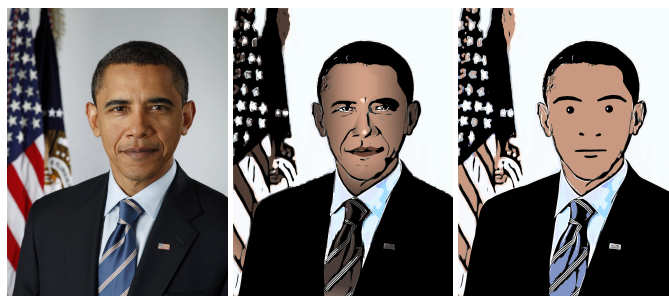# Non-Photorealistic Rendering of Portraits

Paul L. Rosin[†] and Yu-Kun Lai[‡]

Cardiff University, UK

Figure 1: *President Obama re-rendered in "puppet" style and in the style of Julian Opie.*

Abstract

We describe an image-based non-photorealistic rendering pipeline for creating portraits in two styles: The first is a somewhat "puppet" like rendering, that treats the face like a relatively uniform smooth surface, with the geometry being emphasised by shading. The second style is inspired by the artist Julian Opie, in which the human face is reduced to its essentials, i.e. homogeneous skin, thick black lines, and facial features such as eyes and the nose represented in a cartoon manner. Our method is able to automatically generate these stylisations without requiring the input images to be tightly cropped, direct frontal view, and moreover perform abstraction while maintaining the distinctiveness of the portraits (i.e. they should remain recognisable).

Categories and Subject Descriptors (according to ACM CCS): Computer Graphics [I.3.3]: Picture/Image Generation—; Computer Graphics [I.3.6]: Methodology and Techniques—.

1. Introduction

In image-based non-photorealistic rendering (NPR) images are modified to produce new stylised versions. Over the years many NPR styles have been explored such as watercolour, impressionist, oil painting, stippling, mosaic, stained glass, etc. [RC13]. Also, NPR has been used in a variety of ways, such as aiding generation of animated cartoons [BCK*13], production of films (e.g. "A Scanner Darkly", 2006) and computer games (e.g. Borderlands, 2009) [Win13], bas-relief generation [WMR*14], generating stimuli for perceptual experiments [MKHM07], scien-

tific illustration [HTER04], etc. Finally, as well as general purpose NPR techniques that are intended to be applied to a wide variety of types of images, there also exist more specialised techniques that have been designed for images restricted to a small range of objects. The most common such application is to faces (i.e. the generation of NPR portraits).

In this paper we also tackle the generation of NPR portraits, and consider two styles. The first is to create a somewhat "puppet" like rendering of a face that will be used for perceptual experiments. While the image should be stylised in such a way that it is simplified so that the face becomes less realistic, it nevertheless has to remain identifiable as the person in the original photograph. The second rendering style is to emulate Julian Opie, who is a contemporary

---

[†] e-mail: Paul.Rosin@cs.cf.ac.uk
[‡] e-mail: Yukun.Lai@cs.cf.ac.uk

British visual artist whose work includes painting, sculpture and video. In this paper we only consider one of his painting styles which consists of striking graphical portraits. Opie reduces the human face to its essentials, using flat areas of colour, thick black lines, and minimal detail; for instance the eyes are represented by just black circles. Nevertheless, the portraits remain recognisably distinct and (to a degree) recognisable, however given the sparsity of features, the expressions tend to appear fairly neutral. His work spans commercial design, and the graphical portrait style was used to good effect in 2000 for the album cover for the British pop band Blur. Another measure of his success is the large number of copycat images, and tutorials (e.g. for Photoshop) to create graphical portrait images in this style, that can be found online.

Although the two styles covered in this paper are somewhat different, both rendering pipelines share many common elements. They start by generating a basic NPR rendering for the background (i.e. non face) component of the scene. In practice it is more convenient to render the whole image, and then replace the foreground components later. Second, the face is rendered in a more abstracted style, which could be considered to be under-painting. More precisely, the skin regions (which can include the neck and ears) are re-rendered. Finally, the facial features are rendered, either in a more detailed manner for the "puppet" style, or in the extremely abstracted version used by Opie.

Our contribution is a new approach to *automatically* generating highly abstracted yet recognisable portraits. The use of two substantially different rendering styles demonstrates the potential of creating a variety of NPR styles that vary substantially in terms of levels of abstraction whilst using a similar pipeline. Unlike many existing works which assume frontal faces [RL13a, ZDD*14], or even frontal faces with uniform background [CLR*04, CVP08, GMW*09, MZZ10, TL12] and thus often require input images to be tightly cropped, our approach is more general and copes well with non-frontal (up to three quarter view) faces and cluttered backgrounds. This is achieved by combining general NPR rendering for background with a component based approach for foreground faces. Attractive portrait stylisation is obtained for a variety of input images. An example is shown in figure 1.

2. NPR Portraits

There is a wealth of literature relating to non-photorealistic rendering of portraits. Many of them use an active appearance model (AAM) or an active shape model (ASM) to localise facial features, and a dictionary of photorealistic and their corresponding stylised templates of facial features. For instance, Zhao and Zhu [ZZ11] took an example-based method to paint portraits: to create a new image the training strokes were simply warped from the training face to the new face. Wang *et al.* [WCHG13] described a more flexible extension to this approach in which from a set of training image pairs (a source image and a painting of that image) a model of style (brush stroke length, position, colour etc.) was learnt using a segmentation of the input face image provided by an ASM followed by graphcut segmentation. The ASM feature contours were also used to interpolate an orientation field over the face. The image was described by labels learnt from dense SIFT descriptors, and MRF was applied to the labels to improve spatial coherence. Finally, stroke-based rendering was performed. The work by Berger *et al.* [BSM*13] also learned models of stroke parameters from training examples. In their case seven artists provided line sketches at four abstraction levels. They used an ASM to capture the artists' deformations of the face shape with respect to the input photograph. Initial lines were extracted from the image using Kang *et al.*'s [KLC07] method, vectorised, deformed, and finally replaced by the best matching artists' strokes.

Traditional NPR stylisation was applied by Brooks [Bro07] who detected the facial area using a combination of skin filtering, AdaBoost, clustering, and graphcut. After segmentation various image filters could be applied to the regions. Meng *et al.* [MZZ10] considered the more unusual stylisation of the paper-cut, in which the result should be a single connected foreground (binary) region. They localised facial features using an active appearance model. After thresholding, these were matched to a dictionary of facial templates, with global consistency in style being enforced by a hierarchical graph model.

An interesting approach was taken by Colton *et al.* [CVP08] who attempted to stylise portraits in a manner that reflected the emotional content of the image. The appropriate emotion was determined by analysing a video clip to determine the image showing the strongest expression. The NPR pipeline used segmentation, and regions were rendered as layers using strokes simulating different media. Various aspects of the stylising were determined by the emotion, such as: **colour** – red/green were used for anger, vivid colours for happiness; **shape** – faces were stretched for disgust, eyes enlarged for fear; **media** – a sketchy pencil style was used for anger, pastel for fear.

Cartoon stylisation is increasingly popular. Most methods detect facial components in the input image, match them to the photographic database, and assemble the corresponding cartoon parts in their appropriate positions [GMW*09, RL13a, ZDD*14]. In some cases the abstraction is so strong that the individuals are not recognisable from their cartoons. Related to cartoons are caricatures, in which facial features are exaggerated, normally by geometric deformation, followed by (mostly standard) NPR rendering [TL12]. Chen *et al.* [CLR*04] described an interactive system for creating manga style portraits, which was restricted to east Asian female faces.

There has also been interest in generating physical stylised portraits, using robots to perform the rendering. In

many such works the stylisation is primitive [CEB05], but an exception is Tresset and Leymarie [TL13], the first author being an established artist. Their work is executed with a black Biro pen by "Paul the robot", and has had several public exhibitions as well as critical acclaim at an artistic level. A feature of their work is that the system incorporates a camera to monitor the artwork in progress and provides visual feedback which is used to alter shading behaviour.

As an alternative to the automated production of NPR portraits, PortraitSketch [XHLW14] is an interactive drawing system that assists users to draw sketches by adjusting their strokes in real time to improve their aesthetic quality. Facial landmarks are detected using a combination of features located by an ASM and manual annotation. These enable the positions of user strokes to be adjusted according to their distances from the facial features. Also the distances of strokes from facial features can be used to control the emphasis and de-emphasis of strokes by adjusting their width and opacity.

3. Rendering Methodology

Our system involves several steps. Given an input image, we first use a general purpose NPR rendering pipeline to generate a stylised image. This copes well with general input images with potentially cluttered background, but the faces are not particularly attractive: some details (e.g. in eyes) may be lost but could be critical, and sometimes too many details are included and the resulting image does not look sufficiently abstracted. To address this, a facial model is fitted. This provides valuable semantic information for facial components. The model however is not accurate – we refine the model using the local image information as well as skin colour information. We use a simple reflection model to produce the face shading, which gives a basis for highly abstracted faces. This is followed by adding facial components, either using details extracted from the input images with varying levels of details depending on the components (e.g. more details are required for eyes) or using simple primitives to mimic Opie's style. While existing computer vision techniques are adapted to address this problem, we note that a conservative approach is essential to be able to automatically produce good results for a variety of input images. For example, for skin detection, the existing technique is not perfect, and we prefer false positives to false negatives, as missing even a small part of the skin and failing to render it in a manner consistent with other skin regions causes substantial artefacts.

3.1. Underlying NPR Pipeline

The basis for our NPR pipeline is that described in [RL13b] and [LR14] which combines coloured regions and lines. The regions are highly simplified, and produce a stylised effect, which in itself would look rather abstract. The lines are therefore included as an overlay since they provide sparse but sufficient details to ensure that the overall effect is representational and recognisable.
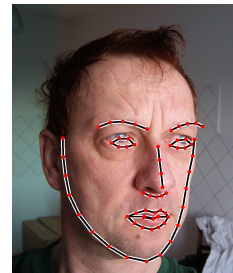
(a)

Figure 2: *Overlay of face model estimated using the HOG-CLM method [AZCP13].*

Region simplification is achieved in two ways [LR14]: by both smoothing and thresholding, i.e. within the spatial and colour domains. The input colour image is converted into HSV colour space, and each channel is blurred with a Gaussian kernel ($\sigma = 8$), which is applied in a circular fashion for the hue channel. Next, each of the HSV channels is thresholded into $C = 3$ classes; again this must be applied in a circular fashion for the hue channel. Near zero saturation values are set to zero so as to obtain a reasonable quantity of white pixels (since in HSV space white only occurs if $V = 1$ and $S = 0$, which tends to be rare for typical images). The standard Otsu algorithm is applied for the remaining $C - 1$ saturation classes. Once the pixels are allocated to HSV classes they are recoloured using the mean HSV values for each class. By increasing or decreasing the amount of blurring and the number of HSV classes the amount of abstraction can be altered, but we have found the setting above to be most effective.

To extract the lines a Laplacian of Gaussian (LoG) is applied to the intensity channel of the input image. Kang *et al.*'s approach is used which performs filtering within kernels that are locally adapted to the image structure, which enables it to produce highly coherent lines [KLC07]. Connected set morphology opening and closing operations [MW02] are applied to the lines to further enhance their coherence. The basic pipeline extracts both dark and light lines since the latter capture highlights and increase the sense of three dimensional structure.

3.2. Fitting Facial Models

Fitting a face model to an image is challenging given the large variation possible in its identity, pose, expression and illumination. Consequently, many methods proposed in the literature are not robust or accurate when applied to unseen data. We have used the state of the art multi-view HOG-CLM method by Asthana *et al.* [AZCP13] which combines the Constrained Local Model (CLM) framework with a discriminative regression framework. Their method showed good performance over a variety of databases, even those
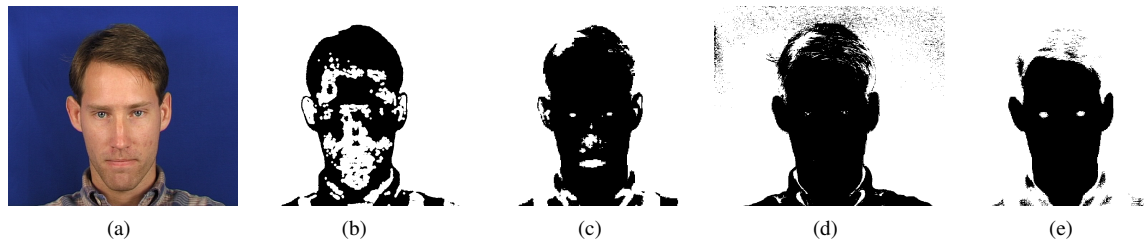
Figure 4: *Skin detection using the following methods: a) input image b) simple hue/saturation thresholding c) [HAMJ02, PKS03] d) [YLW97] e) [JR02].*
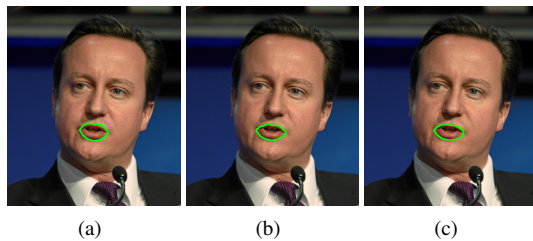
Figure 3: *Refined mouth. a) initial mouth estimated using HOG-CLM b) re-segmented using HOG-CLM initialisation followed by grabcut c) grabcut segmented projected into AAM mouth model.*

Figure 5: *Post-processing skin detection. a) initial skin detection, b) after morphological cleaning, c) after mapping colours to the dominant skin colour.*

containing images captured in uncontrolled natural settings. The CLM framework applies local models corresponding to facial parts to compute response images, from which a more holistic shape model is used to search for the best combined response. The advantage of this local approach is that there is no need to model the complex appearance of the whole face. Asthana *et al.* use HOG features from which classifiers are trained to estimate the probability of the facial parts being located at a given location in the image. The discriminative regression framework estimates the parameters for the face model from the response maps. This is done in two steps by Asthana *et al.* First a dictionary is constructed to approximate the response maps. Second, Linear Support Vector Regression (SVR) is used to learn how to update the model parameters given the dictionary representation of the response maps. The final model is comprised of three view-based CLMs, approximately covering frontal views, and left and right three quarter views. The model contains 66 landmarks covering the jaw, mouth, nose, eyes and eye brows, whose positions are estimated along with pitch, yaw and roll angles. Fitting is preceded by initialisation using a standard face detector. The authors' code is available with pre-trained facial models, and is efficient, with a run time around 2 seconds for a 0.5 megapixel image.

Figure 2 shows a typical example of fitting a face using the multi-view HOG-CLM method. It has done a reasonable

job, and coped with a three-quarter view, but inaccuracies can be clearly seen.

3.3. Refining Facial Models

Since the HOG-CLM is often inaccurate we also consider refining the facial model (currently just the outer mouth contour) using grabcut [RKB04], as did Wang *et al.* [WCHG13]. However, we find that the result is often worse as the segmentation tends to follow local, minor patterns. Therefore we built a more detailed ASM model of the mouth using a dataset that contained 28 landmarks [LBL*12] compared to the 12 landmarks in the HOG-CLM model. The grabcut region boundary was projected onto the ASM model's principal modes to force the segmentation result to conform with the mouth model – see figure 3.

3.4. Skin Detection

The second step in our NPR pipeline is to re-render the skin regions. Ideally the skin would be provided by the facial model, but even though we are using a state of the art face detector with some subsequent refinement, the results are inadequate since the facial model is incomplete, coarse and inaccurate. To correct for this we perform skin detection, although this process on its own is also unreliable. The outputs from four different skin detection algorithms are shown in figure 4, and it can be seen that all of them generate either considerable false positives or false negatives. We use the

(a) (b)
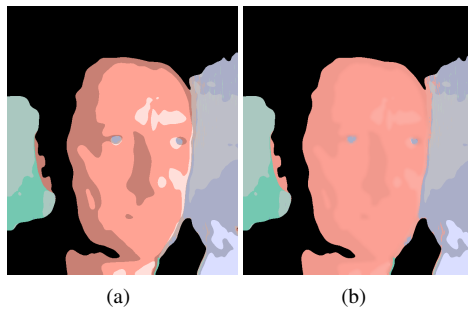
Figure 6: *Re-rendering of skin, a) coloured regions produced by the basic NPR pipeline, b) colours blended towards the dominant skin colour.*

Gaussian Mixture Model method by Jones and Rehg [JR02] (see figure 4e), as it tends to provide the most consistent detection (i.e. primarily the fewest number of false negatives, and secondly the least false positives), and combine it with further post-processing to improve the result.

First morphological opening and closing is performed which reduces the effect of noise, particularly present in the upper example in figure 5. Next the skin mask is combined along with colour information from the basic NPR version of the input image to produce the final skin mask. The initial NPR version of the image has heavily quantised the colours using thresholding to a small number of hues, saturations and intensities. The most frequent quantised colour in the pixels specified by the skin mask is selected as the main skin colour that will be used in the rendering. All other colours in the skin region are considered for mapping to the main skin colour. Mappings are only accepted if the colour difference is below a threshold and the number of pixels with the old colour in the skin region is above a threshold.[†] The updated skin mask consists of all pixels which map to the main skin colour.

The mapping is then applied to all image pixels (including those outside the skin region, but not including black and white pixels). This strategy works better than applying the mapping only to skin pixels as skin detection is not always reliable and it is much more acceptable to overestimate (with some similar non-skin pixels mapped) than underestimate (leaving some skin pixels unmapped which causes significant artefact). Rather than totally replacing the old colours, they are blended with the main skin colour. This is advantageous as retaining some aspects of the old colours increases the faithfulness of the rendering, but eliminates unaccept-

---

[†] The default threshold for the squared RGB colour difference is 30, where RGB values are in the range $[0, 64]$. The default threshold for the number of pixels is 100, where it is assumed that the input image contains approximately 0.5-1 megapixels.

(a) (b)

(c) (d)
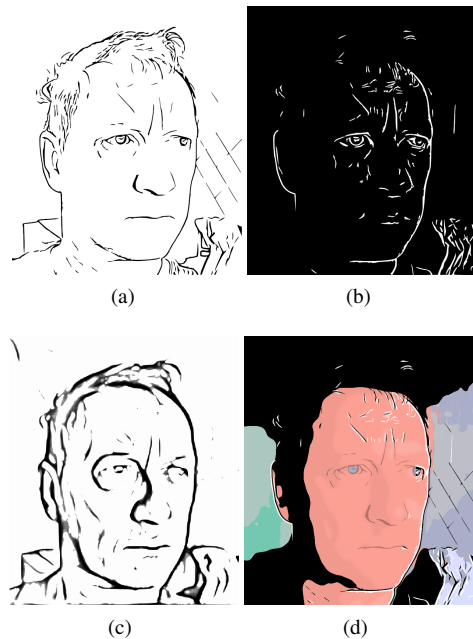
Figure 7: *Extracted and smoothed lines. (a) fine black lines, (b) fine white lines, (c) coarse black lines, (d) fine lines overlaid/blended with background image according to the face mask.*

able, jarring variations in the colour of skin tones. Further blurring is applied along the colour transition boundaries of the mapped and blended colours – see figure 6.

3.5. Rendering Smooth Lines

The lines which are generated using the LoG method described in section 3.1 are coherent in the sense that they are continuous rather than fragmented. However, they still exhibit discretisation artefacts such as "jaggies", which make them aesthetically less attractive compared to those typically produced by artists. To remedy this, we apply the Potrace [Sel03] software (using the author's code in our automatic pipeline) to the images containing the lines. Potrace is a tracing algorithm that converts bitmaps into smooth contours. It first decomposes the bitmap into a set of paths which are approximated by polygons, which in turn are represented by Bézier curves. This sequence of representations enables the algorithm to detect corners and subsequently find plausible curves that approximate the paths. Since the underlying representation is a set of Bézier curves, these can easily be rescaled whilst maintaining smoothness. For our purposes, as output it is more convenient to rasterise the curves back to an image.

For better stylisation we wish to emphasise the most salient lines, and the most straightforward way to achieve
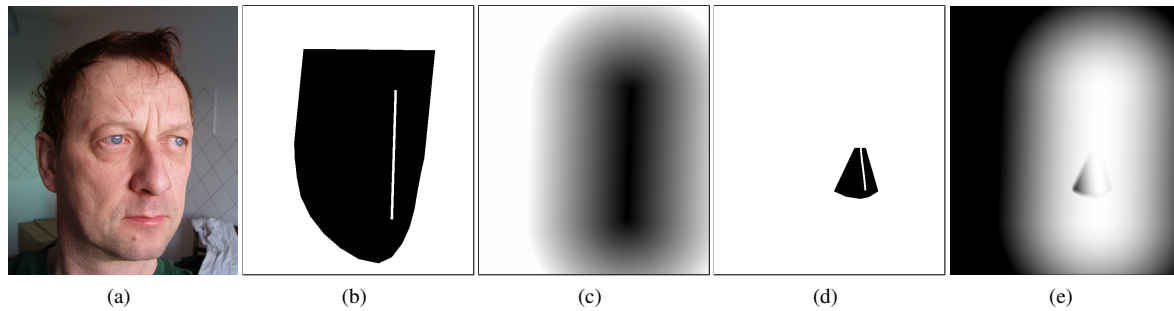
Figure 8: *Facial shading model. a) input image; b) facial region with extended nose line; c) truncated distances from extended nose line; d) nose region with nose line; e) shading factor (shaded image without applying skin colour).*

this is to use a set of low frequency lines. While these could be directly extracted by the LoG method using a large level of smoothing, we found better effects were obtained by applying the LoG to a quarter size version of the image, performing contour tracing with Potrace, with the output directly up-scaled by a factor of two to half the size of the original image, and then performing contour tracing and up-scaling again with Potrace. Since this would produce thicker contours than are desirable for our stylisation, the results of the half size tracing are thinned before the second application of tracing.

The addition of lines to the image is restricted in the following manner:

- outside the face: fine detail black and white lines extracted at full resolution are added according to our standard NPR pipeline
- inside the face: fine detail black lines are added, but blended with the skin colour to reduce their opacity
- across the whole image (both inside and outside the face): coarse black lines are overlaid

The reasoning for these rules is as follows. The generation of black and white lines using the LoG creates lines not only at intensity ridges and valleys, but also at edges [RL13b]; moreover, both black and white lines occur at either side of an edge. This is not suitable for rendering faces in the current context, as it would produce a cluttered effect, and we specifically wish the faces to be highly abstracted. Although Rosin and Lai [RL13b] suggested techniques to remove such double responses they are not totally reliable, and we prefer a more conservative approach where they are not generated in the first place. The white lines were found to be effective in suggesting highlights and three dimensional structure, but since they would be emphasising *fine detail* structure they are not used here since only the gross three dimensional structure is required.

Figure 7 shows the extracted scale lines after smoothing with Potrace, and the overlay (or blending in the case of black lines with the face) with the fine lines is presented in

figure 7d. The coarse scale black lines in figure 7c will be added at the end of the processing pipeline.

3.6. Creating a Shading Effect

The fitted face model is used to extract a mask outlining the extent of the face. Since the model is missing the upper part of the face, it is extended upwards by a fixed proportion and closed to provide an approximate mask. Shading is applied using a simplified Phong reflection model, where we focus only on the ambient and diffusion terms, with the pseudo-specular effect added later. We further simplify the problem by assuming a frontal lighting direction. Rather than use a true three dimensional model for the face surface we approximate the angle θ that a normal makes with the frontal lighting direction by computing a distance field from an extended version of the bridge of the nose, which is obtained from the CLM face model. The distances are rescaled so that they reach the value $\frac{\pi}{2}$ at the border of the face mask (i.e. these have normal directions orthogonal to the lighting direction). For points in the face mask this produces a reasonable simple surface that is approximately a cylinder with rounded ends. The pseudo-normals for all points outside of the face mask are truncated to zero. Shading is then applied to all pixels that have been detected as skin as $C = S(\cos\theta(1-\alpha)+\alpha)$, where C is the obtained colour for the pixel, S is the skin colour, $\alpha = 0.4$ is the weight for the ambient light, and $1-\alpha$ weight for the diffusion term. The shading for the nose is generated in a similar manner and replaces the face shading in the nose region, with blending at the region boundaries. The results at each step of the above process are illustrated in figure 8 and figure 9a.

3.7. Enhancing Eye Details

One of the difficulties in rendering portraits is that the human viewer is very sensitive to details (and consequently rendering errors) in faces. Not only do such errors jump out at the viewer, but for reasonably realistic images there is also the danger that they will lead to the "uncanny valley" effect, in
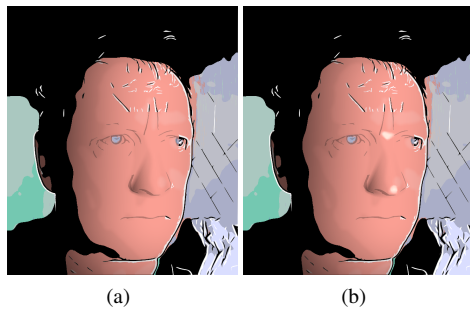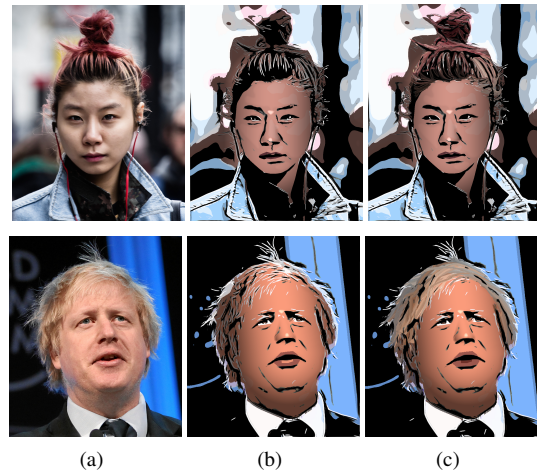
Figure 9: *Shading and highlights added to the background and skin layers.*

Figure 10: *Hair rendered using b) the basic pipeline or c) with added flowlines.*

which viewers will find the images repulsive. Of particular importance in the face are the eyes, and therefore we need to add some additional rendering to try to ensure that the eyes appear natural. The smooth lines that are added (described in section 3.5) only capture coarse detail, and are not suitable for the eyes which are relatively small features. In addition, the whites of the eyes are also relatively small features, and will possibly be lost in the process of blurring and thresholding in the underlying NPR pipeline. To overcome both these issues both black and white lines at the full image resolution (as used in [RL13b] and [LR14]) are extracted from the input image around the eyes. The landmarks for the eyes are taken from the facial model to create the eye mask. Since we have found the CLM method tends to frequently underestimate the eye size, and because it is better to be conservative and ensure that sufficient details are added even at the cost of adding extraneous detail, the eye region is enlarged before applying it to mask the black and white lines. These lines are smoothed (but not rescaled) using Potrace, and combined with the NPR rendering (as a logical AND and OR respectively).

Even with the above procedure there are some instances where the eye region has such low contrast that no white lines are detected, even at full resolution and with a low threshold. Although it would be potentially possible to use the facial model to direct the addition of the whites of the eyes, in practice it is not sufficiently accurate. Instead we add a highlight in each eye when no white lines have been detected in the central region of the eye mask. Unlike the addition of the whites of the eyes which must be applied accurately to prevent detracting from the quality of the rendering, correct localisation of the eye highlight is not critical, as the viewer is insensitive to its position.

3.8. Adding Hair

Our pipeline provides an option of applying a specialised hair rendering module. First the hair region needs to be extracted, which is not straightforward given the large number of possible variations in shape, colour, texture, size, position, etc. Therefore this step is performed semi-automatically by rendering streamlines, which are commonly used in visualization to make flow patterns visible. The only input required from the user is to assist in finding the mask region for hairs. Apart from this optional step, our method is fully automatic. A related approach based on line integral convolution was employed by Mao *et al.* [MKKI00].

We first compute the local edge flow by Kang *et al.*'s [KLC07] method. A set of particles are randomly sampled from within the hair mask; typically 500 points are chosen. These particles are advected in both forwards and backwards directions according to the edge flow until they exit from the hair mask. The particle trajectories define the initial streamlines, which are subsequently traced to provide smooth lines of a given thickness. Each streamline region acts as a mask, and for each streamline the mean colour for the masked pixels in the source image is applied to render the streamline into the NPR image – see figure 10.

3.9. Additional Components

The final steps involve adding a few more components to the rendered face to improve its appearance. Highlights are added to complement the shading effect and enhance the three dimensional nature. This is done by simply defining two templates of highlights, one at the ball of the nose and another above the nose. Using the fitted facial model parameters the templates are translated, scaled and rotated, and then blended with the NPR image – see figure 9b. The mouth is also added using the fitted mouth model, and the region recoloured using the mean HSV values from the source image.

3.10. Julian Opie Stylisation

To achieve a stylisation similar to that of Julian Opie the background is first generated as before. Again, the dominant skin tone is estimated and a colour mapping is applied. However, to obtain a high level of abstraction consistent with Opie's style, the old colours are totally replaced with the main skin colour rather than blended as was done for the puppet style.

Lines outside the face are added in the same manner as the puppet style, but no lines (both black/white and fine/coarse scales) are added inside the face, again to increase the degree of abstraction. If the facial model was more reliable and accurate then it would be used to emphasise the boundaries around the face, head and hair by drawing a uniform thickness black line around them. Since the model is not able to do this, it is therefore important to retain the extracted LoG lines which define the main overall shape of the face (i.e. the occluding boundaries and the chin) as otherwise the desired artistic effect will be lost. Therefore the face model is used to generate a more conservative face mask that more tightly bounds the inner facial area, and it is this mask that is applied to select the LoG lines.

There are four sets of facial features that need to be added to the rendered skin: eyes, eyebrows, nostrils and mouth. The landmarks fitted by the face model are used directly to generate the eyebrows which are pasted into the image. Both eyes are defined by one template, and the fitted facial model parameters determine its appropriate translation and scale. Nostrils are also defined as templates; three versions are used in which one of the nostrils is compressed horizontally by factors $\{1, 0.75, 0.5\}$. These are used to provide suitable frontal and two increasingly side-on views. Depending on the yaw angle estimated by the fitted facial model the appropriate template is selected

$$\text{factor} = \begin{cases} 1 & \text{yaw} = [-5°, 5°] \\ 0.75 & \text{yaw} = [-10°, -5°] \text{ or } [5°, 10°] \\ 0.5 & \text{otherwise.} \end{cases}$$

and applied after reflection if necessary. The minimal rendering of the mouth by two lines favoured by Julian Opie was recreated by extracting the middle, lower section of the outer mouth boundary as one curve, and the main line of the mouth consists of the corners of the mouth connected to the upper section of the inner mouth. Both the mouth and eyebrows are rendered using Potrace to ensure that the lines are smooth.

4. Experimental Results and Discussions

Figures 11-13 show a gallery of results. The source images span a variety of types, from close cropped heads to more distant shots, some containing background clutter, and a variety of poses. Note also the variety of skin colours, due to both ethnicity and lighting conditions, which poses a considerable challenge. The "puppet" style renderings demonstrate that the results are both recognisable yet simplified. The added highlights simulate the effect of specularities, which is consistent with our original aim of mimicking plastic puppets. The Julian Opie results are recognisable to a limited degree, however, this is not unexpected due to the extreme abstraction and also applies to the original work of Opie's.

Figure 14 shows the first six faces from the first disc of the XM2VTS multi-modal face database [MMK*99]. The results are generally successful, although there are some artefacts in the Julian Opie stylisation due to glasses and facial hair. Some more extreme cases of problematic input images are shown in figure 16.

Figure 15 shows that our method works successfully on images with more complicated backgrounds. We have shown results containing faces in frontal and three quarter views and with clear or cluttered background. While the pipeline is general, we are currently unable to process side on views mainly because the HOG-CLM face model we use does not include such poses. We will exploit more general face poses in future.

Programs were run on a 3.40 GHz Intel Core i7. For 0.5 megapixel images the run time of our basic method was about 30 seconds using our unoptimised code.

Our method is fully automatic. The majority of parameters (e.g. for morphological opening and closing, thresholding, etc) are preset, and fixed for all the images presented in this paper. The only parameters that were adjusted for a few of the images were the two parameters in the skin detection method described in section 3.4, which was improved by manually adjusting the thresholds for the colour difference and region size.

5. Limitations and Future Work

It can be seen from the gallery that the skin hue mapping is sometimes over-applied (e.g. figure 12c&f), so that the skin colour bleeds into the rest of the image. However, it is necessary to ensure that all relevant skin pixels are re-mapped, even at the cost of some false positives, as otherwise skin not covered by the fitted face model, such as forehead, ears, neck, etc. will often not appear consistent with the inner face, which is aesthetically unacceptable. Future work will consider alternative constraints to reduce skin bleed without compromising the final appearance.

It is not a problem for our puppet style rendering if the subject in the input images is wearing glasses. However, for the Julian Opie style our current system cannot cope – see figure 16. The difficulty is that ideally the glasses should be detected, localised, and then removed by performing in-painting in order that the image can then be appropriately re-rendered. While this is possible, it is a relatively complex process [WLS*04], and not part of the current pipeline, but will be considered in the future. An additional challenge is to
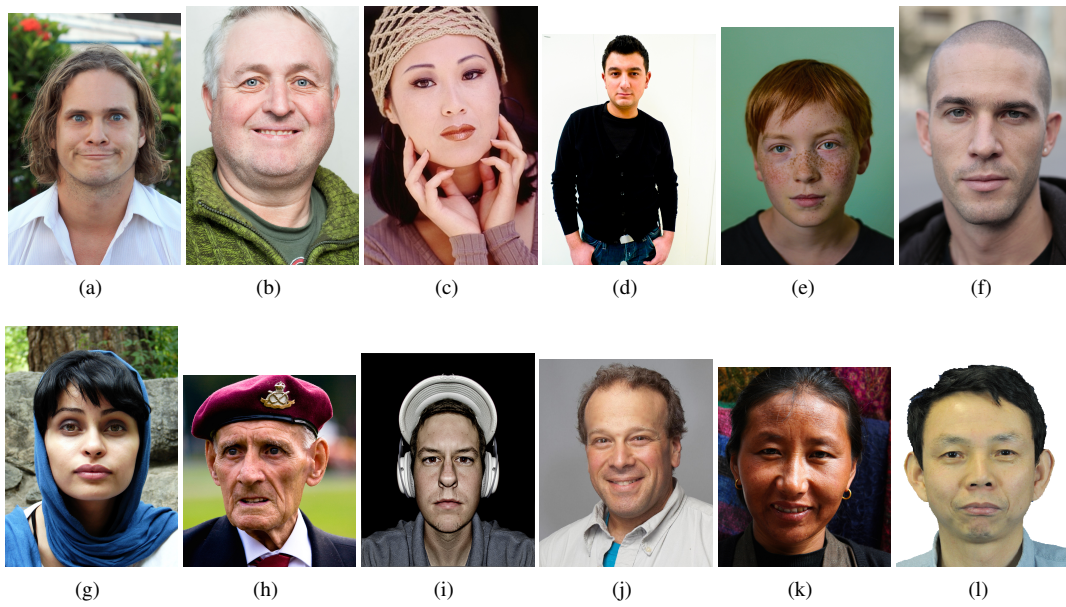
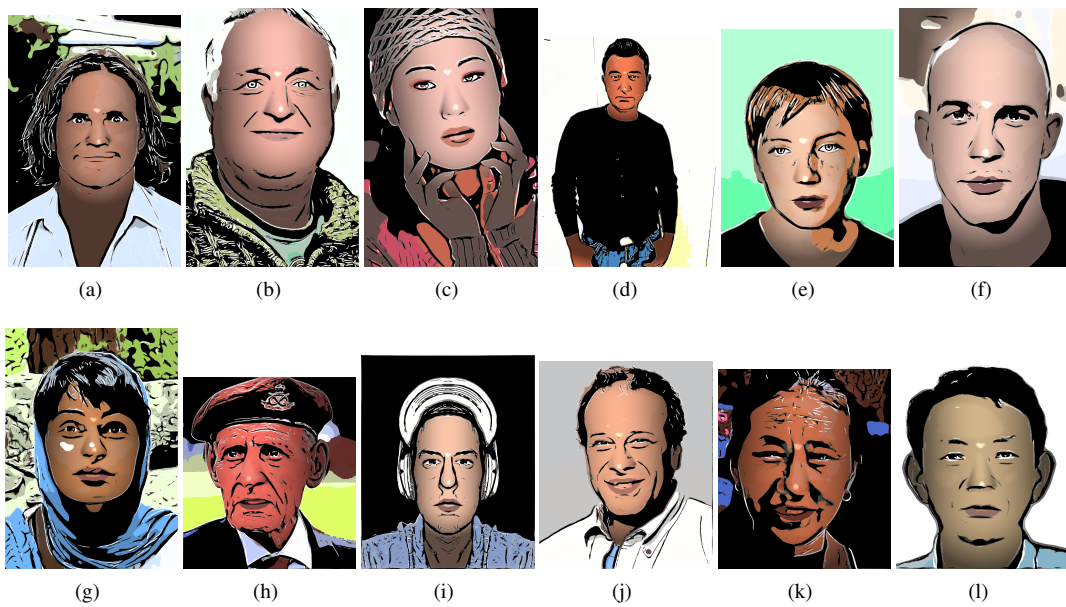**Figure 11:** *Gallery: source images*

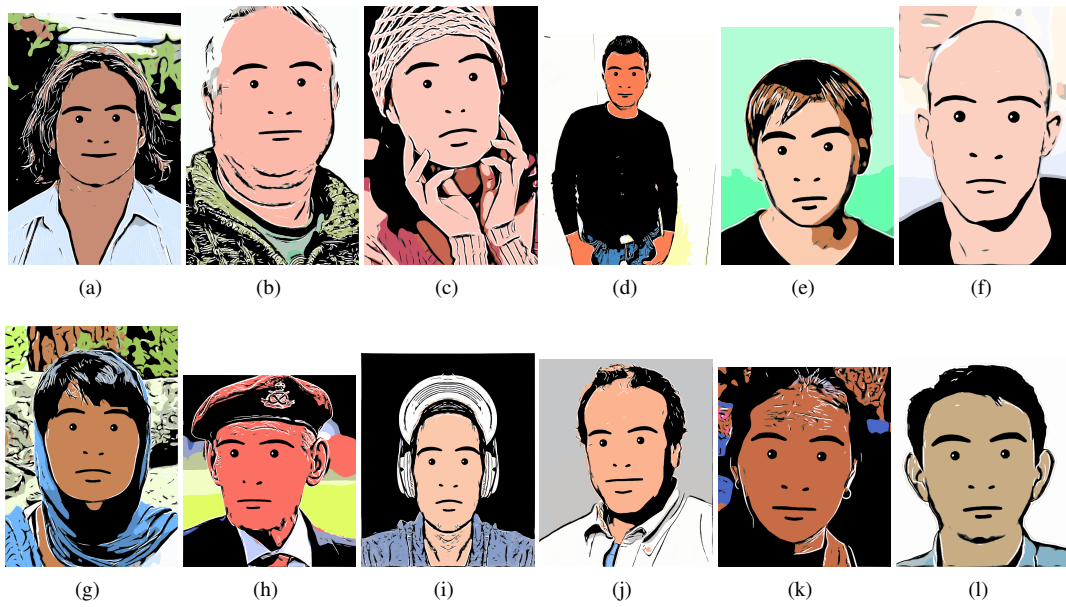**Figure 12:** *Gallery: "puppet" style rendering*

Figure 13: *Gallery: Julian Opie style rendering*

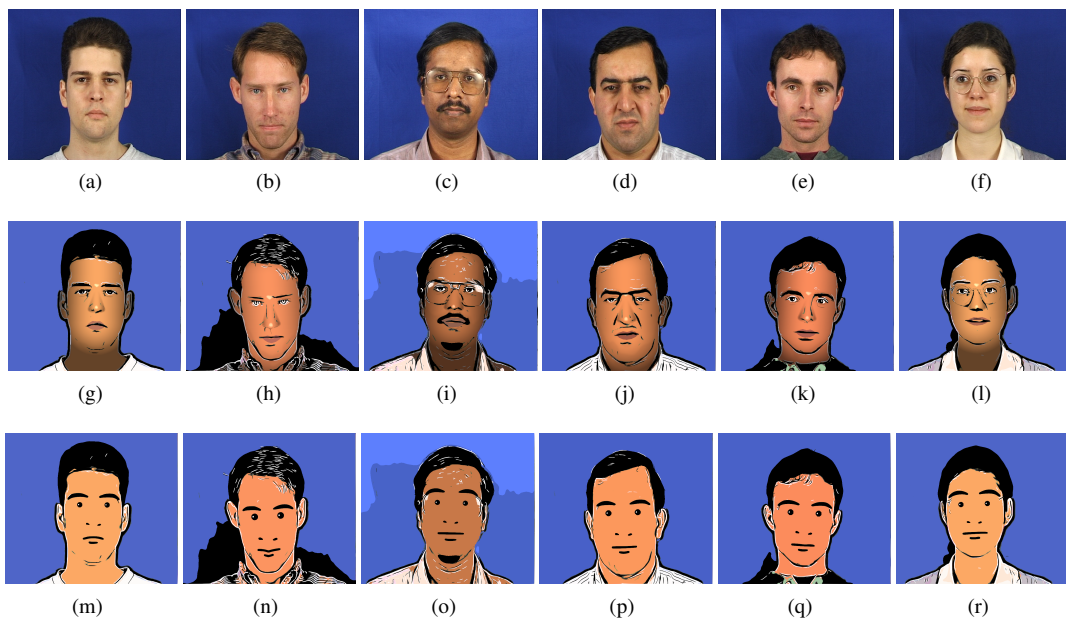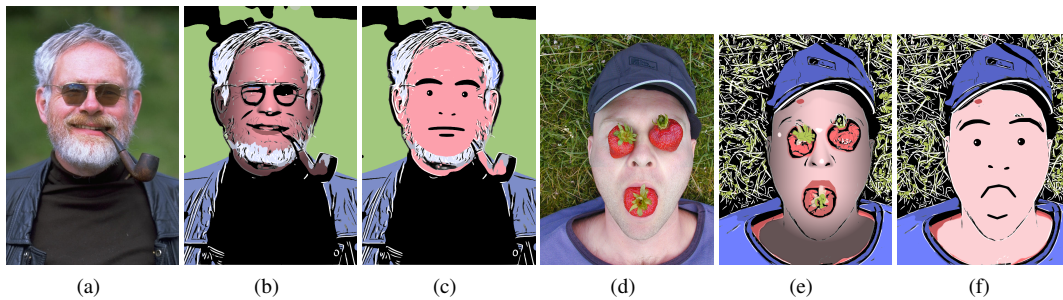**Figure 14:** *Images from the XM2VTS face database rendered in "puppet" style and Julian Opie styles.*

(a)   (b)   (c)

(d)   (e)   (f)

Figure 15: *Results of rendering images showing a more complicated background.*

(a)   (b)   (c)   (d)   (e)   (f)

Figure 16: *Problematic images for our NPR pipeline due to: facial hair, glasses and other objects obscuring the face.*

be able to consistently cope with facial hair. This will require a module to detect moustaches, beards and stubble, and then render them in a specialised way.

For eye rendering, the current Opie implementation simply locates the template at the centre of the eye region returned by the face model. Future work will detect the iris/pupil so that the template can be placed more accurately when the subject was not looking straight on.

The current pipeline assumes that the portrait images contain a single face. Future work will generalise this to cope with input images containing multiple faces.

Finally, even with the refinement of the mouth using grabcut and reprojection into the ASM model, there is scope for further improvements, which would particularly benefit the expressiveness of the Julian Opie stylisation. Also in addition to mouths it is desirable to refine other features such as eyebrows.

Acknowledgements

References

[AZCP13] ASTHANA A., ZAFEIRIOU S., CHENG S., PANTIC M.: Robust discriminative response map fitting with constrained

local models. In *Conf. Computer Vision Pattern Recognition* (2013), pp. 3444–3451. 3

[BCK*13] BÉNARD P., COLE F., KASS M., MORDATCH I., HEGARTY J., SENN M. S., FLEISCHER K., PESARE D., BREE-DEN K.: Stylizing animation by example. *ACM Trans. Graph. 32*, 4 (2013), 119:1–119:12. 1

[Bro07] BROOKS S.: Mixed media painting and portraiture. *IEEE Trans. Vis. Comput. Graph 13*, 5 (2007), 1041–1054. 2

[BSM*13] BERGER I., SHAMIR A., MAHLER M., CARTER E., HODGINS J.: Style and abstraction in portrait sketching. *ACM Trans. Graph. 32*, 4 (2013), 55:1–55:12. 2

[CEB05] CALINON S., EPINEY J., BILLARD A.: A humanoid robot drawing human portraits. In *Int. Conf. on Humanoid Robots (Humanoids)* (2005), pp. 161–166. 3

[CLR*04] CHEN H., LIU Z., ROSE C., XU Y., SHUM H., SALESIN D.: Example-based composite sketching of human portraits. In *Int. Symp. on Non-Photorealistic Animation and Rendering* (2004), pp. 95–153. 2

[CVP08] COLTON S., VALSTAR M. F., PANTIC M.: Emotionally aware automated portrait painting. In *Int. Conf. on Digital Interactive Media in Entertainment and Arts* (2008), pp. 304–311. 2

[GMW*09] GAO W., MO R., WEI L., ZHU Y., PENG Z., ZHANG Y.: Template-based portrait caricature generation with facial components analysis. In *Int. Conf. on Intelligent Computing and Intelligent Systems* (2009), pp. 219–223. 2

[HAMJ02] HSU R., ABDEL-MOTTALEB M., JAIN A.: Face detection in color images. *IEEE Trans. on Pattern Analysis and Machine Intelligence 24*, 5 (2002), 696–706. 4

[HTER04] HEALEY C. G., TATEOSIAN L., ENNS J. T., REMPLE M.: Perceptually based brush strokes for nonphotorealistic visualization. *ACM Trans. Graph. 23*, 1 (2004), 64–96. 1

[JR02] JONES M. J., REHG J. M.: Statistical color models with application to skin detection. *Int. J. Comput. Vision 46*, 1 (2002), 81–96. 4, 5

[KLC07] KANG H., LEE S., CHUI C.: Coherent line drawing. In *ACM Symp. Non-photorealistic Animation and Rendering* (2007), pp. 43–50. 2, 3, 7

[LBL*12] LE V., BRANDT J., LIN Z., BOURDEV L., HUANG T. S.: Interactive facial feature localization. In *Europ. Conf. Computer Vision*. 2012, pp. 679–692. 4

[LR14] LAI Y.-K., ROSIN P. L.: Efficient circular thresholding. *IEEE Trans. Image Processing 23*, 3 (2014), 992–1001. 3, 7

[MKHM07] MAR R. A., KELLEY W. M., HEATHERTON T. F., MACRAE N. C.: Detecting agency from the biological motion of veridical vs animated agents. *Soc. Cogn. Affect. Neurosci. 2*, 3 (2007), 199–205. 1

[MKKI00] MAO X., KIKUKAWA M., KASHIO K., IMAMIYA A.: Automatic generation of hair texture with line integral convolution. In *Int. Conf. on Information Visualization* (2000), pp. 303–308. 7

[MMK*99] MESSER K., MATAS J., KITTLER J., LUETTIN J., MAITRE G.: XM2VTSDB: the extended of M2VTS database. In *Proc. Audio- and Video-Based Person Authentication* (1999), pp. 72–77. 8

[MW02] MEIJSTER A., WILKINSON M.: A comparison of algorithms for connected set openings and closings. *IEEE Trans. PAMI 24*, 4 (2002), 484–494. 3

[MZZ10] MENG M., ZHAO M., ZHU S. C.: Artistic paper-cut of human portraits. In *ACM Multimedia* (2010), pp. 931–934. 2

[PKS03] PEER P., KOVAC J., SOLINA F.: Human skin colour clustering for face detection. In *Int. Conf. on Computer as a Tool* (2003), pp. 144–148. 4

[RC13] ROSIN P. L., COLLOMOSSE J.: *Image and Video-based Artistic Stylisation*. Springer, 2013. 1

[RKB04] ROTHER C., KOLMOGOROV V., BLAKE A.: GrabCut: Interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph. 23*, 3 (2004), 309–314. 4

[RL13a] RHEE C., LEE C.: Cartoon-like avatar generation using facial component matching. *Int. J. of Multimedia and Ubiquitous Engineering 8*, 4 (2013), 69–78. 2

[RL13b] ROSIN P. L., LAI Y.-K.: Artistic minimal rendering with lines and blocks. *Graphical Models 75*, 4 (2013), 208–229. 3, 6, 7

[Sel03] SELINGER P.: *Potrace: a polygon-based tracing algorithm*, 2003. URL: http://potrace.sourceforge.net/. 5

[TL12] TSENG C.-C., LIEN J.-J. J.: Colored exaggerative caricature creation using inter- and intra-correlations of feature shapes and positions. *Image Vision Comput. 30*, 1 (2012), 15–25. 2

[TL13] TRESSET P. A., LEYMARIE F. F.: Portrait drawing by Paul the robot. *Computers & Graphics 37*, 5 (2013), 348–363. 3

[WCHG13] WANG T., COLLOMOSSE J. P., HUNTER A., GREIG D.: Learnable stroke models for example-based portrait painting. In *British Machine Vision Conference* (2013), pp. 36.1–36.11. 2, 4

[Win13] WINNEMÖLLER H.: NPR in the wild. In *Image and Video-Based Artistic Stylisation*, Rosin P. L., Collomosse J. P., (Eds.). Springer, 2013, pp. 353–374. 1

[WLS*04] WU C., LIU C., SHUM H.-Y., XU Y.-Q., ZHANG Z.: Automatic eyeglasses removal from face images. *IEEE Trans. Pattern Anal. Mach. Intell. 26*, 3 (2004), 322–336. 8

[WMR*14] WU J., MARTIN R., ROSIN P. L., SUN X., LAI Y.-K., LIU Y., WALLRAVEN C.: Use of non-photorealistic rendering and photometric stereo in making bas-reliefs from photographs. *Graphical Models 76*, 4 (2014), 202–213. 1

[XHLW14] XIE J., HERTZMANN A., LI W., WINNEMÖLLER H.: PortraitSketch: Face sketching assistance for novices. In *Symp. on User Interface Software and Technology* (2014), pp. 407–417. 3

[YLW97] YANG J., LU W., WAIBEL A.: Skin-color modeling and adaptation. In *Asian Conf. on Computer Vision* (1997), pp. 687–694. 4

[ZDD*14] ZHANG Y., DONG W., DEUSSEN O., HUANG F., LI K., HU B.-G.: Data-driven face cartoon stylization. In *SIGGRAPH Asia Technical Briefs* (2014), pp. 14:1–14:4. 2

[ZZ11] ZHAO M., ZHU S.-C.: Portrait painting using active templates. In *Symp. on Non-Photorealistic Animation and Rendering* (2011), pp. 117–124. 2