ORIGINAL PAPER

Non-photorealistic neural sketching

A case study on frontal-face images

**Francisco de Assis Pereira Vasconcelos de Arruda ·
José Eustáquio Rangel de Queiroz ·
Herman Martins Gomes**

Abstract We present and evaluate a neural network-based technique to automatically enable NPR renderings from digital face images, which resemble semi-detailed sketches. The technique has been experimentally evaluated and compared with traditional approaches to edge detection (Canny and Difference of Gaussians, or DoG) and with a more recent variant, specifically designed for stylization purposes (Flow Difference of Gaussians, or FDoG). An objective evaluation showed, after an ANOVA analysis and a Tukey t-test, that the proposed approach was equivalent to the FDoG technique and superior to the DoG. A subjective experiment involving the opinion of human observers proved to be complementary to the objective analysis.

Keywords Non-photorealistic rendering · Edge detection · Image sketching · Statistical analysis · Neural networks

1 Introduction

Non-photorealistic rendering (NPR) or *stylized rendering* is an important growing area of Computer Graphics and Image Processing. It is mainly concerned with the generation of images or videos that can have an expressive appeal, containing some visual and emotional characteristics, e.g., pencil, pen-and-ink, charcoal and watercolor drawing. Moreover, NPR techniques usually are meant to provide expressive, flexible visual representations, which are characterized by the use of randomness and arbitrary interpretation of features being drawn, rather than total adherence to realistic properties [2].

Within the above context, the goal of this paper is to present a neural network-based technique to automatically enable NPR renderings from digital images, which resembles making semi-detailed sketches. It has experimentally been evaluated in the context of human frontal-face images. Human faces have been subject of extensive study in diverse fields, such as digital image processing (DIP), computer vision (CV), pattern recognition (PR), and computer graphics (CG), for a wide variety of tasks ranging from detection [71], recognition [24], tracking [12], and animation [63], to NPR—portrait and sketch [11].

The following aspects of the sketch generation have been addressed in this work: (i) *pre-processing* (color processing and blurring) of face images; (ii) a *neural network approach* for edge detection, focusing on sketch generation; (iii) *post-processing* for improving final results; (iv) *experimental validation* of produced renderings by objective measures (PSNR, FoM, SSIM); and (v) *validation* of the results by a subjective evaluation, based on a human voting scheme.

The following section gives an overview on related work. Next, in Sect. 3, the proposed approach is presented. Results and evaluation are discussed in Sect. 4. Finally, in Sect. 5 some final considerations and suggestions for future work are provided.

F.A.P.V. de Arruda · J.E.R. de Queiroz · H.M. Gomes (✉)
UFCG, Av. Aprígio Veloso 882, 58429-900 Campina Grande, PB, Brazil
e-mail: hmg@dsc.ufcg.edu.br

F.A.P.V. de Arruda
e-mail: chico@copin.ufcg.edu.br

J.E.R. de Queiroz
e-mail: rangel@dsc.ufcg.edu.br

2 Related work

NPR interest by the scientific community has shown noticeable growth in the 1990s. Some examples that support this observation are: (i) the creation of a dedicated NPR section at the Association for Computing Machinery's Special Interest Group on Graphics and Interactive Techniques (ACM-SIGGRAPH) event in 1998; (ii) the creation of a dedicated NPR section in the 1999 Eurographics Conference; and especially (iii) the emergence of the first symposium dedicated exclusively to the NPR theme—the Non-photorealistic Animation and Rendering in 2000 [34]. Another milestone is the publication of two textbooks that addresses the topic, authored by Gooch and Gooch [19] and Strothotte and Schlechtweg [55].

Several methods for simulating artistic styles were reported in the literature, such as ink illustrations [51, 68], technical illustrations [18], watercolor painting [6], illustrations with graphite [54], impressionist painting [37], cartoons [62], stained glass [42], image abstraction [33, 72], mosaics [14], among others.

In the following subsection, a review focused on generation of sketches is provided.

2.1 NPR sketch generation

Being arguably one of the most important operations in low-level computer vision with a number of available techniques, edge detection is in the forefront of CV, DIP and PR for object detection. Since edges contain a major amount of image information [50], it is essential to have a good understanding on state-of-the-art edge detection techniques. In the context of NPR, edge detection is an important stage in sketch generation.

Additionally, the edge detection method stands out as an auxiliary process for feature extraction [65], segmentation and recognition [28], and image analysis [9]. It is widely used in various CV and DIP algorithms, as it enables the grouping of pixels that divide the image into regions [9]. Such a division is the basis for pre-processing steps, and the edge map serves as input to other algorithms, such as object detection and object recognition.

The creation of sketches and stroke-based illustrations using 3D scene information, instead of using 2D information, was also addressed by several authors in recent years (e.g. [25, 30, 32, 36]). However, the focus of the presented research is restricted to the generation of 2D digital NPR images.

Some NPR techniques use edges when dealing with region outlines, and the detected edges enhance the final look of the rendering. For example, Salisbury et al. [51] developed a system for pen-and-ink illustration that allows the user to use the Canny edge detector to perform the tedious work of generating many pen strokes necessary to form illustrations. Litwinowicz [37] used the Canny edge detector to constrain the user drawn strokes to the object boundaries, thus preserving object silhouettes and fine details of the scenes.

Markosian et al. [38] citing architects that use printed architectural models with overlapped sketch strokes before presenting the models to costumers, gave special attention to sketches, avoiding the impression of completeness of the project. Sayeed and Howard [52] noted that the primary objective of such a representation is to outline objects, using a lower or higher level of detail, allowing recognition using only object boundaries.

DeCarlo and Santella [13] proposed a stylization approach based on eye-tracking data, a robust variant of the Canny edge detector [40], and mean-shift image segmentation. A visual attention map is used to infer information about high-impact areas that are used to create stylized rendering with emphasis on the edges used in the composition of the final stylized image. According to the authors, despite some limitations of the edge detector, the generated edges added expressiveness to the final result.

Tang and Wang [57] proposed a law enforcement system for recognition and indexing of faces using an approach based on face sketches. The sketch generated by their system resembles the real picture, does not portray the hair on the sketch composition and uses elements of shape and texture of the original image. The authors performed validation of the system through experiments using a subjective voting process, as well as numerical experiments, in a large set of images (greater than 300 images).

Tresset and Fol Leymarie [59] developed a study on the flow adopted by an artist to start the drawing process, which begins with observation, includes identification of projections, drawing lines and structuring surfaces, and goes to the composition of the final image. According to the authors, emphasis was given to the lines of the initial draft, which constitute the key element for the composition of the artwork. A computational system was developed for automatic generation of sketches, focusing on human faces. Some implementation details were highlighted, such as applying a more suitable color space for making sketches with fewer false positives for face and skin location, as well as processing the input image by applying an color constancy algorithm [4], such as histogram expansion or retinex.

Kang, Lee and Chui [31] proposed the Flow-based Difference of Gaussians (FDoG) algorithm, which automatically generates line representations of digital images, in a non-photorealistic manner. The technique consists of two parts: (a) computation of a smooth, feature-preserving local edge flow (called Edge Tangent Flow—ETF); and (b) edge detection by using a flow-based difference of Gaussians. The authors emphasized that line drawing is the sim-

plest and oldest form of visual communication. Additionally, the authors stated that several NPR techniques use lines as basis for creating other styles of NPR, but there is scarcely research focusing primarily on line drawing. The work done by Kang, Lee and Chui [31] became an established NPR approach to line drawing picture generation, and thus is used in the objective and subjective comparison of this work, referred to as the FDoG technique.

Xu et al. [70] proposed a method for edge-preserving smoothing that can be used to generate image abstractions. Additionally, this method can be used for edge detection, detail enhancement and clip-art restoration. The method proposed by the authors tries to globally maintain the most prominent set of edges by increasing the steepness of intensity transitions, i.e. attenuating the intensity changes among neighboring pixels. A drawback of the method is oversharpening image regions where there is large illumination variation.

According to Gonzalez and Woods [17], image segmentation plays a crucial role in DIP and CV, especially for nontrivial images and, in general, segmentation algorithms are based on the discontinuity or on the similarity of image gray levels. In the segmentation based on discontinuities, there are three types of feature of interest: isolated points, lines and edges. The authors define edges as significant gray-level variations in an image. Edge images are, in turn, formed by sets of connected edge pixels. Edge detectors are algorithms that aim to identify discontinuities in the image gray levels.

A remarkable aspect of edge detectors is the ability to extract relevant boundaries between objects. Such capacity is the leading inspiration for this research, as the extraction of salient facial features such as eyes, mouth and nose were possible, in order to automatically generate NPR representations of faces. In an earlier work [1], we proposed a method based on multiscale edge detector and a sub-image homomorphic filtering, with parameters optimized by means of a genetic algorithm. An experiment using image measures and a subjective evaluation compared the results of that method with the ones of a plain Canny edge detector. Some benefits of the method were highlighted, but no inferential statistical analysis was presented at that time.

A contribution of this research is the creation of a technique that deals specifically with the generation of sketches, using machine learning, specially focusing on human faces, given the scarcity in the reviewed literature of such techniques. Moreover, the proposed method presents an original and promising approach to the problem of NPR sketching and reports on inferential statistical analyses of objective and subjective comparisons between the outputs of different methods, which we believe is a relevant contribution to the NPR area.

The following subsection includes some related work involving the use of neural networks, a widely disseminated machine learning technique.

2.2 Edge detection using neural networks

Ng, Ong and Noor [43] highlighted issues with classical edge detectors (e.g. rounded edges) and proposed a neural edge detector, with a hybrid approach (partially supervised and partially unsupervised) using a MLP network with input data based on 3×3 image samples, trained with only five images. The detector obtained 1-pixel-width edges when using the supervised input given by a Canny edge detector. Although lacking a more systematic and objective evaluation, the results indicated the validity of a learning-based edge detector.

Suzuki, Horiba and Sugie [56] proposed an edge detector using supervised neural networks for noisy images. Experiments and visual comparison showed that the neural detector generated well connected lines with good noise suppression. However, details regarding the used network architecture were not provided.

Rajab, Woolfson and Morgan [48] proposed an approach for segmenting regions of injured human skin. Their method was based on edge detection using neural networks for pattern recognition of borders, ultimately indicating skin cancer in humans. The training set was made up exclusively of synthetic samples of 3×3 pixels size. After a quantitative evaluation using both synthetic and real-life examples of skin lesions, the technique developed by the authors obtained a better performance for skin-cancer diagnosis when used in a specific kind of problem, namely lesions with different border irregularities.

Becerikli and Demiray [5] have also investigated a neural network edge detector. A Laplacian edge detector was used for supervised training where the training set contained images and edge images corrupted with noise. The authors reported that any classical method of edge detection can be adapted to serve as a basis for training the classifier. The results presented by the authors showed a better visual quality when compared to a Laplacian edge detector, both in graylevel and colored images. However, no experimental objective or systematic evaluation was presented.

Several other authors (e.g. [10, 21, 44, 58, 64]) also used machine learning to detect edges. As previously stated, the lack of specific techniques for generating non-photorealistic representations using machine learning was noticed, specially neural networks.

It is worthy to note that, besides the problems given in the reviewed literature dealing with sketches and NPR techniques, some aspects found in the works that use machine learning to detect edges also emerged and inspired the approach proposed in this research, such as the use of: (i) training samples with noise, (ii) training samples with fixed size, (iii) estimation of contrast of a kernel classification using neural networks, (iv) synthetic samples for training; and (v) several paradigms of neural networks (multilayer perceptron, self-organizing maps, Hopfield networks),

among others. In the following section, the proposed approach is detailed, which was inspired by the related works on NPR, such as the synthetic edge samples used by Rajab, Woolfson and Morgan [48], and the noise-corrupted training samples used in the work of Becerikli and Demiray [5].

3 Proposed approach

The main modules of the proposed approach are shown in Fig. 1. The three main steps, indicated by the numbered boxes in the figure, are:

(1) color pre-processing and smoothing;
(2) neural edge detection; and
(3) post-processing.

Step (1) is subdivided in three pre-processing stages, namely: (i) color correction by color constancy processing; (ii) color space conversion to a more suitable space for sketching; and (iii) smoothing. In step (2), the image is spatially filtered using a 5×5 pixels kernel by means of a multilayer perceptron neural network, trained to classify pixels as edges or non-edges. The result is an edge map that is input to the final step (3), composed of two post-processing stages: (i) brightness and contrast adjustment; and (ii) histogram transformation by gamma correction. The output of this process is the non-photorealistic image sketch. Figure 2 illustrates the main steps regarding the proposed approach for sketch generation.

The specific goals of the research reported in this paper are threefold: (i) defining an approach to generate non-photorealistic sketches, where a validation took place, among a finite set of techniques and parameters; (ii) defining pre- and post-processing algorithms that achieve the best results for a limited set of face images under evaluation; and (iii) experimentally evaluating the proposed approach. In the following subsections we give more details of the steps of the proposed approach, which are aligned with first two specific goals above, whereas the experimental evaluation (iii) is described in Sect. 4.

The pre-processing steps were inspired by the work done and detailed by Tresset and Fol Leymarie [59], as discussed in Sect. 2.1, namely color space transformation and the color constancy algorithm processing. This research was motivated by the scarcity of research focused primarily on line drawings, edge-like abstractions, and related techniques in the NPR area. A neural network approach to such a problem presented an interesting topic for investigation. Post-processing is done to improve visualization of the resulting image.
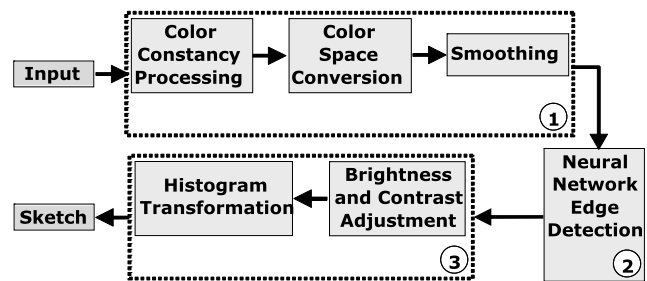

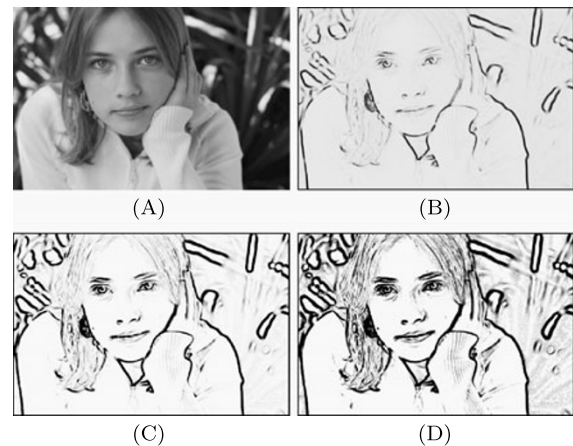**Fig. 1** Main modules of the proposed approach

Fig. 2 Intermediate steps of the proposed approach: (**A**) Smoothed Y channel, (**B**) Neural network result, (**C**) Brightness and contrast adjustment, and (**D**) Histogram transformation by gamma correction

3.1 Pre-processing

Pre-processing is aimed at improving the image for the edge detection step, correcting the image colors to remove illumination artifacts, converting the color space to a space better suitable for sketch generation, and performing smoothing for detail suppression. Additional information is presented next.

3.1.1 Color correction

Scene illumination changes may be a problem for image processing algorithms, because they may introduce false edges caused by shadows, and different illumination conditions. Those problems inspired the use of illumination compensation algorithms to improve the quality of the generated sketch, a feature also highlighted by Tresset and Fol Leymarie [59].

For chromatic correction, experiments using the following algorithms were performed: two color constancy algorithms (Retinex and GrayWorld) and a linear histogram stretch; both are explained next.

The retinex theory assumes that the human visual system consists of three retino-cortical systems [49], responsible for absorption of low, medium and high frequencies of

the visible spectrum. Each system forms a different image, determining the relative brightness of different regions of the scene. According to Land and McCann [35], the borders between these adjacent areas are actively involved in color perception. The division of the brightness between two regions was chosen as a factor that describes the relationship between them. Therefore, if two areas have very different brightness, the ratio deviates from 1, and if they have close luminosities, the ratio approaches 1. If the ratio of the luminosity is measured in various image regions, the dominant color can be estimated, thus allowing an equalization of the overall scene brightness.

Depending on the parameters and the type of input image, the retinex algorithm can provide sharpening, lighting compensation, shadows mitigation, improved color image stability or compression of dynamic range [53]. Several versions of the retinex algorithm have been proposed over time: Single Scale Retinex (SSR), Multi-Scale Retinex (MSR), and Multi-Scale Retinex with Color Restoration (MSRCR) [3]. In this paper, we employed the MSR version.

The gray world hypothesis (GrayWorld assumption) was proposed by Buschsbaum (1980). The hypothesis that the illuminant could be estimated by computing an average value for the light received as a stimulus has a long history, being proposed, additionally, by Edwin Land [8]. Buschsbaum [8] was one of the first to formalize this hypothesis, considering that, on average, the world is gray, i.e., assuming an image with enough variation in color, the average values of the R, G and B components must tend to a common level of gray. Applying the gray world hypothesis to an image is equivalent to force a common average value for each channel R, G and B, which mitigates effects where lighting conditions are not favorable.

The contrast expansion, or linear histogram expansion, is a linear point operation that expands the range of values of the histogram of an image using the full range of available values [7]. In the method proposed in this paper, after the contrast expansion, there is enhancement of the details of the image, improving the dynamic range of overexposed (too bright) images or underexposed (too dark) images.

In this work, the approach that yielded the best results was the one that uses the GrayWorld algorithm, which is supported by the evaluation described in Sect. 4. After the color correction operation, the input image is submitted to a color space conversion as explained next.

3.1.2 Color space conversion

The purpose of a color space is to facilitate the specification of color, standardizing its representation [17]. Several color spaces have been proposed in recent years [45], each one serving a specific purpose. In the luminance–chrominance models (e.g. YIQ, YUV, YCbCr), the image information is divided in a luminance component (brightness, present in the Y component) and two chrominance components (color or tone). The HSV (Hue, Saturation and Value) color space is a model that is related to the human perception of color [7] and presents a more intuitive representation of color than the RGB model. The value component provides the brightness of colors in terms of how much light is reflected or emitted by an object [15]. Other options related to human perception are the CIE LUV and CIE Lab color spaces, which are an attempt to produce a perceptually uniform model in the sense that the distance between two colors in the color space is more closely correlated with the perceptual color distance [7].

Two color spaces were studied and evaluated in this research: (i) HSV, and (ii) YCbCr. The investigation of alternative color spaces is left as a future work. The sketch generation was performed using only the brightness image information (the V channel in HSV model and Y channel in YCbCr model). The use of luminance-only information has been adopted by several other sketch generation approaches (e.g. [31, 41, 57, 59]). After an evaluation procedure described in Sect. 4, the Y channel of the YCbCr color space showed better results for sketch generation.

3.1.3 Smoothing

A smoothing filter is used, in this research, to reduce details of the input image and suppress noise. Two techniques were investigated and tested: (i) Gaussian smoothing [17]; and (ii) median filtering [17]. Although the median filter is well known to be edge preserving, the technique that showed better results was the Gaussian smoothing with $\sigma = 2$, as described in Sect. 4.

3.2 Edge detection

After the pre-processing step described previously, an edge detection step is performed by means of a multi-layer perceptron (MLP) [22] trained with synthetic samples of edges and non-edges, using the supervised learning backpropagation algorithm, with randomly initialized weights and 10 trials for each trained network. The MLP model was chosen after a comparison with two other unsupervised methods, namely Hopfield networks and Kohonen Self-Organizing Maps (SOM). In the comparison, the unsupervised methods presented some issues, such as poor edge and non-edge grouping, over-segmentation (high false positive rate) and discontinuous edge segments. The neural network used in the proposed approach is composed of three layers: input, hidden and output. More details are given in the following paragraphs.

The neural network has two output neurons (x_{edge}, $x_{\neg edge}$). For non-edge training samples, $x_{edge} = 0$ and

$x_{\neg\text{edge}} = 1$, whereas for edge samples, $x_{\neg\text{edge}} = 0$ and x_{edge} was trained with an estimate of the edge's local contrast c (with 1 being the maximum contrast). This value was obtained using the same method as Gomes and Fisher [16]:

$$c = \frac{|L_{\max} - L_{\min}|}{L_{\max} + L_{\min}} \tag{1}$$

where $L_{\max}$ and $L_{\min}$ are the maximum and minimum pixel intensities of a given image region. For training purposes, the contrast calculation is done only on the edges samples. The use of a contrast estimate, rather than the pure image binarization for the neural network training, helped producing a sketch-like effect, not only a pure binary edge map.

The pixel values of a fixed image neighborhood of 5×5 pixels are input to the proposed approach t. In order to reduce the size of the input space and to select discriminative features, a Principal Component Analysis (PCA) [29] was performed, which reduced the input space from 25 to 11 dimensions (the size of the neural network input layer). The PCA steps are now described.

1. Transform all samples into a vector: a 5×5 sample matrix becomes a 1×25 vector (obtaining $\mathbf{X}_{\text{T}}$).
2. Create a matrix $\mathbf{X}^{\text{edge}}$ from the concatenation of all 1×25 input sample vectors.
3. Obtain the mean vector: each column of step 2 forms a 1×25 matrix, with each cell containing the mean value for each linear column values (obtaining $\bar{\mathbf{X}}$).
4. Subtract the mean vector from each sample to obtain a vector $\mathbf{A}$, which has zero mean: $\mathbf{A} = \mathbf{X}_{\text{T}} - \bar{\mathbf{X}}$.
5. Find the covariance matrix $\mathbf{C}$ of size $wh \times wh$, created as follows: $\mathbf{C} = \mathbf{A}\mathbf{A}^{\text{T}}(\mathbf{N} - \mathbf{1})^{-1}$, where $w = 5$ and $h = 5$, given the sample size nature (5×5 dimension).
6. Calculate the normalized eigenvalues and eigenvectors of the covariance C matrix: $\mathbf{C} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\text{T}}$, where matrix $\mathbf{U}$ contain the eigenvectors $[u_1, \ldots, u_D]$ and $\boldsymbol{\Lambda}$ is a diagonal matrix with the associated eigenvalues $[\lambda_1, \ldots, \lambda_D]$. The eigenvalues are ordered in the ascending order by magnitude. The eigenvector associated with the higher eigenvalue is discarded, because this eigenvector typically does not contain discriminative characteristics.
7. Reduce the dimensionality: this reduction is performed by choosing an eigenvalue pivot: eigenvectors with indices higher than the pivot are discarded. The pivot selection, namely the retention factor for the PCA, was chosen as the point representing 99% the data variance.
8. Create a projection subspace $\mathbf{H}$ by concatenating edge and non-edge samples: $\mathbf{H} = [\mathbf{U}^{\text{edge}}\ \mathbf{U}^{\neg\text{edge}}]$, where $\mathbf{H}$ is of size 25×11. The value 25 comes from the sample size (5×5), and 11 is the amount of retained eigenvectors (seven edges and four non-edges).
9. Project the input vector $\mathbf{X}$ into the $\mathbf{H}$ subspace, yielding a projection vector p of 11 positions: $p = \mathbf{X} \times \mathbf{H}$.

Table 1 Samples distribution

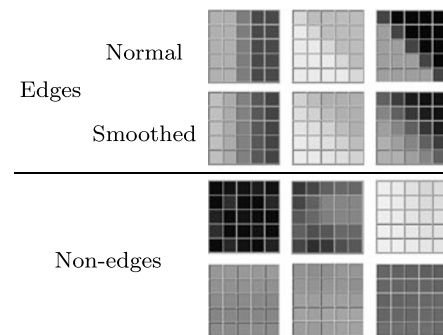| Set	Number of synthetic samples	%
Training	766	60
Test	191	15
Validation	319	25

Fig. 3 Synthetic training samples

The edge and non-edge samples were divided into training, validation, and testing data. The distribution of each of these sets is presented in Table 1.

Four edge orientations were used in the sample training task: 0°, 45°, 90° and 135°. Smoothed versions of the synthetic edge samples were used to increase the training set variability. The non-edges have been defined as random values ranging in the interval [0, 255]. A variation of no more than 20% between the higher and the lower pixel values was allowed in the generation of non-edge patterns. Some of the training samples are shown in Fig. 3.

Finally, in order to define the hidden layer size, the number of hidden neurons was varied in the range [1, 50], and the training process was repeated 10 times. Thus 10 different neural networks were obtained, by randomly initializing the connection weights. The network that presented the higher classification rate against the test set was selected (98.5% hit rate, two neurons in the hidden layer) to compose the proposed approach. The neural network training was implemented in MathWorks MATLAB®, and the application of the trained network was done by a custom made C++ implementation, for fast processing time.

3.3 Post-processing

The post-processing step was composed of two sub-stages: (i) contrast and brightness adjustment, and (ii) histogram transformation. This post-processing stage was applied to each image pixel to modify the dynamic range of the sketch, with the aim of emphasizing higher intensity pixels and facilitate the user visualization of the previous step output. According to experimental tests conducted after a calibration

procedure (see Sect. 4), the brightness value was set to −80, the contrast value was set to 80, and gamma was set to 4.

4 Experimental evaluation, results and discussion

The resulting sketches were submitted to an objective and a subjective evaluation process, as discussed in this section. Additionally, some sketch images are presented at the end of the discussion.

4.1 Objective evaluation

The objective evaluation involved three steps:

(i) the choice of a set of face images taken in non controlled conditions of illumination and viewpoint, followed by an artist ground-truth edge labeling;
(ii) a statistical comparison of the results obtained by some combinations of the pre-processing algorithms, using three different measures; and
(iii) comparison of three techniques for sketch generation.

A ground-truth set of fourteen images was manually labeled by a single artist. All of these images consisted of one pixel-width line drawing sketch of frontal-face images. A pair of ground-truth images used in the comparative experiments is given in Fig. 4. See Annex A of the supplied electronic supplementary material for images of the complete ground-truth set. Martin et al. [39] provided an empirical database for research on image segmentation and boundary detection, but with very few single-person frontal-face images, and thus that database was not used on the experiments.

As mentioned above, one of the purposes of the objective evaluation was to obtain a set of optimal combinations of available algorithms to be used in the pre-processing step of the proposed approach. Three edge similarity measures were considered: (i) **P**eak **S**ignal to **N**oise **R**atio (**PSNR**)

Fig. 4 Examples of ground-truth images used for evaluation: (**A**) Lena image; and (**B**) Thaís image

[47]; (ii) Pratt's **F**igure of **M**erit (**FoM**) [46]; and **S**tructural **SIM**ilarity (**SSIM**) [66, 67]. Those measures calculate a matching score between two images by using either a pixel-based approach or a neighborhood approach.

PSNR uses a pixel-based method for similarity evaluation, usually measured in decibels; thus a higher PSNR means higher similarity. The mean square error (MSE) between images I1 and I2 can be computed as described in (2). This measure and the next one are typically used to evaluate the degradation caused in images by compression algorithms:

$$MSE = \frac{\sum_{m=1,n=1}^{M,N} [I_1(m,n) - I_2(m,n)]^2}{M \times N} \qquad (2)$$

where M and N are the image dimensions. The PSNR measure can be obtained from the MSE as follows:

$$PSNR = 10 \log_{10} \left(\frac{MAX_I^2}{MSE} \right) \qquad (3)$$

where MAX_I is the maximum value that a pixel can retain, e.g., if using an 8 bit pixel representation, the maximum value will be 255.

Pratt's FoM (see (4)) is a measure for edge deviation between an ideal edge map and another one produced by an edge detector [47]. FoM varies in the interval [0, 1], where 1 is a perfect match between the detected edge image and the ground-truth image:

$$FoM = \frac{1}{I_N} \sum_{i=1}^{I_A} \frac{1}{(1 + a \times d^2(i))} \qquad (4)$$

where $I_N = \max(I_I, I_A)$, I_I represents the number of pixels in the ground-truth, I_A represents the number of pixels in the evaluated image, a is an constant scale factor (usually 1/9), $d(i)$ is the maximum distance between the position of the pixel in the original image and the ground-truth image.

SSIM is a similarity measure that uses local patterns of pixel intensities, previously normalized for luminance and contrast, and is computed as a function of three comparisons: luminance, contrast and structure. SSIM varies in the interval [0, 1] [66]. The SSIM calculation for a pair of test images x and y is given by (5):

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\,cov_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \qquad (5)$$

where μ_x is the x mean; μ_y is the y mean; σ_x^2 is the x variance; σ_y^2 is the y variance; cov_{xy} is the covariance of x and y; $c_1 = (k_1 L)^2$, $c_2 = (k_2 L)^2$ are used to stabilize the division with small denominators, being L the dynamic range of pixels values, typically $2^{\#bitsperpixel} - 1$; $k_1 = 0.01$ and $k_2 = 0.03$ by default.

Numerical-experimental results involving the above mentioned metrics are shown in Table 2, in which each group of algorithms is named as G_Z, where Z represents the group
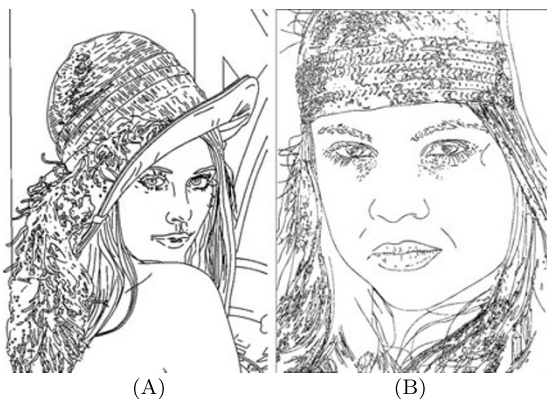
Table 2 Objective evaluation of proposed approach

Measure/ Statistics		GrayWorld				Retinex				Histogram stretch			
		Y Channel		V Channel		Y Channel		V Channel		Y Channel		V Channel	
		Gauss. G1	Mean G2	Gauss. G3	Mean G4	Gauss. G5	Mean G6	Gauss. G7	Mean G8	Gauss. G9	Mean G10	Gauss. G11	Mean G12
PSNR	$\bar{X}$	8.0598	8.1130	8.0096	8.0603	7.0912	7.2380	7.2193	7.3628	7.9911	8.0472	7.9899	8.0395
	s	1.3808	1.3611	1.3914	1.3781	1.0045	1.0051	1.0499	1.0480	1.3433	1.3270	1.1182	1.1041
SSIM	$\bar{X}$	0.1763	0.1756	0.1744	0.1739	0.1667	0.1672	0.1681	0.1684	0.1757	0.1750	0.1765	0.1759
	s	0.0364	0.0367	0.0382	0.0384	0.0297	0.0298	0.0301	0.0306	0.0361	0.0364	0.0329	0.0330
FoM	$\bar{X}$	0.5199	0.4801	0.4386	0.4395	0.4221	0.4228	0.4136	0.4161	0.4722	0.5103	0.4445	0.4426
	s	0.0733	0.0723	0.0744	0.0743	0.0974	0.0943	0.0977	0.0949	0.0727	0.0718	0.0758	0.0751

Table 3 Objective evaluation of available techniques

Statistics	Proposed approach			DoG			FDoG		
	PSNR	FoM	SSIM	PSNR	FoM	SSIM	PSNR	FoM	SSIM
$\bar{X}$	8.05989	0.51997	0.17630	7.23072	0.42980	0.10765	8.72429	0.51053	0.17226
s	1.38082	0.07332	0.03649	1.14493	0.10110	0.03947	1.26687	0.11386	0.03461

number. Each group is composed of three pre-processing algorithms, e.g., the G1 (Group 1) algorithms are: (i) Gray-World color constancy processing; (ii) a RGB to YCrCb conversion and the use of Y channel; and (iii) a Gaussian smoothing, as indicated in the table. The table lines contain the mean value $\bar{X}$ and standard deviation s for the indicated measure, obtained for 14 images.

After computing Table 2, an Analysis of Variance (ANOVA) with $\alpha = 0.05$ was performed in order to test the null hypothesis that there were no differences between the means of the measure values.

In the context of this research, the ANOVA test was used where one way (or factor) is to be evaluated (e.g., the measure result) and more than 2 categorized groups were presented for evaluation (e.g., the proposed approach groups G1, G2, …, G12).

The ANOVA values for each measure were as follows: (i) PSNR: $F(11, 156) = 1.54252$ and **p-value** < 0.12145; (ii) SIM: $F(11, 156) = 0.184008$ and **p-value** < 0.99824; (iii) FoM: $F(11, 156) = 2.693397$ and **p-value** < 0.00336. A statistically significant difference was found only in the FoM results; thus, the null hypothesis was rejected.

After the ANOVA test, a multiple comparison Tukey t-test [61] was conducted in order to identify which group of available algorithms was significantly different from the others. The Tukey t-test showed that group 1 of algorithms (G1) was statistically different from the others when the FoM measure is taken as reference. Thus, the chosen algorithms for the pre-processing step were GrayWorld color constancy processing, use of the Y Channel and Gaussian smoothing.

After defining the pre-processing algorithms, an objective comparison between the proposed approach the DoG, and the FDoG edge detectors was performed. Although the proposed approach is aimed at producing gray-level representations, the objective evaluation described above was done with binary ones, given that the ground-truth was provided as binary images. The comparison was feasible because the DoG and FDoG outputs resemble a sketch. As Canny edges present single-pixel thicknesses, which interfere in FoM calculation, the Canny detector was not included in this comparison. Nonetheless it was included in the subjective evaluation process (see Sect. 4.2).

For each comparative experiment, the mean and the standard deviation of the 14 ground-truth images involving the three considered measures and the three chosen algorithms are given in Table 3.

Similarly to the experiment aimed to define the pre-processing algorithms for the proposed approach, an ANOVA test was performed to acknowledge statistical differences between the mean values of the three measures, regarding the chosen algorithms, using $\alpha = 0.05$. According to that analysis, significant statistical differences between the means of the three measures exist: (i) PSNR: $F(2, 39) = 4.876696$ and **p-value** < 0.01287; (ii) SSIM: $F(2, 39) = 15.24669$ and **p-value** < 0.00012; (iii) FoM: $F(2, 39) = 3.61137$ and **p-value** < 0.03639.

The Tukey t-test for pairwise comparison (see Table 4), using $\alpha = 0.05$, showed that no statistically significant difference existed between the proposed approach and the FDoG algorithm, considering all the three measures. Additionally, a statistically significant difference was found on

Table 4 Tukey t-test results for objective evaluation of available techniques (P.A. is the proposed approach)

Contrast	Significant difference for $\alpha = 0.05$		
	PSNR	FoM	SSIM
P.A. Vs. FDoG	No	No	No
P.A. Vs. DoG	No	Yes	Yes
FDoG Vs. DoG	Yes	No	Yes

the following cases: (i) DoG and FDoG (PSNR); (ii) the approach proposed and DoG (SSIM); and (iii) the approach proposed and DoG (FoM).

Nevertheless, the objective assessment cannot exhaust the nuances and variability in the results of techniques that aim at visual appeal to human observers, inherently ambiguous and subjective. This aspect does not invalidate the objective evaluation held earlier, but raises the need for finding a method or tool in order to better evaluate those nuances. Subjective evaluation appears as a second necessary technique for evaluation.

4.2 Subjective evaluation

Wallraven et al. [60] provided an experimental framework in which perceptual features of animated and real facial expressions can be addressed. Such a framework allows the evaluation of different animation methods.

Gooch et al. [20] also studied facial illustration evaluation, and showed that illustrations and caricatures are as effective as photographs in recognition tasks.

Winnemöller et al. [69] also evaluated stylization algorithms using a memory style game, reported a performance improvement when measuring recognition speed and memory of abstracted images of familiar faces.

Heath et al. [23] applied a visual rating score method for testing a set of edge detectors for a number of natural images. The evaluation criterion was a subjective rating as to how well humans could easily, quickly, and accurately recognize objects within a test image from the edge map produced by an edge detector. The evaluation process consisted of two parts:

(i) parameter optimization of each edge detection algorithm; and
(ii) human visual assessment of the results.

The subjective evaluation process done on this research was inspired by the above studies in human facial illustrations and specially by the work by Heath et al. [23], with some modifications, namely: (i) the method for calibration of algorithm parameters; (ii) the choice of the image set; (iii) the use of a web-based polling application; and (iv) the choice of algorithms (FDoG, DoG, Canny and the approach proposed).

Fig. 5 Subjective evaluation—Part 1—User Interface

Table 5 First part results: parameter calibration

Method	Parameter	Value
Canny	*Maximum threshold*	101
	Minimum threshold	24
DoG	*Kernel 1*	15
	Kernel 2	9
FDoG	*Half_w*	9
	M	1
	ρ	3.9937902
	$\sigma 1$	1.3616940
	$\sigma 2$	4.6370154
	Gray threshold	0.99
Proposed approach	Color constancy algorithm	GrayWorld
	Channel	Y
	Smoothing	Gaussian (5×5)
	Brightness	-80
	Contrast	80
	Gamma	4

In order to avoid biases when defining the parameters for each edge detector, a parameter calibration (the *first part* of the subjective evaluation) was conducted by a voting scheme. An excerpt of the user interface used in this part of the experiment is given in Fig. 5.

Twenty test users, with ages varying between 18 and 28 years, calibrated the four algorithms. In this *first part* of the experiment all users were software engineers and with image processing programming skills. Twenty face images were used (Annex B of the supplied electronic supplementary material contains those images). For each method, a random variation of parameters was generated, resulting in a set of eight sketch images for each of the 20 face images.

In the *first part* of the experiment, each user has evaluated 640 images (20 face images $\times$ 8 sketches $\times$ 4 algorithms). The parameters that showed the best results for the evaluated face set are given in Table 5.
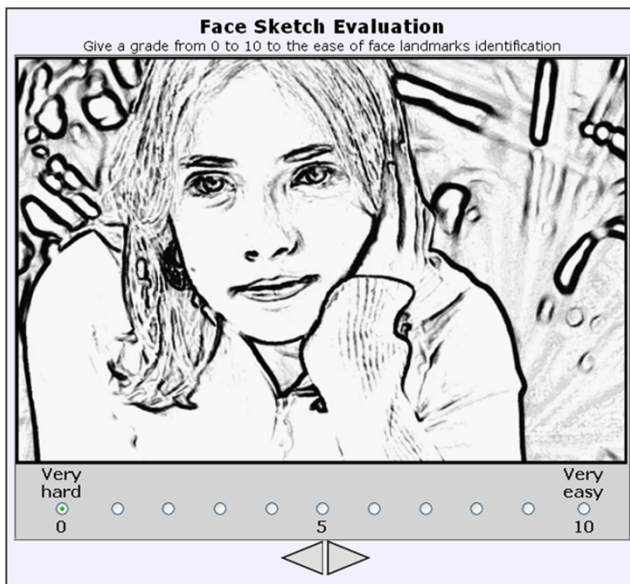
Fig. 6 Subjective evaluation—Part 2—User Interface

Table 6 Second part results: relative performance for each subjectively evaluated technique

Detector	$\bar{X}$	s
FDoG	7,0917	2,0880
Proposed approach	5,2375	1,8449
DoG	4,9917	1,9210
Canny	3,8375	2,2301

Table 7 Tukey t-test results for objective evaluation of available techniques

Contrast	Significant for $\alpha = 0.05$
FDoG Vs. Canny	Yes
FDoG Vs. DoG	Yes
FDoG Vs. Proposed approach	Yes
Proposed approach Vs. Canny	Yes
Proposed approach Vs. DoG	No
DoG Vs. Canny	Yes

The number of participants is aligned with previous NPR evaluation work; e.g., when performing recognition and learning tasks regarding caricature renderings, Gooch [20] performed the evaluation with 42 users for the recognition speed task (divided in three two-part experiments, each with 14 participants), and 30 students for the learning speed task (divided into three groups of 10), Winnemöller et al. [69] conducted two task-based studies, with 10 participants in each study, and Isenberg et al. [27] performed an observational study with 24 participants in total.

It is important to note that, after the calibration phase in the subjective evaluation, the results showed that the pre-processing modules for the proposed approach were equal in both evaluation assessments (objective and subjective), i.e., the objective evaluation results (Table 2, group G1 of algorithms) did not differ from subjective evaluation (Table 5—proposed approach parameters) when defining the best combination of pre-processing algorithms.

Similarly to the work of Heath et al. [23], the *second part* of the experiment considered a voting scheme where the images generated by the calibrated parameters (obtained in the first phase) were used. An excerpt of the user interface used in this part of the experiment is presented in Fig. 6.

In Heath et al. [23], 16 users evaluated 280 edge images, (2 edge images for each detector × 28 images of varied themes × 5 detectors). Moreover, the comparison of each edge detectors was made indirectly, taking into account a score in the interval [0, 7], using sheets of paper for voting.

The range used in the present work differed from the range proposed by Heath et al. [23], because we used the interval [0, 10], and the user was instructed to vote 0 (zero) if the face characteristics (eyes, eyebrows, nose, mouth, ears) were "very hard" to identify, and to vote 10 (ten) if the face

characteristics were "very easy" to identify. In summary, the goal was to ask the participants to rate the relative performance of each algorithm when rendering frontal-face images in a non-photorealistic way, i.e., whether the algorithm outputs contained recognizable face landmarks.

In this paper the *second part* of the subjective evaluation was assessed by 25 users, without any image processing background, with ages varying between 18 and 28 years, who evaluated 10 face images generated by the four compared methods, with one image per method (totaling 40 images per person).

During a pilot test, it was found that using the same amount of images (20) used in step 1 of the subjective experiment resulted in a prohibitively time-consuming task (about 1 hour and 10 minutes per user). After reducing the number of images to 10 and after reformulating the interface of the experiment, the average participation time was reduced to 25 minutes.

Table 6 contains the mean ($\bar{X}$) and the standard deviation (s) of the performance for each evaluated algorithm after the computation of the subjective evaluation results.

Once again, a single factor ANOVA was performed to acknowledge statistical differences between the mean values of the votes computed for each algorithm, using $\alpha = 0.05$, relatively to the results presented in the second part result. The ANOVA results were $\mathbf{F}(3, 956) = 106.1174$ and **p-value** $< 2.65E-59$. After the evaluation of ANOVA results, it can be observed that there is significant statistical difference between the means of the three measures. Finally, a multiple comparison Tukey t-test was performed, as shown in Table 7.

The Tukey t-test results indicate that there is a statistically significant difference between the FDoG and all other approaches, using $\alpha = 0.05$, and that there are no statistically significant differences between the proposed approach and the DoG algorithm.

A visual comparison of the evaluated algorithms is presented in Fig. 7, where it can be seen that the proposed technique presents a sketchy aspect. This sketchy aspect can be observed by the highly variable line width that is obtained by the edge importance and by the facial components highlighting, like stronger lines presented on drawing of the mouth and eyes.

Additional results obtained using the proposed approach are presented in Annex C of the supplied electronic supplementary material.

5 Final considerations and further work

An approach for digital image stylization by means of a neural network sketching process was presented in this paper. In an objective evaluation, a set of test images with faces was used to compare the proposed approach with existing sketch-like rendering systems. The objective evaluation showed, after an ANOVA analysis and Tukey t-test, that the proposed approach did not differ from a state-of-the-art technique for sketch rendering (namely FDoG).

A subjective experiment proved to be complementary to the objective analysis. In the subjective evaluation, the calibration parameters for the four algorithms (Canny, DoG, proposed approach and FDoG) strengthened the conclusions derived from the objective evaluation. The preprocessing modules used in the proposed approach are equal in both evaluation assessments (objective and subjective). The Canny algorithm was considered unsuitable for generating non-photorealistic representations of human faces, given the low average rank value (3.84) obtained within the [0, 10] scale. Moreover, the proposed approach presented a higher average rank when compared to the DoG and Canny algorithms, and a lower rank when compared to the FDoG algorithm.

Hertzmann [26] states that a missing ingredient from most NPR research is the lack of experimental study and evaluation of results, not only the simplistic evaluation of several test images and the visual appeal of results produced by a technique. Within that context, the research presented brings about an important contribution by applying a formal analysis in the evaluation of methods for human face sketch generation.

The average execution time for the four algorithms (Canny, DoG, FDoG and the proposed approach) are shown in Table 8. This time was computed after 20 runs on different input images with average resolution of 780k pixels. The results were obtained using a standard 2 GHz ×86

Fig. 7 Visual comparison between techniques: (**A**) Canny, (**B**) DoG, (**C**) FDoG, and (**D**) proposed approach

personal computer with 2 GB of RAM, running Windows XP operating system. The programs were written in C/C++ and compiled using Microsoft Visual Studio 2008. Standard OpenCV functions were used for color space conversion and thresholding. The execution time measurements did not take into account any i/o or memory allocation procedures.

Canny and DoG run at nearly real time due to the low complexity of these algorithms. The proposed approach took 3.645 seconds on average to process an image, whereas the average execution time for the FDoG algorithm was 3.235 seconds. The slightly higher execution time of the proposed approach, when compared to the FDoG, is partly

explained by the pre- and post-processing steps of the proposed approach, which are not required in the other compared algorithms.

Table 9 allows a visual comparison to be made of the evaluated algorithms when processing noise-corrupted input images. Two types of noise were considered: (i) Gaussian white noise with zero mean and three different variances (s); and (ii) salt & pepper noise with three different densities (D). Three different noise levels were considered, varying from nearly imperceptible noise artifacts to a very noisy image. Visually, the FDoG and the proposed approach showed some degree of robustness against the two types of noise considered, whereas the Canny and the DoG produced comparatively poor results.

In order to further improve the subjective evaluation results (as reported on Tables 6 and 7), alternative color spaces as well as other pre- and post-processing techniques will be investigated, such as homomorphic and bilateral filtering. Moreover, we intend to evaluate the impact of training with

Table 8 Average execution time

Detector	Canny	DoG	FDoG	Proposed approach
Average time (s)	0.005	0.005	3.235	3.645

Table 9 Outputs of the evaluated algorithms when considering different levels of Gaussian and salt & pepper noise

additional synthetic edge orientations, using different kernel sizes and applying an edge thinning process to the output.

Another consideration regarding the neural network step is that the algorithm was trained on synthetic data, but was applied in real-world data. Some neural network trials were done with real-world data, with samples taken directly from stylized images. Unfortunately, the neural network failed at getting a consistent edge mapping, given that some test samples (among millions of samples) were ambiguous, leading at the same time to edge and non-edge mapping. Nevertheless, future work might include filtered real-world data in the neural network step of the proposed approach.

Future work may also involve increasing the number of images in the first step of the objective evaluation, including noise-corrupted images, in order to strengthen the result that placed group 1 (G1) as the top pre-processing strategy among all investigated groups.

Moreover, the second part of the subjective evaluation may be extended by means of additional scales to capture other relevant dimensions, such as aesthetics, complexity, face quality and overall image quality. Finally, evaluations based on levels of expertise may be performed (e.g. users with and without prior experience with illustration or NPR) in order to broaden the conclusions drawn from the subjective evaluation.

References

1. Arruda FA, Porto VA, Gomes HM, de Queiroz JER, Moroney N (2007) Facial sketching based on sub-image illumination removal and multiscale edge filtering. In: Proc IEEE SITIS 2007. IEEE Comp Soc, Los Alamitos, pp 520–527
2. Barile P, Ciesielski V, Trist K (2008) Non-photorealistic rendering using genetic programming. In: Proc 7th int conf on sim evolution and learning, vol 5361. Springer, Berlin, pp 299–308
3. Barnard K, Funt B (1999) Investigations into multi-scale retinex. In: Color imaging in multimedia. Wiley, New York, pp 9–17
4. Barnard K, Cardei V, Funt B (2002) A comparison of computational color constancy algorithms. IEEE Trans Image Process 11(9):972–984
5. Becerikli Y, Demiray HE, Ayhan M, Aktas K (2006) Alternative neural network based edge detection. Neural Inf Process – Lett Rev 10:193–199
6. Bousseau A, Kaplan M, Thollot J, Sillion F (2006) Interactive watercolor rendering with temporal coherence and abstraction. In: Proc NPAR 2006. ACM, New York, pp 141–149
7. Bovik AC (2005) Handbook of image and video processing. Academic Press, San Diego
8. Buschsbaum G (1980) A spatial processor model for object colour perception. J Franklin Inst 310(1):1–26
9. Chabrier S, Laurent H, Rosenberger C, Emile B (2008) Comparative study of contour detection evaluation criteria based on dissimilarity measures. J Image Video Proc 8(2):1–13
10. Chang CY (2004) A contextual-based hopfield neural network for medical image edge detection. In: Proc IEEE ICME 2004, vol 2, pp 1011–1014
11. Chen H, Liu Z, Rose C, Xu Y, Shum HY, Salesin D (2004) Example-based composite sketching of human portraits. In: Proc NPAR 2004. ACM, New York, pp 95–153
12. Cootes TF, Edwards GJ, Taylor CJ (2001) Active appearance models. IEEE Trans Pattern Anal Mach Intell 23(6):681–685
13. DeCarlo D, Santella A (2002) Stylization and abstraction of photographs. In: Proc SIGGRAPH 2002. ACM, New York, pp 769–776
14. Dobashi Y, Haga T, Johan H, Nishita T (2002) A method for creating mosaic images using Voronoi diagrams. In: Proc Eurographics short presentations, pp 341–348
15. Ebner M (2007) Color constancy. Wiley, New York
16. Gomes HM, Fisher R (2003) Primal-sketch feature extraction from a log-polar image. Pattern Recognit Lett 24(7):983–992
17. Gonzalez RC, RE Woods (2007) Digital image processing. Prentice Hall, New York
18. Gooch A, Gooch B, Shirley P, Cohen E (1998) A non-photorealistic lighting model for automatic technical illustration. In: Proc SIGGRAPH'98, pp 447–452, ACM, New York
19. Gooch B, Gooch A (2001) Non-photorealistic rendering. Peters, Wellesley
20. Gooch B, Reinhard E, Gooch A (2004) Human facial illustrations: Creation and psychophysical evaluation. ACM Trans Graph 23(1):27–44
21. Gupta L, Sukhendu D (2006) Texture edge detection using multi-resolution features and som. In: Proc ICPR 2006. IEEE Press, New York, pp 199–202
22. Haykin S (2008) Neural networks: a comprehensive foundation. Prentice Hall, New York
23. Heath MD, Sarkar S, Sanocki T, Bowyer KW (1997) Robust visual method for assessing the relative performance of edge-detection algorithms. IEEE Trans Pattern Anal Mach Intell 19(12):1338–1359
24. Heisele B, Ho P, Wu J, Poggio T (2003) Face recognition: Component-based versus global approaches. Comput Vis Image Underst 91(1–2):6–21
25. Hertzmann A (1999) Introduction to 3d non-photorealistic rendering: Silhouettes and outlines. Springer, Berlin
26. Hertzmann A (2010) Non-photorealistic rendering and the science of art. In: Proc NPAR 2010. ACM, New York, pp 147–157
27. Isenberg T, Neumann P, Carpendale S, Sousa MC, Jorge JA (2006) Non-photorealistic rendering in context: an observational study. In: Proc NPAR 2006. ACM, New York, pp 115–126
28. Jiang X, Marti C, Irniger C, Bunke H (2006) Distance measures for image segmentation evaluation. EURASIP J Appl Signal Process 2006(1):1–10
29. Jolliffe IT (2002) Principal component analysis. Springer, Berlin
30. Kalnins R, Markosian L, Meier BJ, Kowalski MA, Lee JC, Davidson PL, Webb M, Hughes JF, Finkelstein A (2002) Wysiwyg npr: drawing strokes directly on 3d models. In: Proc SIGGRAPH 2002. ACM, New York, pp 755–762
31. Kang H, Lee S, Chui CK (2007) Coherent line drawing. In: Proc NPAR 2007. ACM, New York, pp 43–50
32. Kolliopoulos A, Wang JM, Hertzmann A (2006) Segmentation-based 3d artistic rendering. In: Proc EGSR 2006, pp 361–370
33. Kyprianidis JE, Kang H, Döllner J (2009) Image and video abstraction by anisotropic Kuwahara filtering. Comput Graph Forum 28(7):1955–1963
34. Lake A, Marshall C, Harris M, Blackstein M (2000) Stylized rendering techniques for scalable real-time 3d animation. In: Proc NPAR 2000. ACM, New York, pp 13–20
35. Land EH, McCann JJ (1971) Lightness and retinex theory. J Opt Soc Am 61(1):1–11
36. Lee H, Kwon S, Lee S (2006) Real-time pencil rendering. In: Proc NPAR 2006. ACM, New York, pp 37–45

37. Litwinowicz P (1997) Processing images and video for an impressionist effect. In: Proc SIGGRAPH'97. ACM/Addison-Wesley, New York, pp 407–414

38. Markosian L, Kowalski MA, Goldstein D, Trychin SJ, Hughes JF, Bourdev LD (1997) Real-time nonphotorealistic rendering. In: Proc SIGGRAPH'97. ACM/Addison-Wesley, New York, pp 415–420

39. Martin D, Fowlkes C, Tal D, Malik J (2001) A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Proc 8th int'l conf comp vision, vol 2, pp 416–423

40. Meer P, Georgescu B (2001) Edge detection with embedded confidence. IEEE Trans Pattern Anal Mach Intell 23(12):1351–1365

41. Mignotte M (2003) Unsupervised statistical sketching for non-photorealistic rendering models. In: Proc ICIP 2003, vol 3, pp 573–576. IEEE Comp Soc, Los Alamitos

42. Mould D (2003) A stained glass image filter. In: Proc 13th EGWR, eurographics association, pp 20–25

43. Ng G, Ong FL, Noor NM (1995) Neural edge detector. ASEAN J Sci Tech Dev 12(1):35–42

44. Pian Z, Gao L, Wang K, Guo L, Wu J (2007) Edge enhancement post-processing using hopfield neural net. In: Proc ISNN 2007. Springer, Berlin, pp 846–852

45. Poynton C (2003) Digital video and HDTV algorithms and interfaces. San Mateo, Morgan Kaufmann

46. Pratt WK (2007) Digital image processing: PIKS scientific inside. Wiley, New York

47. Prieto MS, AR Allen (2003) A similarity metric for edge images. IEEE Trans Pattern Anal Mach Intell 25(10):1265–1273

48. Rajab MI, Woolfson MS, Morgan SP (2004) Application of region-based segmentation and neural network edge detection to skin lesions. Comput Med Imaging Graph 28(1-2):61–68

49. Rizzi A, Gatta C, Marini D (2004) From retinex to automatic color equalization: issues in developing a new algorithm for unsupervised color equalization. J Electron Imaging 13(1):75–84

50. Roushdy M (2006) Comparative study of edge detection algorithms applying on the grayscale noisy image using morphological filter. Int J Graph Vis Image Process 6(4):17–23

51. Salisbury MP, Anderson SE, Barzel R, Salesin DH (1994) Interactive pen-and-ink illustration. In: Proc SIGGRAPH'94, pp 101–108. ACM, New York

52. Sayeed R, Howard T (2006) State-of-the-art of non-photorealistic rendering for visualisation. In: Proc TP.CG.06. Pergamon, Elmsford, pp 1–10

53. Sharma G (2002) Digital color imaging handbook. CRC Press, Boca Raton

54. Sousa MC, Buchanan JW (1999) Computer-generated graphite pencil rendering of 3d polygonal models. Comput Graph Forum 18(3):195–208

55. Strothotte T, Schlechtweg S (2002) Non-photorealistic computer graphics: modeling, rendering and animation. San Mateo, Morgan Kaufmann

56. Suzuki K, Horiba T, Sugie N (2000) Edge detection from noisy images using a neural edge detector. In: Proc IEEE SPS workshop, vol 2. IEEE Comp Soc, Los Alamitos, pp 487–496

57. Tang X, Wang X (2003) Face sketch synthesis and recognition. In: Proc IEEE ICCV 2003. IEEE, New York, p 687

58. Toivanen PJ, Ansamaki J, Parkkinen JPS, Mielikainen J (2003) Edge detection in multispectral images using the self-organizing map. Pattern Recognit Lett 24(16):2987–2994

59. Tresset P, Leymarie FF (2005) Generative portrait sketching. In: Proc VSMM 2005, pp 739–748, Hal Twaites

60. Wallraven C, Breidt M, Cunningham DW, Bülthoff HH (2008) Evaluating the perceptual realism of animated facial expressions. ACM Trans Appl Percept 4:1–20

61. Walpole RE, Myers RH, Myers SL, Ye K (2010) Probability and statistics for engineers and scientists. Prentice Hall, New York

62. Wang J, Xu Y, Shum HY, Cohen MF (2004a) Video tooning. In: Proc SIGGRAPH 2004. ACM, New York, pp 574–583

63. Wang K, Yin B, Guo J, Ma S (2004b) Face animation parameters extraction and driving. In: Proc IEEE ISCIT 2004, vol 2. IEEE, New York, pp 1242–1245

64. Wang K, Gao L, Pian Z, Guo L, Wu J (2007) Edge detection combined entropy threshold and self-organizing map(som). In: Proc ISNN 2007. Springer, Berlin, pp 931–937

65. Wang S, Ge F, Liu T (2006) Evaluating edge detection through boundary detection. EURASIP J Appl Signal Proc 1–15

66. Wang Z, Bovik AC (2002) A universal image quality index. IEEE Signal Process Lett 9(3):81–84

67. Wang Z, Simoncelli E (2005) An adaptive linear system framework for image distortion analysis. In: Proc IEEE ICIP 2005, vol 2. IEEE, New York, pp 1160–1163

68. Winkenbach G, Salesin DH (1994) Computer-generated pen-and-ink illustration. In: Proc SIGGRAPH'94. ACM, New York, pp 91–100

69. Winnemöller H (2006) Perceptually-motivated non-photorealistic graphics. PhD thesis, Northwestern University

70. Xu L, Lu C, Xu Y, Jia J (2011) Image smoothing via l0 gradient minimization. ACM Trans Graph 30(6):174

71. Yang MH, Kriegman DJ, Ahuja N (2002) Detecting faces in images: A survey. IEEE Trans Pattern Anal Mach Intell 24(1):34–58

72. Zeng K, Zhao M, Xiong C, Zhu SC (2009) From image parsing to painterly rendering. ACM Trans Graph 29(1):1–11