Empathic Painting: Interactive stylization through observed emotional state

Maria Shugrina*
Computer Science Department
Boston University
Boston MA, USA.

Margrit Betke
Computer Science Department
Boston University
Boston MA, USA.

John Collomosse
Department of Computer Science
University of Bath
Bath, England.

Abstract

We present the "empathic painting" — an interactive painterly rendering whose appearance adapts in real time to reflect the perceived emotional state of the viewer. The empathic painting is an experiment into the feasibility of using high level control parameters (namely, emotional state) to replace the plethora of low-level constraints users must typically set to affect the output of artistic rendering algorithms. We describe a suite of Computer Vision algorithms capable of recognising users' facial expressions through the detection of facial action units derived from the FACS scheme. Action units are mapped to vectors within a continuous 2D space representing emotional state, from which we in turn derive a continuous mapping to the style parameters of a simple but fast segmentation-based painterly rendering algorithm. The result is a digital canvas capable of smoothly varying its painterly style at approximately 4 frames per second, providing a novel user interactive experience using only commodity hardware.

CR Categories: I.3.4 [Computer Graphics]: Graphics Utilities—Paint Systems; I.4.8 [Image Processing and Computer Vision]: Motion, Time-varying imagery, Tracking—; I.4.9 [Image Processing and Computer Vision]: Non-photorealistic Rendering—;

Keywords: Painterly rendering, Animation, Emotion, FACS.

1 Introduction

The development of image based non-photorealistic rendering (NPR) techniques, in particular painterly rendering algorithms, has gathered considerable momentum in recent years. A number of automatic painting algorithms now exist, guided by heuristics tailored to emulate particular media or stroke placement styles. Although the majority of algorithms tend to focus on one particular artistic style (for example impressionism [Litwinowicz 1997]), some seek to emulate a broader range of styles using a plethora of user configurable low-level parameters (for example, [Hertzmann 1998] varies brush size, colour jitter, and stroke length to move between pseudo "expressionist" and "pointillist" styles). Often these parameters can be time consuming to set — both due to their number, and due to their low-level nature, which can make them non-intuitive for inexperienced users to manipulate when aiming for a conceptually higher level effect (e.g. a gloomy painting, or an energetic, cheerful composition). This can result in a slow, iterative trial and error process before the user is able to instantiate their desired results.

This paper reports an experiment into the feasibility of using high level style parameters to express control over a painterly stylization. Specifically, we allow the user to interactively specify the emotional ambiance, or "mood" that they wish to convey through a particular artistic rendering; examples might include despair, anger, or elation. In the spirit of earlier work by Hertzmann and Perlin [2000] we have developed a novel interface and interactive NPR installation for this experiment, which we term the "empathic painting". This system is capable of estimating the viewer's purposefully displayed emotional state through automatic facial expression recognition, and affecting rendering parameters to mirror that state on a "live" digital canvas (Figure 1). The purpose of the exhibit is to allow users to both explore the design space of a painterly rendering parameterised by higher level concepts such as emotion, and to experience a novel means of interacting with digital artwork. The two principal technical contributions of this work are therefore:

- A computer vision component capable of estimating the emotional state that the viewing user tries to convey through observation of facial expression. We describe novel approaches to detecting three facial action units: mouth curl, wideness of eyes, raised and furrowed brows, as defined under the Facial Action Coding Scheme (FACS) [Ekman and Friesen 1978]. These action units are mapped to vectors within the 2D pleasure-arousal emotional space proposed by Russell [1997].

- A fast, segmentation based painting algorithm capable of real-time stylisation of photographs using commodity hardware. Mappings are created from the pleasure-arousal space to stylization parameters of the algorithm, so enabling the painting to react to the state output from the vision component. We ensure that stroke attributes vary in a temporally coherent manner as the emotional parameters expressed by the viewer change.

Figure 2 summaries the basic architecture of the system — we give a detailed description of the vision and painting algorithms in Sections 3 and 4 respectively. The papers concludes with a gallery of results and discussion in Section 5.

2 Related Work

A number of machine-assisted 2D painting environments were developed in the early '90s for the purpose of painterly rendering [Haeberli 1990; Haggerty 1991], however Litwinowicz [1997] was the first to propose a fully automated algorithm. Drawing

*This work was undertaken during a summer internship at the Department of Computer Science, University of Bath, England

upon Haeberli's [1990] earlier semi-automatic paint systems, Litwinowicz produced convincing *impressionist style* paintings by aligning small rectangular strokes tangential to Sobel edge gradients in the image, and stochastically perturbing their colour. A multi-scale approach to painting using curved β-spline strokes was later proposed in [Hertzmann 1998]. Spline control points were obtained by hopping between pixels in directions tangential to Sobel edges. The process operated at several discrete spatial scales, concentrating stroke detail in high frequency areas of the image. An iterative adaptation of this algorithm that produced more accurate paintings via active contour relaxation was presented in [Hertzmann 2001]. Other early painterly rendering algorithms such as [Treavett and Chen 1997] and [Shiraishi and Yamaguchi 2000] also made use of local image processing operators to guide stroke placement, specifically pixel variance within a window. Our approach is most closely aligned with more recent work that harnesses mid-level computer vision techniques to model scene content, with the aim of refining aesthetics. Segmentation of the image into homogeneous greyscale regions was first proposed by [Gooch et al. 2002]; strokes were painted along medial axes of each segmented region leading to a significant reduction in the number of brush strokes whilst still preserving fine detail. Segmentation was also used by [DeCarlo and Santella 2002; Santella and DeCarlo 2004] to produce painterly abstractions in which a human gaze tracker was used to correlate level of detail in the painting with perceptually salient detail in the source image. An automatic system for salience adaptive painting, driven by machine learning rather than run-time interaction, was recently presented in [Collomosse and Hall 2005].

The majority of painterly rendering algorithms focus upon a particular media type or artistic style — predominantly through a procedural approach, but in some cases by learning from example [Hertzmann et al. 2001]. Instead, our work is aligned with algorithms encompassing a range of visual styles selectable via user parameterisation. For example, Hertzmann claims expressionism, pointillism, impressionism and "abstract" styles through the variation of low level parameters such as stroke length. Similarly low-level parameters may be used to tune the visual style of paintings in [Hays and Essa 2004]. We form a mapping onto such parameters using a high level emotional parameterisation derived from a facial tracking system. Facial tracking has been used previously to drive NPR animation by piecewise retargetting of tracked motion to move components of NPR facial avatars [Buck et al. 2000; Li et al. 2001]. However our work focusses on the derivation of emotional context from the state of a facial tracker, and the reflection of that context in the style of a painterly visualisation. As such our work is aligned with recent studies exploring the affective qualities of NPR [Duke et al. 2003; Halper et al. 2003].

Our work also draws parallels with recent literature addressing painterly animation; we too are concerned with the smooth animation of paintings as style parameters vary over time. A central problem to any painterly animation is that of suppressing stroke flicker, caused by process non-determinism (either due to image noise or pseudo-random elements of the algorithm). This is typically addressed by maintaining as much visual state as possible between frames, and preventing sharp changes in that state over time. Although little attention has been devoted to the problem of real-time painterly animation for interaction, notably Hertzmann [2000] adapted his earlier static rendering technique [1998] to "paint over" regions containing significant motion — so preserving strokes from previous frames to mitigate against flicker. Optical flow has also been used to translate strokes between frames whilst preserving visual attributes such as orientation and colour [Litwinowicz 1997; Kovacs and Sziranyi 2002]. Recently spatio-temporal constraints

Figure 1: A user interacts with the painting in a live installation with a camera mounted on top of the monitor; the rendering's visual style adapts in real time to reflect the perceived emotional state of the user.

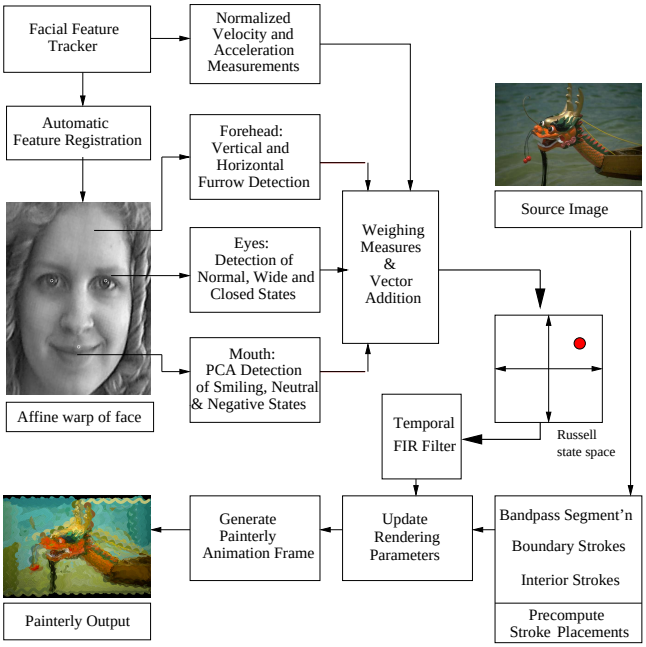


Figure 2: Architecture of the empathic painting system. User emotional state is estimated by detecting and combining visual cues derived from the Facial Action Coding Scheme (FACS). The state estimate is then passed to the painterly rendering algorithm for visualisation.

have been enforced through video to smooth the trajectories of strokes [Hays and Essa 2004; Wang et al. 2004; Collomosse et al. 2005]. However, the large temporal windows required by these techniques make them unsuitable for real-time processing.

3 Estimation of Emotional State

In this section we first describe the mapping from visual cues to emotional state and then explain the process by which those cues are detected in a monocular view of the user's face.

Several approaches exist for deriving emotional context from facial expressions. Componential approaches associate emotional interpretations with distinct units of facial expression, such as those defined by the Facial Action Coding System (FACS) [Ekman

and Friesen 1978]. Expressions are then typically classified by their proximity to one of a set of prototypical emotions in an emotional space (usually accommodating joy, anger, sadness, surprise, fear and disgust [Black and Yacoob 1997]). To facilitate smooth animation of rendering style we do not wish to categorise expressions, but instead require a continuous mapping from observations to the emotional space. To this end, we have adapted an approach discussed in [Russel and Fernández-Dols 1997] for our empathic painting application, under which observations of facial action units (AUs) are mapped to vectors in Russell's 2D pleasure-arousal emotional space, adapted from Plutchik's activation-evaluation "emotional wheel" [Plutchik 1980]) — see Figure 3. Notwithstanding our real-time performance restrictions, the task of accurately recovering all 46 of the FACS action units is a significant challenge to contemporary vision techniques. Rather than adopt an invasive marker-based system which might deter incidental interactions with our system, we have adapted three significant visual cues used by FACS: forehead furrows (to measure raised or lowered brows), curvature of the mouth and eye openness. Variation in these cues is manifested in seven individual facial action units detectable by our system. Emotional expression, however, is not accomplished solely through facial deformations, and so we have included an additional eighth "agitation" action unit that relies on motion data derived from the elementary tracker embedded in our system.

Individual video frames are submitted to the vision component for classification at successive instants written as time t. Facial features are first located and subjected to affine registration (Section 3.1). The presence of our eight action units is then measured by a bank of four independent classifiers (each dealing with a distinct visual cue) described in Sections 3.2-3.5 respectively. Mid-level feature analysis within each classifier allows us to infer the likelihood of a particular action unit being present at the current instant, which is used to produce a signed scalar weighting $w_i(t)$ on corresponding vectors $\underline{v}_i$ associated with each action unit. In order to cover the entire emotional range, we have adapted the pleasure and arousal values suggested by Snodgrass [1997] to the vectors $\underline{v}_i$ used in our system (the mappings between action units and $\underline{v}_i$ are tabulated in Figure 3). A simple weighted summation of these vectors (center of Figure 2) yields a 2D point $\underline{\mathscr{P}}(t)$:

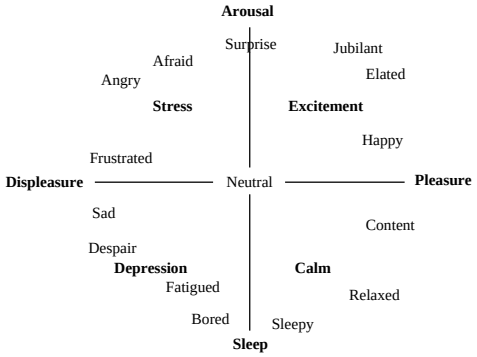
$$\underline{\mathscr{P}}(t) = \sum_{i=1}^{8} w_i(t)\underline{v}_i \qquad (1)$$

in Russell's emotional space that is passed to the painterly rendering component (bottom of Figure 2 and Section 4).

3.1 Feature Tracking and Registration

Our system incorporates a fully automated eye and mouth tracker that compensates for affine variation of the face with respect to the camera plane. The dominant facial region within the frame is first detected using a standard Haar wavelet based approach [Viola and Jones 2001]. Three points corresponding to the centroids of the eye and mouth regions are computed and used to recover an affine transform that warps the face to a normalised reference frame for subsequent analysis by the facial action unit classifiers.

The eyes and mouth regions are tracked using deformable template matching with the normalized correlation coefficient [Chau and Betke 2005]. The initial few frames of video are used to bootstrap the tracker, during which the subject is assumed to be facing the camera with neutral expression and open eyes. Pre-defined templates combined with anthropometric properties of the average face are used to estimate the location of the subject's eyes and mouth.



Our Action Units (Ekman's AUs)	Vector	Pleasure	Arousal
Brow Fully Raised (1+2)	$\underline{v}_1$	0.0	0.7
Inner Brow Raised (1)	$\underline{v}_2$	-0.5	-0.7
Brow Furrowed (4)	$\underline{v}_3$	-0.5	0.7
Negative Mouth (various)	$\underline{v}_4$	-1.0	0.0
Positive Mouth / Smile (12+25)	$\underline{v}_5$	1.0	0.0
Wide Eyes (5)	$\underline{v}_6$	0.0	0.4
Closed Eyes (41, 43, 45)	$\underline{v}_7$	0.0	-1.0
Agitation (N/A)	$\underline{v}_8$	0.0	0.5

Figure 3: Above: Russell's 2D pleasure-arousal space used to express emotional state in our system. Below: Our facial action units (AUs) adapted from the FACS scheme (Ekman's original AU noted in parentheses) and their vector mappings in Russell's 2D space.

These locations are refined via thresholding, and subject-specific templates are then cut for use by the tracker throughout the remainder of the session. To improve tracking accuracy, templates are scaled and rotated according to the moments of the facial region identified within each frame. We also use a Kalman filter to estimate the second order motion parameters of each tracked feature. This enables us to both restrict template search around a predicted feature location, so reducing the high costs of template-matching and improve robustness to occlusions.

3.2 Mouth Shape Analysis

The first of our classifiers considers the detection of mouth curl using a data-driven approach based on principal component analysis (PCA). Principal component analysis has been applied extensively to the problem of face and expression recognition (for example [Pentland et al. 1994]), and although such approaches require an *a priori* supervised training step, their run-time efficiency is appealing for our application. Our classifier accepts a mouth image isolated by the tracker at time t, and returns a signed value on the normalised continuum spanning the two extremes of upward and downward mouth curl. Our basic technique is to model the statistical distribution of each of three training image classes, each class representing a discrete value on this continuum (specifically downward, neutral and upward curl). These distributions are built offline using a representative sample of users and under varying lighting conditions (we used around 800 frames in our experiments). Given a novel image at run time we are able to infer the likelihood of membership to each distribution, and so the state of the mouth.

We represent each mouth image as a vector of concatenated pixel grey-values which we write $\underline{x} \in \Re^n$ where n is the number of of pixels. Training images $\underline{x}_i$ are analysed using PCA and projected via their major eigenvectors into a lower dimensional space containing 97% of training set variation. We have observed the point distributions of training classes to be poorly approximated by simple linear

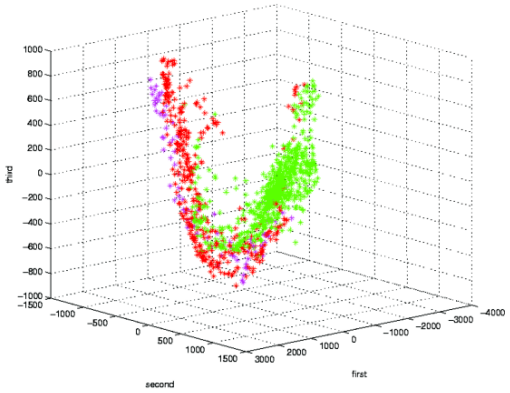


Figure 4: Mouth curl distributions. A visualization of the upward curl (red), neutral (magenta) and downward curl (green) training image sets, projected onto the first three principal components of the distribution. Gaussian mixture models were used to model each class.

models, and so cluster each of the three classes in piecewise fashion using Gaussian Mixture Models (see Figure 4). Labelling classes $c \in \{1,2,3\}$, this results in three models $\mathscr{M}_c$, each comprising a set of eigenmodels $\{\underline{\mu}_{c,i}, \underline{\underline{U}}_{c,i}, \underline{\underline{V}}_{c,i}\}$ where $\underline{\mu}_{c,i}$ is a mean, $\underline{\underline{U}}_{c,i}$ a set of eigenvectors, and $\underline{\underline{V}}_{c,i}$ their corresponding eigenvalues. Given a mouth image $\underline{x}_t$ at run time we may determine the Mahalanobis distance to model $\mathscr{M}_c$ (that is, the inverse likelihood of $\underline{x}_t$ being a member of class c) using:

$$l(c,\underline{x}_t) = min_i \left[(\underline{\mu}_{c,i} - \underline{x}_t)^T \underline{\underline{U}}_{c,i} \underline{\underline{V}}_{c,i}^{-1} \underline{\underline{U}}_{c,i}^T (\underline{\mu}_{c,i} - \underline{x}_t) \right] \quad (2)$$

By computing simple linear combinations of class memberships we obtain the parameters $w_4(t)$ and $w_5(t)$ used to weigh the positive $\underline{v}_4$ and negative $\underline{v}_5$ mouth vectors in the emotional state space.

3.3 Brow Analysis

Raised or lowered brows feature in many of the FACS action units, and are detected in our system by measuring the presence of furrows in the forehead region (located above the eyes using simple anthropometric properties of the face). Specifically we aim to measure fully raised brows (derived from FACS AUs 1 and 2, $\underline{v}_1$), furrowed brows (FACS AU 1, $\underline{v}_2$), raised inner brows (FACS AU 4, $\underline{v}_3$).

The forehead is often susceptible to specularities which can adversely affect classification based on a data-driven approaches such as PCA used in Section 3.2. We have instead adopted a feature-based approach in which line segments are detected using a low-resolution Hough transform operating over the Canny edge detected forehead region. Vertical wrinkles often indicate furrowed brows while horizontal wrinkles indicate raised brows [Tian et al. 2001]. In addition, a raised inner brow can also result in wrinkles of arbitrary orientation. Pixels detected within line segments are therefore grouped into three sets according to line orientation: approximately vertical lines (set V), horizontal lines (H) and "neither" (N). This yields the following respective weights:

$$w_1 = min\left(0, \omega\left(\frac{\|H\|}{T} - \frac{\|V\|}{T}\right)\right)$$

$$w_2 = min\left(0, \omega\left(\frac{\|V\|}{T} - \frac{\|H\|}{T}\right)\right)$$

$$w_3 = min\left(0, \omega\left(1 - |(\frac{\|H\|}{T} - \frac{\|V\|}{T})|\right)\right) \quad (3)$$

on our action unit vectors $\underline{v}_{1..3}$, where $\|.\|$ indicates set cardinality, $T = \|H \cup V \cup N\|$, and ω is a weighting factor introduced to prevent misclassification of wrinkles that have become permanent with age (computed as a ratio of edge pixels detected in the current frame, to those detected in the "neutral face" used to bootstrap the tracker).

3.4 Eye State Analysis

The degree to which the user's eye is open is a powerful cue to alertness (arousal). We aim to measure a range that spans the closed, normal and wide-open states of the eye. Much work has been done in blink detection, and we extend the approach of [Chau and Betke 2005] in the analysis of open and closed eye states to accommodate wide eyes as well. The correlation score used during tracking of the eye region generally adheres to the inequality $1 \geq C_{normal} > C_{wide} > C_{closed} \geq 0$, where C subscript indicates the correlation score of a typical eye state. However this is insufficient to accurately differentiate between eye states. We refine accuracy by combining this measure with a further normalised "iris correlation measure", computed only within the iris region of the eyes, located using the Hough Transform. We write the iris and eye region correlation measures as $z(t)$ and $c(t)$ respectively, computed between the tracker templates and the current frame at time t. We observed the behaviour of $c(.)$ and $z(.)$ over each of our three discrete eye states and modelled their likelihoods by six Gaussian probability distributions $P_i(c(.))$ and $Q_i(z(.))$ respectively (where $i \in \{closed, normal, wide\}$). The likelihood of the eye being in discrete state i at time t is then the product:

$$p_i(t) = P_i(c(t))Q_i(z(t)) \quad (4)$$

where $p_{closed}(t)$ and $p_{wide}(t)$ correspond to the facial action unit weights $w_6(t)$ and $w_7(t)$ used in Equation 1. Note that in order to minimise computational overhead we calculate the pixel difference within eye regions in consecutive frames to determine if eye state has changed significantly. If the inter-frame difference is less than a sensitivity threshold then we simply return $p_i(t)$=$p_i(t-1)$.

3.5 Agitation

Observations of human interaction suggest that indicators of a person's arousal state include not only speed (how quickly the face moves), but also acceleration magnitude (how frequently the velocity of the face changes, for example head-shaking). We base a simple agitation classifier on these observations. Using the Kalman filter state computed by the tracker (Section 3.1) we obtain values for the speed and acceleration of the midpoint $\underline{m}$ between the eyes. These measures are normalised with respect to scale of the face, yielding $|\dot{\underline{m}}|$ and $|\ddot{\underline{m}}|$ respectively. The weight of agitation vector $\underline{v}_8$ is then computed by:

$$w_8 = \frac{|\dot{\underline{m}}|}{2} + 2(|\ddot{\underline{m}}| - \varepsilon) \quad (5)$$

which introduces a bias towards rapidly changing motion, and allows very low acceleration to result in a slightly negative arousal value (we have found $\varepsilon = 0.15$ to be a suitable value through experimentation).

4 Image Stylization

We now describe the fast multi-resolution algorithm used to generate frames of painterly animation from a source photograph. In

Figure 5: Sample output generated by our painterly rendering algorithm (Section 4), original photographs inset. Rendering parameters were: $p_1 = 0.3, p_2 = (-0.3, 0.0), p_3 = 0.31, p_4 = 0.0, p_5 = 0.5$ (left image), and $p_1 = 0.0, p_2 = (0,0), p_3 = 0.0, p_4 = 0.0, p_5 = 0.0$ (right image).

contrast to the majority of painterly rendering algorithms we operate using multi-scale segmentation only, eschewing image gradient measures for region shape properties to guide stroke placement. Most image based painterly algorithms aim to conserve detail by aligning strokes tangential to edges detected in the image using intensity gradient direction (e.g. [Litwinowicz 1997; Hertzmann 1998; Collomosse and Hall 2005]) or statistical moments [Treavett and Chen 1997; Shiraishi and Yamaguchi 2000]. However, such measures often become noisy in relatively flat (non-edge) areas resulting in chaotic orientation of strokes where there is no predominant direction of intensity gradient or variance. This can result in poor aesthetics in object interiors. Some approaches have used expensive interpolation techniques, e.g. thin-plate splines [Litwinowicz 1997] and radial-basis functions [Hays and Essa 2004], to create smooth direction fields between sparse, irregular samples of strong edge direction. Here we borrow from our previous work in painterly video stylisation [Collomosse and Hall 2005] — performing mid-level image analysis through segmentation and consistent alignment of strokes within segmented regions. Regions are rendered by laying down "interior" strokes of similar orientation and "boundary" strokes placed around the region's perimeter (including any holes that may exist within the region interior). A similar approach to stroke placement was recently applied to the generation of paintings from object-space by [Kolliopoulos 2005], although this work focused primarily upon deriving a temporally coherent geometry segmentation of scenes.

4.1 Stroke placement

We begin by creating a colour band-pass pyramid segmentation of the source image using the EDISON algorithm [Christoudias et al. 2002] (after [DeCarlo and Santella 2002]). Successively coarser layers of the pyramid are sub-sampled without low-pass filtering to preserve corners and discontinuities in the boundaries of large regions. Pyramid layers are then rendered in coarse to fine order. For each layer, we first render the "interior" strokes of all regions, then the "boundary" strokes of all regions. Brush strokes are formed using Catmull-Rom piecewise cubic splines, the control points of which are computed from the binary image of each region. We now describe the stroke placement process for rendering one such region.

4.1.1 Interior Strokes

The interior of a segmented region is first filled using a modified boundary-fill algorithm that paints strokes tangential to the region's principal axis. We compute the eigenvectors of pixel coordinates inside the region, and temporarily warp the region so that its principal eigenvector is parallel to the horizontal. Scan-lines are then traversed, and strokes are started and terminated as region boundaries are encountered; the vertical interval between scan-lines during processing is proportional to stroke thickness. Stroke control points are distributed uniformly over the stroke's length, and jittered via small translations to disguise the regularity of the stroke placement process (see Figure 6). Stroke colour is computed using the mean colour of the original image, sampled at pixels corresponding to the stroke's control points. Stroke thickness is set on a per region basis, in proportion to area. In the case of very large regions, thickness is capped and strokes are painted horizontally (after [Kolliopoulos 2005]) to preserve natural appearance.

4.1.2 Boundary Strokes

The boundary of the segmented region is encoded as a run-length compressed Freeman (chain) code, and stroke control points generated by vectorizing this code sequence. Points on the chain code are visited one at a time and added to an initially empty "working set". Upon each point's addition, we sum the distance between all points in the working set to a line drawn between the first and last points in that set. If the distance is above a threshold (or no further points remain in the chain code), we output the most recently added point as a stroke control point. The working set is then emptied. Figure 6 illustrates the control points (magenta) generated from a chain code (cyan). A brush stroke is terminated, and a new stroke started, when the angle between adjacent control points rises above a preset threshold (we use 50°). A stroke may also be terminated if the colour of a new control point differs from the mean colour of those already present to the stroke by more than a preset threshold

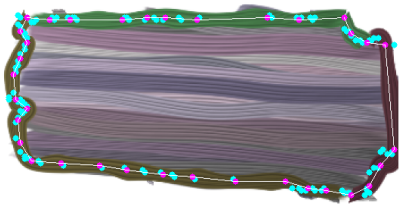


Figure 6: False colour illustration of interior and boundary stroke placement within the brown region of the *DRAGON* image (see Figure 7, bottom). Chain code entries indicated in cyan, and stroke control points deduced from these in magenta. Stroke medial axes indicated in white.

(we use 0.3 in a normalised CIELAB* space). Stroke thickness and colour are set as with the interior stroke placement process.

4.2 Rendering Parameters

Although it can be difficult to separate the subject of the painting from the emotional atmosphere it expresses, several psychological studies suggest strong correlations between certain types of strokes or colour combinations and the emotional context or "mood" portrayed by a visual artwork. For instance, color psychology suggests that bright colors are exciting, while cooler colours such as blues and green are calming [Mahnke 1996]. We have incorporated a gamut of such psychological cues from the literature within our painting algorithm. The expression of these cues is governed by the state estimate $\underline{\mathscr{P}}(t)$ determined by the vision component in Section 3.2. In order to prevent prevent flicker and temporal discontinuity between frames, we first low-pass filter this signal using a finite impulse response (FIR) filter of the form:

$$\underline{\mathscr{P}'}(t) = \alpha\underline{\mathscr{P}}(t) + (1-\alpha)\underline{\mathscr{P}'}(t-1) \tag{6}$$

where α represents the expected speed of emotional state variation; we found $\alpha = 0.3$ suitable for our installation. We create a mapping from $\underline{\mathscr{P}'}(t)$ to a number of normalised stylisation parameters governing stroke placement, tonal variation, stroke denotation style and accuracy, detailed descriptions of which are given in Sections 4.2.1–4.2.3. In the majority of cases we create a continuous mapping from $\underline{\mathscr{P}'}$ to the i^{th} parameter p_i, by allowing the user to sketch a trajectory across the pleasure-arousal space between the two extrema of the parameter, $p_i = [0,1]$. The trajectory is expressed as a parametric curve $\underline{T}_i(p_i)$ where p_i is an arc-length parameterisation. With each trajectory $\underline{T}_i(.)$ defined *a priori*, cue expression parameters are recovered by solving:

$$p_i = argmin_x(|\underline{\mathscr{P}'} - \underline{T}_i(x)|) \tag{7}$$

The exception is the cue governing tonal variation, which is expressed as a non-linear function of $\underline{\mathscr{P}'}$ to reflect the separate effects of pleasure and arousal with respect to colour. We now describe each of the style parameters in turn.

4.2.1 Region Turbulence (p_1)

We wish to encompass a gamut of brush stroke styles ranging from the calm, serene washes of a watercolour to the energetic swirls of a Van Gogh oil or the chaotic strokes of a Turner or Dzigurski seascape. Such effects are often manifested within expansive regions (e.g. skies), whilst also maintaining fidelity around the edges of regions [Butler et al. 1994]. We can introduce a range of similar effects by repeatedly performing boundary stroke placement (Section 4.1.2) with regions being subjected to morphological erosion

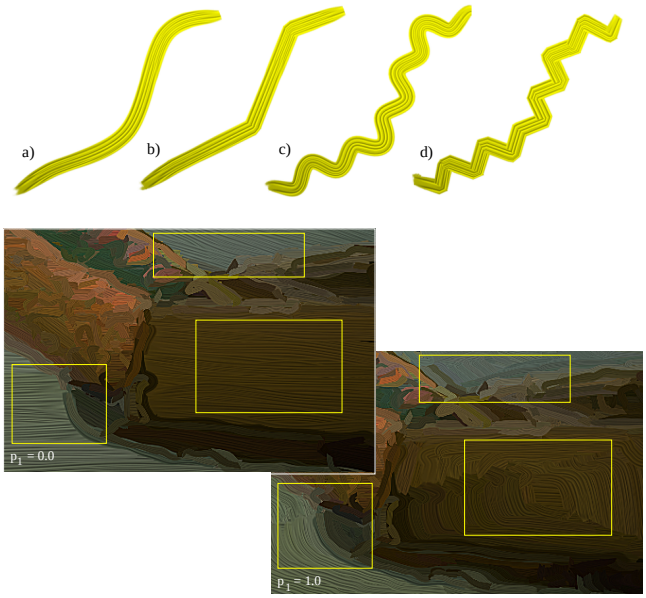


Figure 7: Some of the effects generated by varying style parameters. Top: A single brush stroke rendered with a range of undulation (p_4) and jaggedness (p_5) parameters. From left to right; (a) $p_4 = 0.0, p_5 = 0.0$, (b) $p_4 = 0.0, p_5 = 1.0$, (c) $p_4 = 0.7, p_5 = 0.0$, (d) $p_4 = 0.7, p_5 = 0.8$. Bottom: The effect of the turbulence (p_1) parameter on interior stroke regions (most pronounced in the regions indicated).

prior to each iteration. This has the effect of allowing boundary strokes to encroach upon the interiors of regions in an unstructured manner, so breaking up smooth expanses. The number of process iterations is proportional to rendering parameter p_1, the trajectory of which is typically sketched in broad alignment with the arousal axis in the pleasure-arousal space. Figure 7 (bottom) illustrates the effect of this parameter on a section of the *DRAGON* painting.

4.2.2 Tonal Variation ($\underline{p_2} = \underline{\mathscr{P}'}(t)$)

It is well known that certain combinations of colours can evoke particular emotions, so helping to convey a particular mood to a composition. Although a number of studies attempt to attribute emotional semantics to particular colour ranges, there is little general consensus except for a few special cases (such as expressions of anger, or sadness) [Pickford 1972]. Rather, Wright and Rainwater [1962] have found the notion of happiness (pleasantness) to be primarily dependent on colour brightness (luminance), and to a lesser degree on saturation. Intuitively arousal corresponds to colour saturation, but can also be linked to hue. Wright and Rainwater's study has shown calmness to be blue-correlated [Wright and Rainwater 1962], but according to Mahnke blue may also suggest depression and cold [Mahnke 1996].

We have defined a number of transfer functions that operate upon hue, saturation and luminance as a mechanism for instantiating the colour heuristics we have distilled from the literature. The complex psychological theories underpinning colour and emotion generate non-linear mappings of hue, saturation and luminance variation to the pleasure-arousal space. We approximate these piecewise with a collection of linear transfer functions — different functions are applied in each of six regions of the space. Figure 8 illustrates the boundaries of these regions, and the transfer

functions used over the pleasure-arousal space. Functions $G(x)$ and $U(x)$ correspond to greying and un-greying (scaling saturation in proportion to x), while $D(x)$ and $L(x)$ correspond to lightening and darkening (scaling luminance in proportion to x). The operation of the latter function is capped for "boundary" brush strokes to prevent bleaching of fine detail. Care is taken in blending the constants of proportionality to prevent visible discontinuities near the boundaries defined over the pleasure-arousal space.

Functions $T_1(x)$ and $T_2(x)$, indicated in Figure 8, are two special cases that encode hue variation consistent with aroused displeasure (anger) and apathetic displeasure (depression). Hue is manipulated via an RGB space transformation prior to saturation and luminance manipulations. In the former case $T_1(x)$, predominantly red colors are reddened and green (associated with calm) is reduced (in proportion to x). These effects combine with the saturation and luminance transformations already present to produce the combination of aroused reds and dismal darks that appear in psychological literature in association with anger. In the latter case $T_2(x)$ we increase the blue in proportion to x to generate a monotonous shift into the blue spectrum, associated with sadness and calm. Colours are also desaturated and darkened in accordance with transformations already present in that quadrant of the space.

4.2.3 Stroke Denotation Style ($p_3 - p_5$)

Henver [1935] surveyed artists' use of line and found a correlation between denotation style and the emotional context conveyed to viewers by a drawing. Gently sloping curves were observed to depict serenity, laziness, or tender-sentimentality in a subject — whereas harsh angles or jagged strokes depicted vigorousity, or power in a subject, expressing fury or agitation. These studies expanded upon earlier investigations into line style by [Poffenberger and Barrows 1924] and [Lundholm 1921] who observed impressions of agitation or unpleasantness to be conveyed by lines exhibiting discontinuities and angularities. Halper *et al* form similar conclusions in a recent observational study [Halper et al. 2003] of the affective nature of NPR, and similar visual cues remain common in contemporary comic-strip inking. In response, we have introduced two parameters to control stroke undulation (p_3) and jaggedness (p_4). The former is typically sketched along the arousal axis, and the latter diagonally from the displeasure-aroused quadrant to the pleasure-sleep quadrant on Russell's emotional space.

When rendering a given stroke we create an arc-length parameterisation over the piecewise Catmull-Rom spline that smoothly interpolates stroke control points, which we write as $\underline{P}(s)$. To introduce stroke jaggedness we create a further, linear interpolation over the control points $\underline{L}(s)$ using the same arc-length parameterisation. We now form a new stroke plotting function $\underline{Q}(s)$ by combining these functions:

$$\underline{Q}(s) = (1 - p_4)\underline{P}(s) + p_4\underline{L}(s) \tag{8}$$

We introduce undulations into the stroke by translating each point on the stroke curve $\underline{P'}(s)$ along its normal, according to a periodic function $\underline{O}(s)$. The frequency and amplitude of this function are proportional to p_3, but are randomly perturbed to avoid introducing regularity into the painting. To maintain the desired jaggedness of the stroke, this function is a weighted sum of a smooth and discontinuous periodic signals $\underline{O_P}(s)$ and $\underline{O_L}(s)$; in our case sine and triangular waves respectively:

$$\underline{Q}(s) = (1 - p_4)(\underline{P}(s) + \underline{O_P}(s, p_3)) + p_4(\underline{L}(s) + \underline{O_L}(s, p_3)) \tag{9}$$

We have also introduced a further parameter (p_5) to dampen the effects of undulation (p_3) on interior strokes, which can often lead to highly chaotic stroke placements in the backgrounds of paintings. We have found that this parameter's mapping can be sketched either in parallel with the arousal axis, or on the diagonal between the pleasure-aroused and displeasure-sleep quadrants, depending on user preference.

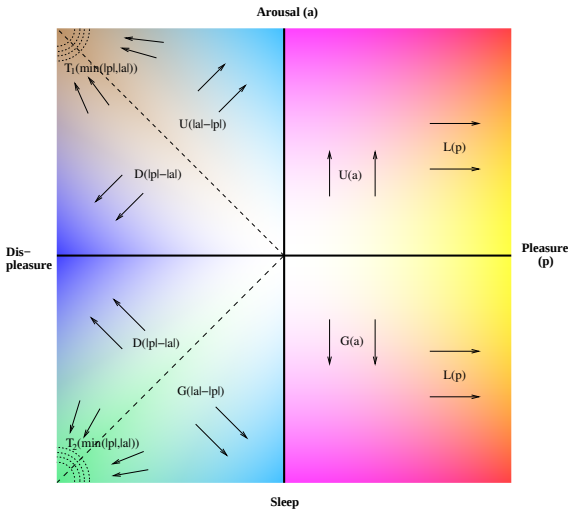


Figure 8: Schematic illustrating the various colour transformations performed within regions of the pleasure-arousal space. False colour is used here to represent intensity of particular colour transfer function and is not related to colour changes manifested in paintings. Functions G, U, L, D, T_1 and T_2 are defined in Section 4.2.2.

4.3 Rendering Process

To enable real-time rendering, computation of the band-pass segmentation and stroke placements is performed as a pre-processing step during system initialisation. During the interactive phase of execution, the stroke list need only been rendered and stroke attributes modulated in accordance with parameters $p_{1..5}$. Strokes are then textured and bump-mapped using standard graphics hardware to give an oil painted appearance [Hertzmann 2002]. Unfortunately parameter p_1 alters the number of strokes in the painting, so introducing complications under this optimization. Real-time operation can be maintained by quantising the range of p_1 and labelling portions of the stroke list as visible only in particular discrete intervals of p_1. The requirement for a stochastic process to drive some aspects of painting (for example stroke undulation) could introduce temporal incoherence as parameters vary over time. We avoid this by assigning unique identifiers to each stroke on creation. These numbers are used to seed the pseudo-random number generator before plotting each stroke, so ensuring a reproducible but seemingly "random" series of perturbations for each stroke. Because we introduce artifacts such as jaggedness and undulation at the stroke rendering stage (rather than by adding additional control points to the stroke), we need only maintain a static stroke list for rendering — this improves both real-time performance and also temporal coherence of the painting.

5 Results and Discussion

We have described an interactive system for creating "empathic paintings" the styles of which react in real time to reflect the perceived emotional state of the viewer. The system comprises a computer vision component for expression recognition, and a computer

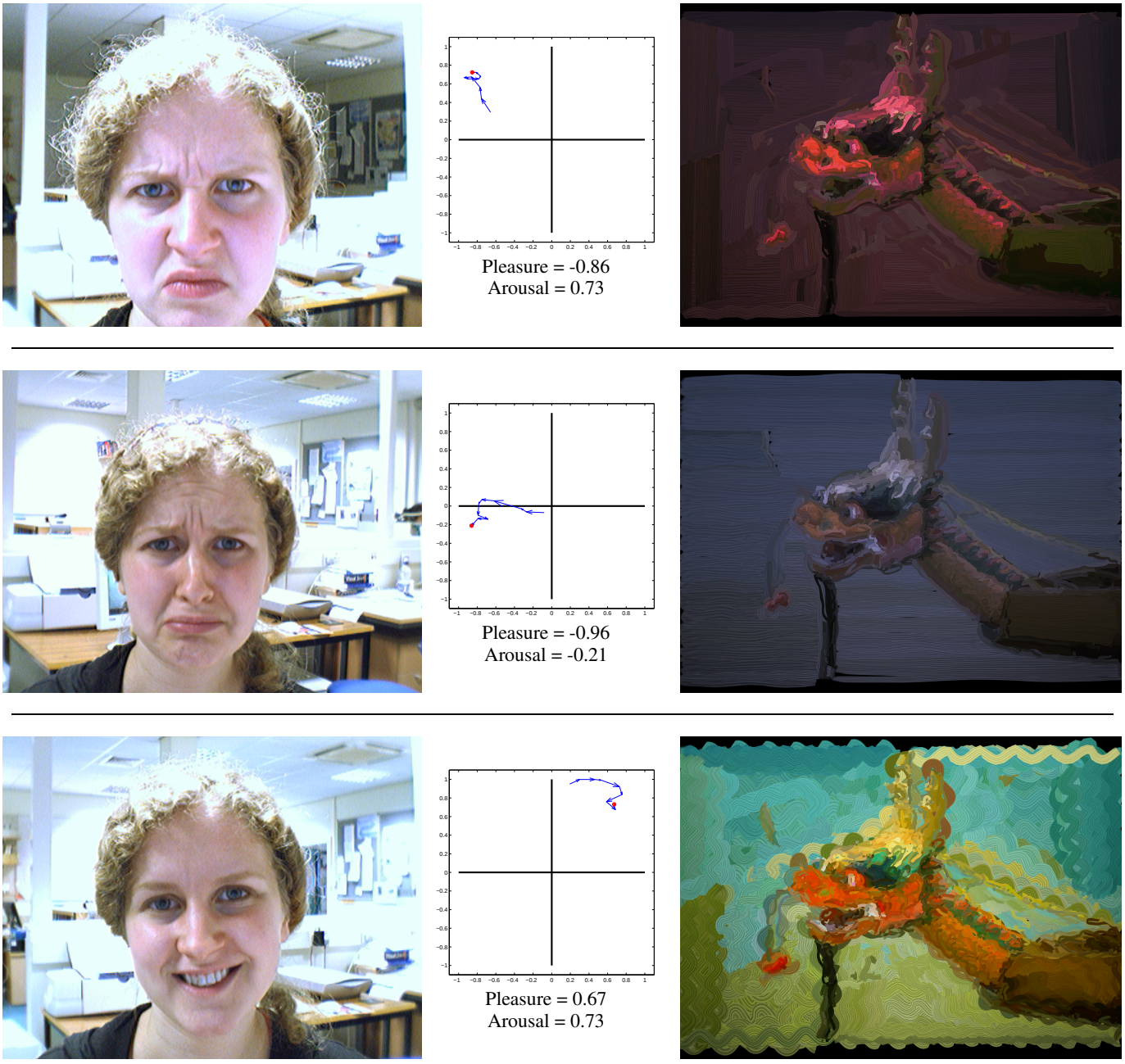


Figure 9: Still frames demonstrating typical operation of the "empathic painting" using the *DRAGON* photograph (Figure 10). Three different facial expressions from top to bottom depicting anger, sadness/despair and jubilation. The measured coordinates in Russell pleasure-arousal state space (centre), arrows indicate trajectory over the previous 10 frames. The resulting painterly renderings (right). Note these paintings have been gamma corrected for printing — high resolution originals are available at **http://www.cs.bath.ac.uk/∼vision/empaint**.

graphics component to visualise the estimated emotional state of the viewer through a non-photorealistic painterly rendering algorithm. We have implemented the system on a standard Pentium 4 2.8GHz machine, with NVidia GeForce 6600 graphics accelerator. Painting initialisation took approximately 30 seconds, and by leveraging commonly available libraries (OpenGL, OpenCV) a maximum frame rate of four frames per second was achieved during interaction. Similar frame rates were found suitable for painterly interaction in [Hertzmann and Perlin 2000].

Figure 9 shows a gallery of facial expressions, the perceived emotional state in Russell's pleasure-arousal space, and the corresponding painterly rendering generated from the *DRAGON* image (Figure 10, top-left). The user is expressing anger in Figure 9 (top), corresponding to high displeasure and moderately high arousal. The painterly output mirrors this through a general red-shift and luminance reduction of colours used in the rendering. Strokes become moderately undulated, and noticeably jagged and chaotic. Rendering parameters were set at $p_{\{1..5\}} = [0.64, (-0.86, 0.73), 0.62, 0.85, 0.58]$. In Figure 9 (centre) the user expresses sadness or despair, reflected in high displeasure and neutral to negative arousal; falling within the third quadrant of the emotional space. The presence of user state in the third quad-

rant generates a colour shift towards the calmer blues and greens, with low arousal also manifesting itself through both low saturation and stroke undulation, as well as a much calmer, less chaotic background. As with anger, the high displeasure score has generated a general darkening of the image. Rendering parameters were set at $p_{\{1..5\}} = [0.33, (-0.96, -0.22), 0.31, 0.52, 0.24]$. In Figure 9 (bottom) we give an example of a cheerful, jubilant expression corresponding to moderately high scores on both arousal and pleasure axes. In this example we see highly agitated and chaotic stroke placement, again reflecting high arousal scores. This output is slightly over-agitated for the current emotional state (indicated by the red circle on the plot), and this can be attributed to recent historical states expressing much higher arousal (indicated by blue vector trail) — the FIR filter imposed smoothing constraints on the NPR style parameters, in this case introducing a short lag to mitigate against temporal incoherence in the animation (this trade-off is controlled by parameter α, see Section 4.2). The cheerful, vivid colour selection is a result of the combination of high pleasure and arousal values within the first quadrant. Rendering parameters were set at $p_{\{1..5\}} = [0.87, (0.67, 0.73), 0.85, 0.47, 0.91]$. A "control" painting, generated from a neutral expression state ($\underline{\mathscr{P}} \approx (0,0)$) is shown in Figure 10.

For the purposes of evaluation we deployed our empathic painting system in a live installation, allowing users to experiment with our interaction method and explore the gamut of painterly styles available. User feedback was broadly positive, both in terms of painting aesthetics and method of interaction. In particular users felt engaged with the system, remarking on the intuitive nature of the interface and that they felt able to easily control the style of the painting to produce their desired results. To attempt to quantify this notion of intuitiveness, a small group of users were taken aside and presented with a series of sliders corresponding to rendering parameters $p_{\{1..6\}}$ and, following instruction on their operation, asked to generate paintings exhibiting particular a emotional atmosphere — for example, a cheerful, energetic painting. Whilst users of the standard system were able to produce such a result in one or two seconds, the same users manipulating the sliders were observed to revert to several cycles of trial and error experimentation; often taking between twenty to thirty seconds to produce paintings of similar style. However we note that the configuration space of the painting was less constrained when using the sliders.

We believe there are a number of interesting avenues for building upon this experimental system. Certainly flexibility of the computer vision and graphics components could be improved. Facial expression recognition is currently trained on the features of a single user, and accuracy varies greatly between participants (most notably with age, and to a lesser extent gender). This often requires recalibration between users to obtain acceptable results. Likewise the graphics component relies on a series of subjective mappings between the emotional space and the style parameters of the painterly algorithm. We have found that the mappings specified in Section 4.2 are appropriate for the majority of users, however one might imagine a scenario where the user would prefer to see a cheerful painting produced in response to a despondent input expression. These mappings might also require modification to take into account cross-cultural variation in the way visual cues are used to depict emotional state; such variation would require reconfiguration of the trajectories suggested in Sections 4.2.1–4.2.3.

We have recently experimented with the addition of a parameter controlling stroke colour jitter (aligned to the arousal axis) to produce paintings in a more abstract style, and also experimented with alternative stroke textures (Figure 11). In future work we hope to explore more rigorously the extension of our system to alternative artistic genres such as pen-and-ink or watercolour. However we do not believe such enhancements necessary to demonstrate that conceptually higher-level parameterisation of NPR algorithms can play a valuable role in intuitive user control of artistic renderings. We look forward to further experimentation in the design of novel user-interaction methods for NPR.

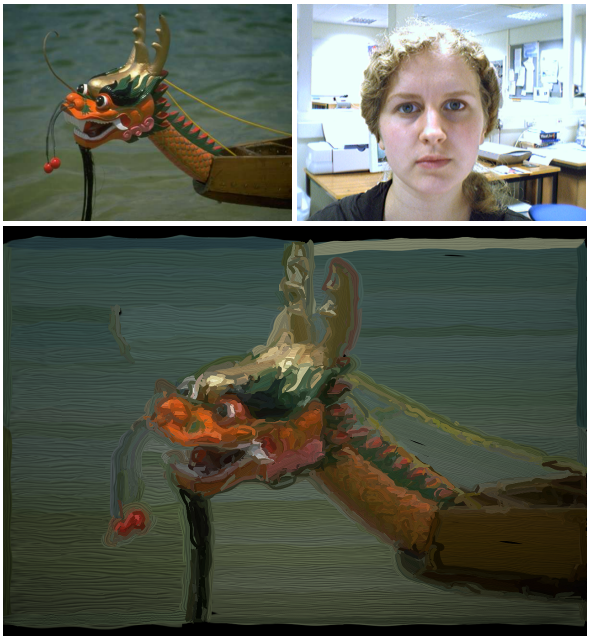


Figure 10: Control data. The source *DRAGON* image (top-left) used to generate paintings in the gallery of Figure 9. A neutral expression corresponding to approximately $\underline{\mathscr{P}} \approx (0,0)$ in the Russell space (top-right) and the resulting neutral painting (bottom).

A video of the system's operation is available at **http://www.cs.bath.ac.uk/∼vision/empaint**.

Acknowledgements

We are grateful to Emmanuel Tanguy and Joanna Bryson for early discussions. This work was supported by an EPSRC/VVG Network summer bursary (GR/T06032/01) and in part by NSF grant IIS-0208876.

References

BLACK, M. J., AND YACOOB, Y. 1997. Recognizing facial expressions in image sequences using local parameterized models of image motion. *Intl. Journal of Comp. Vision (IJCV) 25*, 1, 23–48.

BUCK, I., FINKELSTEIN, A., JACOBS, C., KLEIN, A., SALESIN, D. H., SEIMS, J., SZELISKI, R., AND TOYAMA, K. 2000. Performance-driven hand-drawn animation. In *Proc. ACM NPAR*, 101–018.

BUTLER, A., CLEAVE, C. V., AND STIRLING, S. 1994. *The Art Book*. The Phaidon Press. ISBN: 0714829846.

CHAU, M., AND BETKE, M. 2005. Real time eye tracking and blink detection with USB cameras. Tech. Rep. 2005-12, Boston University Computer Science.

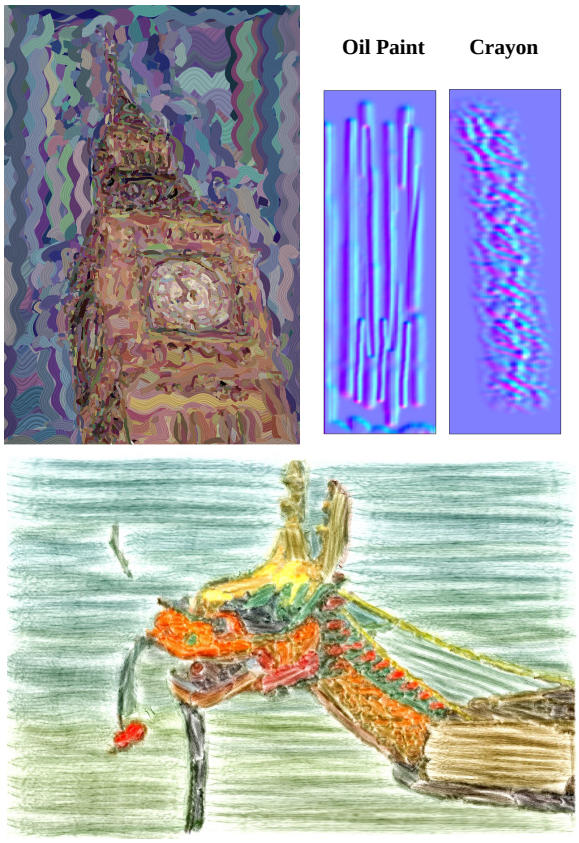


Figure 11: Experiments in style diversification. Introducing a new colour jitter parameter encompasses abstract painting styles (top-left), and substitution of stroke bump map texture (top-right) produces output reminiscient of child-like crayon renderings (bottom).

CHRISTOUDIAS, C., GEORGESCU, B., AND MEER, P. 2002. Synergism in low level vision. In 16th *Intl. Conf. on Pattern Recognition (ICPR)*, vol. 4, 150–155.

COLLOMOSSE, J. P., AND HALL, P. M. 2005. Genetic paint: A search for salient paintings. In *Proc. EvoMUSART (at EuroGP), Springer LNCS*, vol. 3449, 437–447.

COLLOMOSSE, J. P., ROWNTREE, D., AND HALL, P. M. 2005. Stroke surfaces: Temporally coherent artistic animations from video. *IEEE Transactions on Visualization and Comp. Graphics 11*, 5 (Sept.), 540–549.

DECARLO, D., AND SANTELLA, A. 2002. Abstracted painterly renderings using eye-tracking data. In *Proc. ACM SIGGRAPH*, 769–776.

DUKE, D., BARNARD, P., HALPER, N., AND MELLIN, M. 2003. Rendering and affect. In *Proc. Eurographics*, 359–368.

EKMAN, P., AND FRIESEN, W. 1978. *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Consulting Psychologists Press, Palo Alto, CA.

GOOCH, B., COOMBE, G., AND SHIRLEY, P. 2002. Artistic vision: Painterly rendering using computer vision techniques. In *Proc.* 2nd *ACM Sympos. on NPAR*, 83–90.

HAEBERLI, P. 1990. Paint by numbers: abstract image representations. In *Proc. ACM SIGGRAPH*, vol. 4, 207–214.

HAGGERTY, M. 1991. Almost automatic computer painting. *IEEE Computer Graphics and Applications 11*, 6 (Nov.), 11–12.

HALPER, N., MELLIN, M., HERRMANN, C. S., LINNEWEBER, V., AND STROTHOTTE, T. 2003. Towards an understanding of the psychology of non-photorealistic rendering. In *Proc. Workshop on Computational Visualistics*, 67–78.

HAYS, J., AND ESSA, I. 2004. Image and video based painterly animation. In *Proc.* 3rd *ACM Sympos. on NPAR*, 113–120.

HENVER, K. 1935. Experimental studies of the affective value of colors and lines. *Journal of Applied Psychology*, 385–398.

HERTZMANN, A., AND PERLIN, K. 2000. Painterly rendering for video and interaction. In *Proc.* 1st *ACM Sympos. on NPAR*, 7–12.

HERTZMANN, A., JACOBS, C., OLIVER, N., CURLESS, B., AND SALESIN, D. H. 2001. Image analogies. In *Proc. ACM SIGRGAPH*, 327–340.

HERTZMANN, A. 1998. Painterly rendering with curved brush strokes of multiple sizes. In *Proc. ACM SIGGRAPH*, 453–460.

HERTZMANN, A. 2001. Paint by relaxation. In *Proc. Computer Graphics Intl. (CGI)*, 47–54.

HERTZMANN, A. 2002. Fast paint texture. In *Proc.* 2nd *ACM Sympos. on NPAR*, 91–96.

KOLLIOPOULOS, A. 2005. *Image segmentation for stylized non-photorealistic rendering and animation*. Master's thesis, Univ. Toronto.

KOVACS, L., AND SZIRANYI, T. 2002. Creating video animations combining stochastic paintbrush transformation and motion detection. In *Proc.* 16th *Intl. Conference on Pattern Recognition (ICPR)*, vol. II, 1090–1093.

LI, Y., YU, F., XU, Y.-Q., CHANG, E., AND SHUM, H.-Y. 2001. Speech-driven cartoon animation with emotion. In *Proc. ACM Intl. Multimedia Conf.*, 365–371.

LITWINOWICZ, P. 1997. Processing images and video for an impressionist effect. In *Proc. ACM SIGGRAPH*, 407–414.

LUNDHOLM, H. 1921. The affective tone of lines: experimental researches. *The Psychological Review 28*, 60.

MAHNKE, F. 1996. *Color, Environment, and Human Response*. Van Nostrand Reinhold.

PENTLAND, A., MOGHADDAM, B., AND STARNER, T. 1994. View-based and modular eigenspaces for face recognition. In *Proc. Intl. Conf. on Comp. Vision and Pattern Recognition*.

PICKFORD, R. W. 1972. *Psychology and Visual Aesthetics*.

PLUTCHIK, R. 1980. A general psychoevolutionary theory of emotion. In *Emotion: Theory, research, and experience*, R. Plutchik and H. Kellerman, Eds. Academic press, Inc, 3–33.

POFFENBERGER, A. T., AND BARROWS, B. E. 1924. The feeling value of lines. *Journal of Applied Psychology 8*, 187–205.

RUSSEL, J. A., AND FERNÁNDEZ-DOLS, J. M. 1997. *The Psychology of Facial Expression*. Cambridge University Press.

RUSSELL, J. A. 1997. Reading emotion from and into faces: Resurrecting a dimensional-contextual perspective. In Russel and Fernández-Dols [Russel and Fernández-Dols 1997], 295–320.

SANTELLA, A., AND DECARLO, D. 2004. Visual interest and NPR: an evaluation and manifesto. In *Proc.* 3rd *ACM Sympos. on NPAR*, 71–78.

SHIRAISHI, M., AND YAMAGUCHI, Y. 2000. An algorithm for automatic painterly rendering based on local source image approximation. In *Proc.* 1st *ACM Sympos. on NPAR*, 53–58.

SMITH, C., AND SCOTT, H. 1997. A componential approach to the meaning of facial expression. In *The Psychology of Facial Expression*, J. A. Russel and J. M. Fernández-Dols, Eds. Cambridge, 232–249.

TIAN, Y., KANADE, T., AND COHN, J. 2001. Recognizing action units for facial expression analysis. In *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 97–115.

TREAVETT, S., AND CHEN, M. 1997. Statistical techniques for the automated synthesis of non-photorealistic images. In *Proc.* 15th *Eurographics UK Conference*, 201–210.

VIOLA, P., AND JONES, M. 2001. Rapid object detection using a boosted cascade of simple features. In *Proc. Comp. Vision and Pattern Recognition (CVPR)*, vol. 1, 511–5128.

WANG, J., XU, Y., SHUM, H.-Y., AND COHEN, M. 2004. Video tooning. In *Proc. ACM SIGGRAPH*, 574–583.

WRIGHT, B., AND RAINWATER, L. 1962. The meaning of colour. *Journal of General Psychology 67*, 89–99.