955–969.

ZENG, K., ZHAO, M., XIONG, C., AND ZHU, S.-C. 2009. From image parsing to painterly rendering. *ACM Trans. Graph. 29,* 1, 2:1–2:11.

ZHAO, M. AND ZHU, S.-C. 2010. Sisley the abstract painter. In *NPAR '10: Proceedings of the 8th International Symposium on Non-Photorealistic Animation and Rendering*. ACM, New York, NY, USA, 99–107.

ZHAO, M. AND ZHU, S.-C. 2011. Portrait painting using active templates. In *NPAR '11: Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Non-Photorealistic Animation and Rendering*. ACM, New York, NY, USA, 117–124.

ZHAO, M. AND ZHU, S.-C. 2013. *Artistic Rendering of Portraits, Book chapter in Image and Video based Artistic Stylization.* Springer.

ZHU, S.-C. AND MUMFORD, D. 2006. A stochastic grammar of images. *Found. Trends. Comput. Graph. Vis. 2,* 4 (Jan.), 259–362.