Evolutionary Expression of Emotions in Virtual Humans using Lights and Pixels

Celso M. de Melo[1], Ana Paiva[2]

[1] USC, University of Southern California
demelo@usc.edu

[2] IST – Technical University of Lisbon and INESC-ID,
Avenida Prof. Cavaco Silva – Taguspark,
2780-990 Porto Salvo, Portugal
ana.paiva@inesc-id.pt

Abstract Artists express emotions through art. To accomplish this they rely on lines, shapes, textures, color, light, sounds, music, words and the body. The virtual humans field has been neglecting the kind of expression we see in the arts. In fact, researchers tend to focus on gesture, face and voice for the expression of emotions. But why limit ourselves to the body? In this context, drawing on the accumulated knowledge from the arts, this chapter describes an evolutionary model for the expression of emotions in virtual humans using lights, shadows, filters and composition. Lighting expression relies on a local pixel-based model supporting light and shadows parameters regularly manipulated in the visual arts. Screen expression uses filters and composition to manipulate the virtual human's pixels themselves in a way akin to painting. Emotions are synthesized using the OCC model. Finally, to learn mappings between affective states and lighting and screen expression, an evolutionary model which relies on genetic algorithms is proposed. The model generates alternatives using crossover and mutation and selects alternatives based on feedback from artificial and human critics.

1 Introduction

> "The anger which I feel here and now (...) is no doubt an instance of anger (...); but it is much more than mere anger: it is a peculiar anger, not quite like any anger that I ever felt before"

In this passage by Collingwood (1938), the artist is trying to express an emotion. But, this isn't just any emotion. This is a unique emotion. As he tries to make sense of it, he shall express it using lines, shapes, textures, color, light, sound, music, words and the body. The perspective of art as the creative expression of emo-

tions is not new and exists, at least, since the Romanticism (Oatley 2003; Sayre 2007). The idea is that when the artist is confronted with unexpected events, emotions are elicited and a creative response is demanded (Averill et al. 1995). Thus, through the creative expression of its feelings, the artist is trying to understand their peculiarity. But art is not simply an outlet for the artist's emotions. From the perspective of the receiver, through its empathic emotional response to a work of art, it is also seen as a means to learn about the human condition (Elliot 1966; Oatley 2003). Emotions are, therefore, intrinsically related to the value of art.

Affective computing has been neglecting the kind of expression we see in the arts. The state-of-the-art in the related virtual humans field is a case in point. Virtual humans are embodied characters which inhabit virtual worlds and look, think and act like humans (Gratch et al. 2002). Thus far, researchers tended to focus on gesture (Cassell 2000), face (Noh and Neumann 1998) and voice (Schroder 2004) for emotion expression. But, in the digital medium we need not be limited to the body.

In this context, drawing on accumulated knowledge from art theory, this work proposes to go beyond embodiment and synthesize expression of emotions in virtual humans using lights, shadows, composition and filters. This approach, therefore, focuses on two expression channels: lighting and screen. In the first case, the work inspires on the principles of lighting, regularly explored in theatre or film production (Alton 1949; Malkiewicz and Grybosky 1986; Millerson 1999; Birn 2006), to convey the virtual human's affective state through the environment's light sources. In the second case, the work acknowledges that, at the meta level, virtual humans are no more than pixels in the screen which can be manipulated, in a way akin to the visual arts (Birn 2006; Gross 2007; Zettl 2008), to emotions. Now, having defined which expression channels to explore, what remains to be defined is how to map affective states into lighting and screen expression.

This work explores genetic algorithms (GAs) (Mitchell 1999) to learn the mappings between affective states and the expression channels. Genetic algorithms seem appropriate for several reasons. First, there are no available datasets exemplifying what correct expression of emotions using lights or screen is. Thus, standard supervised machine learning algorithms, which rely on a teacher, seem unsuitable. Furthermore, art varies according to time, individual, culture and what has been done before (Sayre 2007). Therefore, the artistic space should be explored in search of creative - i.e., new and aesthetic - expression. Genetic algorithms, defining a guided search, are, thus, appropriate. Second, the virtual humans field is new and novel forms of expression are available. Here, the GAs clear separation between generation and evaluation of alternatives is appropriate. Alternatives, in this new artistic space, can be generated using biologically inspired operators - selection, mutation, crossover, etc. Evaluation, in turn, could rely on fitness functions drawn from art theory. Finally, it has been argued that art is adaptive as it contributes to the survival of the artist (Dissanayake 1987). This, of course, meets the GAs biological motivation.

The remainder of the chapter is organized as follows. Section 2 provides background on virtual humans and describes the digital medium's potential for expression of emotions, focusing on lighting and screen expression. Section 3 describes the virtual human model, detailing the lighting and screen expression channels. Section 4 describes the evolutionary approach which maps affective states into the expression channels. Section 5 describes some of the results. Finally, section 6 draws some conclusions and discusses future work.

2 Background

2.1 Expression in the Arts

There are several conceptions about what expression in the arts is. First, it relates to beauty as the expression of beauty in nature (Batteux 1969). Second, it relates to culture as the expression of the values of any given society (Geertz 1976). Third, it relates to individuality as the expression of the artists' liberties and values (Kant 1951). Finally, it relates to emotions as the expression of the artists' feelings. In fact, many acknowledge the importance of emotions for appreciating and attributing value to the arts. From the perspective of the creator, expression is seen as a way of understanding and coming to terms with what he is experiencing affectively (Collingwood 1938). From the perspective of the receiver, through its empathetic emotional responses to a work of art, it is seen as means to learn about the human condition (Elliot 1966; Oatley 2003). Finally, Artistic expression is a creative endeavor (Kant 1951; Gombrich 1960; Batteux, 1969; Sayre 2007). Art is not a craft were artists can simply follow a set of rules to reach a result (Collingwood 1938) and, according to the Romanticism's view, art is the creative expression of latent affective states (Oatley 2003). In fact, Gombrich (1960) argues that this idiosyncrasy is inescapable as the artist's visual perceptions are necessarily confronted with its mental schemas, including ideas and preconceptions. This work tries to explore the kind of expression we see in the arts in the context of virtual humans. Furthermore, a simple rule-based approach is avoided and, instead, a machine-learning approach, which is more likely to be able to adapt to dynamic artistic values, is pursued to learn mappings between emotional states and light and screen expression.

2.2 Expression of Emotions in the Digital Medium

Digital technology is a flexible medium for the expression of emotions. In virtual worlds, inhabited by virtual humans, besides embodiment, at least four expression channels can be identified: camera, lights, sound, and screen. The camera defines the view into the virtual world. Expressive control, which inspires on cinema and photography, is achieved through selection of shot, shot transitions, shot framing and manipulation of lens properties (Arijon 1976; Katz 1991; Block 2001; Malkiewicz and Mullen 2005). Lights define which areas of the scene are illuminated and which are in shadow. Furthermore, lights define the color in the scene. Expressive control, which inspires in the visual arts, is achieved through manipulation of (Alton 1949; Malkiewicz and Grybosky 1986; Millerson 1999; Birn 2006): light type, placement and angle; shadow softness and falloff; color properties such as hue, brightness and saturation. Sound refers to literal sounds (e.g., dialogues), non-literal sounds (e.g., effects) and music. Expressive control, which inspires in drama and music, is achieved through selection of appropriate content for each kind of sound (Juslin and Sloboda 2001; Zettl 2008). Finally, the screen is a meta channel referring to the pixel-based screen itself. Expression control, which inspires on cinema and photography, is achieved through manipulation of pixel properties such as depth and color (Birn 2006; Zettl 2008). This work focuses on lighting and screen expression.

2.3 Expression of Emotions in Virtual Humans

Virtual humans are embodied characters which inhabit virtual worlds (Gratch et al. 2002). First, virtual humans look like humans. Thus, research draws on computer graphics for models to control the body and face. Second, virtual humans think and act like humans. Thus, research draws on the social sciences for models to produce synchronized verbal and nonverbal communication as well as convey emotions and personality. Emotion synthesis usually resorts to cognitive appraisal theories of emotion, being the Ortony, Clore and Collins (OCC) theory (Ortony et al. 1988) one of the most commonly used. Emotion expression tends to focus on conveying emotions through synchronized and integrated gesture (Cassell 2000), facial (Noh and Neumann 1998) and vocal (Schroder 2004) expression. In contrast, this work goes beyond the body using lights, shadows, composition and filters to express emotions.

A different line of research explores *motion modifiers* which add emotive qualities to neutral expression. Amaya (Amaya et al. 1996) uses signal processing techniques to capture the difference between neutral and emotional movement which would, then, be used to confer emotive properties to other motion data. Chi and colleagues (Chi et al. 2000) propose a system which adds expressiveness to exis-

tent motion data based on the effort and shape parameters of a dance movement observation technique called Laban Movement Analysis. Hartmann (Hartmann et al. 2005) draws from psychology six parameters for gesture modification: overall activation, spatial extent, temporal extent, fluidity, power and repetition. Finally, de Melo (de Melo and Paiva 2005) proposes a model for expression of emotions using the camera, light and sound expression channels. However, this model did not focus on virtual humans, used a less sophisticated light channel than the one proposed here, did not explore screen expression and used simple rules instead of an evolutionary approach.

2.4 Expression of Emotions using Light

This work explores *lighting* to express virtual humans' emotions. Lighting is the deliberate control of light to achieve expressive goals. Lighting can be used for the purpose of expression of emotions and aesthetics (Alton 1949; Malkiewicz and Grybosky 1986; Millerson 1999; Birn 2006). To achieve these goals, the following elements are manipulated (Millerson 1999; Birn 2006): (a) type, which defines whether the light is a point, directional or spotlight; (b) direction, which defines the angle. Illumination at eye-level or above is neutral, whereas below eye-level is unnatural, bizarre or scary; (c) color, which defines color properties. Color definition based on hue, saturation and brightness (Hunt 2004) is convenient as these are, in Western culture, regularly manipulated to convey emotions (Fraser 2004); (d) intensity, which defines exposure level; (e) softness, which defines how hard or soft the light is. Hard light, with crisp shadows, confers a harsh, mysterious, environment. Soft light, with soft transparent shadows, confers a happy, smooth, untextured environment; (f) decay, which defines how light decays with distance; (g) throw pattern, which defines the shape of the light. Shadows occur in the absence of light. Though strictly related to lights, they tend to be independently controlled by artists. Shadows can also be used to express emotions and aesthetics (Alton 1949; Malkiewicz and Grybosky 1986; Millerson 1999; Birn 2006) through manipulation of the following elements: (a) softness, which defines how sharp and transparent the shadow is. The denser the shadow, the more dramatic it is; (b) size, which defines the shadow size. Big shadows confer the impression of an ominous, dramatic character. Small shadows confer the opposite impression. Lighting transitions change the elements of light and shadow in time. Transitions can be used to express the mood of the scene (Millerson 1999; Birn 2006). Digital lighting provides more expressive control to the artist as, besides giving free control of all light and shadow elements, the image's pixels can now be manipulated in a way akin to painting.

2.5 Expression of Emotions using Pixels

At a meta level, virtual humans and virtual worlds can be seen as pixels in a screen. Thus, as in painting, photography or cinema, it is possible to manipulate the image itself for expressive reasons. In this view, this work explores composition and filtering for the expression of emotions. Composition refers to the process of arranging different aspects of the objects in the scene into layers which are then manipulated and combined to form the final image (Birn 2006). Here, aspects refer to the ambient, diffuse, specular, shadow, alpha or depth object components. Composition has two main advantages: increases efficiency as different aspects can be held fixed for several frames; and, increases expressiveness as each aspect can be controlled independently. Composition is a standard technique in film production. Filtering is a technique where the scene is rendered into a temporary texture which is then manipulated using *shaders* before being presented to the user (Zettl 2008). Shaders replace parts of the traditional pipeline with programmable units (Moller and Haines 2002). Vertex shaders modify vertex data such as position, normal and texture coordinates. Pixel shaders modify pixel data such as color and depth. Filtering has many advantages: has constant performance independently of scene complexity; can be very expressive due to the variety of available filters (St-Laurent 2004); and, is scalable as several filters can be concatenated.

2.6 Evolutionary Approaches

We are not aware of any prior attempt to use evolutionary algorithms to express emotions in virtual humans. Nevertheless, they have been widely explored in computer graphics. Karl Sims (Sims 1991) explores a genetic programming approach using symbolic lisp expressions to generate images, solid textures and animations. The artist Steven Rooke (World 1996) uses a set of low and high-level primitives to guide his genetic programming approach to generate images within his own style. Contrasting to the previous approaches, genetic algorithms have been used to evolve shaders (Lewis 2001), fractals (Angeline 1996), animations (Ventrella 1995) and complex three-dimensional objects (Todd and Latham 1992). In all previous systems, the user interactively guides the evolution process. However, attempts have been made to automate this process. Representative is the NEvAr system (Machado 2006) which proposes an artificial critic which, first, extracts features from the images in the population and, then, applies a neural network, trained with appropriate examples, to select the fittest. This project explores both the interactive and artificial critic approaches. In the latter case, the critic is based on rules from art theory.

3 The Model

The virtual human model is summarized in Fig.1. The virtual human itself is structured according to a three-layer architecture (Blumberg and Galyean 1995; Perlin and Goldberg 1996). The *geometry layer* defines a 54-bone human-based skeleton which deforms the skin. The skin, in turn, is divided into body groups – head, torso, arms, hands and legs. The *animation layer* defines keyframe and procedural animation mechanisms. The *behavior layer* defines speech and gesticulation expression and supports a language for multimodal expression control. Finally, several expression modalities are built on top of this layered architecture. Bodily expression manipulates face, postures and gestures. Further details on bodily expression, which will not be addressed here, can be found in (de Melo and Paiva 2006a; de Melo and Paiva, 2006b). Lighting expression explores the surrounding environment and manipulates lights and shadows. Screen expression manipulates the virtual human pixels themselves. These expression channels are detailed next.

3.1 Lighting Expression

Lighting expression relies on a local pixel-based lighting model. The model supports multiple sources, three light types and shadows using the shadow map technique (Moller and Haines 2002). The detailed equations for the lighting model can be found in (de Melo and Paiva 2007). Manipulation of light and shadow elements (subsection 2.4) is based on the following parameters: (a) *type*, which defines whether to use a directional, point or spotlight; (b) *direction* and *position*, which, according to type, control the light angle; (c) *ambient*, *diffuse* and *specular colors*, which define the color of each of the light's components in either RGB (red, green, blue) or HSB (hue, saturation and brightness) space; (d) *ambient*, *diffuse* and *specular intensity*, which define the intensity of each of the components' color. Setting intensity to 0 disables the component; (e) *attenuation*, *attnPower*, *attnMin*, *attnMax*, which simulate light falloff. Falloff is defined as $attenuation^{-attnPower}$ and is 0 if the distance is less than *attnMin* and 1 beyond a distance of *attnMax*; (f) *throw pattern*, which constraints the light to a texture using component-wise multiplication; (g) *shadow color*, which defines the shadow color. If set to grays, shadows become transparent; if set to white, shadows are disabled; (h) *shadow softness*, which defines the falloff between light and shadow areas. Finally, sophisticated lighting transitions, such as accelerations and decelerations, are based on parametric cubic curve interpolation of parameters (Moller and Haines 2002).

3.2 Screen Expression

Screen expression explores composition and filtering. Filtering consists of rendering the scene to a temporary texture, modifying it using shaders and, then, presenting it to the user. Several filters have been explored in the literature (St-Laurent 2004) and this work explores some of them: (a) the *contrast filter*, Fig.2-(b), which controls virtual human contrast and can be used to simulate exposure effects; (b) the *motion blur filter*, Fig.2-(c), which simulates motion blur and is usually used in film to convey nervousness; (c) the *style filter*, Fig.2-(d), which manipulates the virtual human's color properties to convey a stylized look; (d) the *HSB filter*, which controls the virtual human hue, saturation and brightness. Filters can be concatenated to create compound effects and its parameters interpolated using parametric cubic curve interpolation (Moller and Haines, 2002).

Composition refers to the process of (Birn 2006): arranging different aspects of the objects in the scene into layers; independently manipulating the layers for expressive reasons; combining the layers to form the final image. A layer is characterized as follows: (a) is associated with a subset of the objects which are rendered when the layer is rendered. These subsets need not be mutually exclusive; (b) can be rendered to a texture or the backbuffer. If rendered to a texture, filtering can be applied; (c) has an ordered list of filters which are successively applied to the objects. Only applies if the layer is being rendered to a texture; (d) is associated with a subset of the lights in the scene. Objects in the layer are only affected by these lights; (e) defines a lighting mask, which defines which components of the associated lights apply to the objects; (f) can render only a subset of the virtual human's skin body groups. Finally, layer combination is defined by order and blending operation. The former defines the order in which layers are rendered into the backbuffer. The latter defines how are the pixels to be combined.

3.3 Synthesis of Emotions

Virtual human emotion synthesis is based on the Ortony, Clore and Collins (OCC) model (Ortony et al. 1988). All 22 emotion types, local and global variables are implemented. Furthermore, emotion decay, reinforcement, arousal and mood are also considered. Emotion decay is, as suggested by Picard (1997), represented by an inverse exponential function. Emotion reinforcement is, so as to simulate the saturation effect (Picard 1997), represented by a logarithmic function. Arousal, which relates to the physiological manifestation of emotions, is characterized as follows: is positive; decays linearly in time; reinforces with emotion eliciting; and, increases the elicited emotions' potential. Mood, which refers to the longer-term effects of emotions, is characterized as follows: can be negative or positive; con-

verges to zero linearly in time; reinforces with emotion eliciting; if positive, increases the elicited emotions' potential, if negative, decreases it. Further details about this model can be found in (de Melo and Paiva 2005).

3.4 Expression of Emotions

A markup language, called *Expression Markup Language (EML)*, is used to control multimodal expression. The language supports arbitrary mappings of emotional state conditions and synchronized body, light and screen expression. The language is structured into modules. The core module defines the main elements. The time and synchronization module defines multimodal synchronization mechanisms based on the W3C's SMIL 2.0 specification[1]. The body, gesticulation, voice and face modules control bodily expression. The light module controls light expression, supporting modification of light parameters according to specific transition conditions. The screen module controls screen expression, supporting modification of the composition layers' filter lists. Finally, the emotion module supports emotion synthesis and emotion expression. Regarding emotion synthesis, any of the OCC emotion types can be elicited. Regarding emotion expression, the module supports the specification of rules of the form:

$$\{emotionConditions\}* \rightarrow \{bodyAc \mid lightAc \mid screenAc \mid emotionAc\}*$$

where: emotional conditions – *emotionConditions* – evaluate mood, arousal or active emotions' intensity or valence; expressive actions – *bodyAc*, *lightAc* and *screenAc* – refer to body, light or screen actions as defined by its respective modules; and, emotion actions – *emotionAc* – elicit further emotions.

Even though convenient, the definition of rules is unlikely to capture the way artistic expression works (subsection 2.1). There are, in fact, several rules and guidelines for effective artistic expression available in the literature. However, the expression of emotions in the arts is essentially a creative endeavor and artists are known to break these rules regularly (Sayre 2007). Thus, a better approach should rely on machine learning theory, which would support automatic learning of new rules and more sophisticated mappings between emotional states and bodily, environment and screen expression. In the next section an approach for learning such mappings which is based on genetic algorithms is described.

[1] SMIL: Synchronized Multimedia Integration Language (SMIL). http://www.w3.org/AudioVideo/

4 Evolutionary Expression of Emotions

In this section an evolutionary model which learns mappings between affective states and multimodal expression is presented. The model revolves around two key entities: the *virtual human* and the *critic ensemble*. The virtual human tries to evolve the best way to express some affective state. For every possible state, it begins by generating a random set of *hypotheses*, which constitute a *population*. The population evolves resorting to a genetic algorithm under the influence of feedback from the critic ensemble. The ensemble is composed of *human* and *artificial critics*. The set of evolving populations (one per affective state) are kept on the working memory. The genetic algorithm only operates on populations in *working memory*. These can be saved persistently in the *long-term memory*. Furthermore, high fitness hypotheses (not necessarily from the same population) are saved in the long-term memory's *gallery*. Hypotheses from the gallery can, then, provide elements to the initial population thus, promoting high fitness hypotheses evolution. This model is summarized in Fig.3 and detailed in the following sections.

4.1 Genetic Algorithm

At the core of the model lies a standard implementation of the genetic algorithm (Mitchell 1999). The algorithm's inputs are:

- the *critic ensemble* for ranking candidate hypotheses;
- *stopping criteria* to end the algorithm. This can be a number of iterations, a fitness threshold or both;
- the *size of the population*, *p*, to be maintained;
- the *selection method*, *SM*, to select probabilistically among the hypotheses in a population. Two methods can be used: *roulette wheel*, which selects a hypothesis according to the ratio of its fitness to the sum of all hypotheses fitness, see (1); *tournament selection*, which selects with some probability p′ the most fit of two hypotheses selected according to (1). Tournament selection often yields a more diverse population than roulette wheel (Mitchell 1999);

$$\Pr(h_i) = \frac{fitness(h_i)}{\sum_{j=1}^{p} fitness(h_j)} \tag{1}$$

- *r*, the *crossover rate*;
- *m*, the *mutation rate*;
- *e*, the *elitism rate*.

The algorithm begins by setting up the initial population. This can be generated at random or loaded from long-term memory (subsection 4.4). Thereafter, the al-

gorithm enters an infinite loop, evolving populations, until the stopping criterion is met. At each iteration, first, *(1−r)p* percent of the population is selected for the next generation; second, *rp/2* pairs of hypotheses are selected for crossover and the offspring are added to the next generation; third, *m* percent of the population are randomly mutated; fourth, *e* percent of hypotheses from the population are copied unchanged to the next generation. The rationale behind elitism is to avoid losing the best hypotheses when a new generation is evolved (Mitchell 1999). Evaluation, throughout, is based on feedback from the critic ensemble (subsection 4.5). The algorithm is summarized as follows:

```
GA(criticEnsemble, stoppingCriteria, p, sel, r, m, e)
// criticEnsemble: A group of fitness functions
// stoppingCriteria: Criteria to end the algorithm
// (no. of iterations or threshold)
// p: The number of hypothesis per population
// SM: The selection method (roulette wheel or tournament selection)
// r: The crossover rate
// m: The mutation rate
// e: The elitism rate
{
 Initialize population: P := Generate p hypotheses
    at random or load from LTM
 Evaluate: For each h in P, compute fitness(h)

 while(!stop)
 {
  Create a new Generation, Ps
  Select prob. according to SM, (1-r)p members of P to add to Ps
  Crossover rp/2 prob. selected pairs and add offspring to Ps
  Mutate m percent of Ps
  Select probabilistically e percent of P and replace in Ps
  Update P := Ps
  Evaluate, according to criticEnsemble, each h in P
 }
}
```

4.2 Hypotheses

The hypothesis encoding is structured according to expression modalities, see Fig.4. At the top level the virtual human hypothesis is subdivided into the lighting, screen and body hypotheses. The lighting hypothesis refers to a *three-point lighting* configuration (Millerson 1999; Birn 2006). This technique is composed of the following light roles: (a) *key light*, which is the main source of light; (b) *fill light*, which is a low-intensity light that fills an area that is otherwise too dark; (c) *back light*, which is used to separate the character from the background. In this case, only the key and fill lights are used, as these define the main illumination in the scene (Millerson 1999) and the back light can be computationally expensive (Birn 2006). Both lights are modeled as directional lights and only the key light is set to

cast shadows, according to standard lighting practice (Birn 2006). Only a subset of the parameters, defined in subsection 3.1, is evolved: (a) *direction*, which corresponds to a bidimensional float vector corresponding to angles about the *x* and *y* axis w.r.t. the camera-character direction. The angles are kept in the range [-75.0; 75.0] as these correspond to good illumination angles (Millerson, 1999); (b) *diffuse color*, which corresponds to a RGB vector; (c) K_d, which defines the diffuse color intensity; (d) K_s, which defines the specular color intensity; (e) *shadow opacity*, which defines how transparent the shadow is. Finally, the fill light parameters are similar to the key light's, except that all shadow parameters are ignored. The screen hypothesis is structured according to the virtual human's skin body groups. For each body group, a sequence of filters is applied. Filters were defined in subsection 3.2. For each filter, a field is defined for each of its parameters, including whether it is active. Filters are applied to the body groups in the order they appear. Notice order is subject to change through the crossover operation.

4.3 Operators

This work makes use of two genetic operators, Fig.5: *crossover* and *mutation*. Crossover takes two parent hypotheses from the current generation and creates two offspring by recombining portions of the parents. Recombination is parameter-wise. The parent hypotheses are chosen probabilistically from the current population. Thus, the idea is to combine highly fit parents to try to generate offspring with even higher fitness. The percentage of the current population which is subjected to crossover is defined by the crossover rate, *r*. Mutation exists to provide a continuous source of variation in the population. This operator essentially randomizes the values of a random number of the hypothesis' parameters. The operator is constrained to generate within-domain values for each parameter. The percentage of the current population which is subjected to mutation is defined by the mutation rate, *m*.

4.4 Working and Long-Term Memory

Ultimately, the evolutionary model tries to learn multiple mappings relating multimodal expression and affective states. Therefore, the virtual human needs to keep track of several populations, one per affective state, even though only a single one is evolving at any instant of time. The working memory keeps the current state of evolving populations. In real life, creating an artistic product may take a long time (Sayre 2007). Therefore, to accommodate this characteristic, the whole set of evolving populations can be saved, at any time, in long-term memory. Im-

plementation-wise this corresponds to saving all information about the population in XML format. Furthermore, the interaction between working and long-term memory provide the foundations for life-long learning. Finally, a gallery is saved in long-term memory to accommodate the higher fitness hypotheses, independently of the population they originated from. The gallery can be used afterwards to feed some hypotheses to the initial population, thus, promoting rapid generation of highly fit hypotheses. More importantly, the gallery is a dataset which could be used to learn models using supervised learning. This use of the gallery is not currently implemented and is further addressed in the future work section.

4.5 Critic Ensemble

The critic ensemble defines several *critics* per affective state. Critics can be artificial, in which case fitness is inspired on art theory, or human, in which case fitness reflects the subjective opinion of the critic. An artificial critic consists of a set of *population fitness functions* and a set of *hypothesis fitness functions*. A population fitness function calculates the hypothesis' fitness with respect to the others in the population. A hypothesis fitness function assigns an absolute fitness to each hypothesis, independently of the others. Both kinds of fitness function are normalized to lie in the range [0.0;1.0]. As the critic may define several functions, weights are used to differentiate their relative importance. Thus, for each kind, the set fitness is the weighted average among all constituent functions, as in (2). Finally, the final fitness is the weighted combination of the population and hypothesis functions sets, as in (3). Section 5 presents one example of an artificial critic.

$$f^{set} = \frac{w_i * f_i}{\sum_i w_i} \tag{2}$$

$$f^{final} = W_{pop} * f^{set}_{pop} + W_{hyp} * f^{set}_{hyp} \tag{3}$$

The model supports interactive evolution where humans can influence the selection process by assigning subjective fitness to the hypotheses. There are several advantages to bringing humans into the evaluation process (Sayre 2007): (a) art literature is far from being able to fully explain what is valued in the arts; (b) art is dynamic and values different things at different times. Furthermore, bringing humans into the evaluation process accommodates individual, social and cultural differences in expression of emotions (Keltner et al. 2003; Mesquita 2003). Furthermore, as discussed in future work, if the fitness functions are unknown, then the model might be made to learn from human experts. Two disadvantages are that humans may reduce variety in the population, causing convergence to a specific

style, and that the evolution process becomes much slower. For these reasons, the human fitness function (as well as any other) may be selectively deactivated.

5 Results

This work proposes a model which can evolve expression of emotions in virtual humans using lighting and screen expression. In this section this model is used to learn a mapping between the emotion 'anger' and lighting expression. Having described the lighting expression hypothesis encoding (subsection 4.2), what is needed is to define the set of critics which will guide the selection process. In this example, human critics are ignored and an artificial critic is defined. To build an artificial critic it is necessary to define an appropriate set of population and hypothesis fitness functions which reflect the expression of anger as well as their weights (subsection 4.5). However, to demonstrate the flexibility of the current approach, besides aiming at effective expression of anger, we'll also add fitness functions which reflect lighting aesthetics. In both cases, these fitness functions shall try to reflect guidelines from art theory regarding the expression of anger, effectively and aesthetically, through lighting. In some cases, these functions might be conflicting, which is perfectly acceptable in art production (Sayre, 2007). Conflicts are handled by assigning appropriate weights to the functions.

Six hypothesis fitness functions and one population fitness function are proposed. The hypotheses set weight is set to 0.75 and the population set weight to 0.25. The fitness functions are as follows:

- The *red color function*, with weight 4.0, assigns higher fitness the smaller the Euclidean distance between the light's diffuse color to pure red, as in (4). This function is applied both to the key and fill lights. Red was chosen because, in Western culture, red tends to be associated with excitation or nervousness (Fraser and Banks 2004);

$$f = \frac{1}{dist + 1} \tag{4}$$

- The *low-angle illumination function*, with weight 1.5, assigns higher fitness the closer the Euclidean distance between the key light's angle about the x axis to 20^{o}. The rationale is that illumination from below is unnatural, bizarre and frightening (Millerson 1999);
- The *opaque shadow function*, with weight 1.0, assigns higher fitness the closer the key light's shadow opacity parameter is to 0. This favors darker shadows. The rationale is that hard, crisp shadows convey a mysterious and harsh character (Millerson 1999; Birn 2006);

- The *low complexity function*, with weight 0.5, assigns higher fitness to less complex hypotheses. The rationale is that humans naturally value artifacts which can express many things in a simple manner (Machado 2006). What constitutes low complexity in lighting is hard to define but, here, will define a low complexity hypothesis as having: diffuse color equal to a grayscale value (i.e. with equal R,G,B components); K_d equal to 2.0, giving a neutral diffuse component; and K_s equal to 0.0, giving a null specular component;
- The *key high-angle* function, with weight 0.5, assigns higher fitness the closer the Euclidean distance between the key light's angle about the x axis to $\pm 30^o$. This is a standard guideline for good illumination (Millerson 1999). Notice, first, this is an aesthetics function and, second, it contradicts the low-angle illumination function;
- The *key-fill symmetry* function, with weight 0.5, which assigns higher fitness if the fill light angles are symmetrical to the key light's. This is also a standard guideline for good illumination (Millerson 1999);
- The *novelty* function, with weight 0.5, which assigns higher fitness the more novel the hypothesis is w.r.t. the rest of the population. This is, therefore, a population fitness function. The rationale in this case is that novelty is usually appreciated in the arts (Sayre, 2007). Notice also that this function puts some internal pressure on the population, forcing the hypotheses to keep changing to differ from each other.

5.1 Parameters Selection

Having defined the critics, before actually running the genetic algorithm, it is necessary to configure the following parameters: p, the population size; r, the crossover rate; m, the mutation rate; e, the elitism rate; and, SM, the selection method. As the value chosen for these parameters influences the speed and efficacy of the evolution, it is important to choose optimal values. In this sense, we began by running the genetic algorithm, for a small number of iterations, with typical values of these parameters (Mitchell, 1999): p in {25,50}; r in {0.6,0.7}; m in {0.0,0.1,0.2}; e in {0.0,0.1}; and, SM in {roulette wheel, tournament selection}. This amounts to 48 different configurations. Each configuration was ran for 10 iterations and ranked according to the best *population value* among all iterations. Population value is defined as follows:

$$value = 2 * HF + 2 * AF - HD - G \tag{5}$$

where HF is the highest fitness hypothesis, AF is the average fitness, HD is the highest hypothesis dominance, which refers to the ratio of number of occurrences of a certain hypothesis to population size, and G is the ratio between generation

number and maximum number of generations. Thus, population value favors populations with highly fit hypotheses, low dominance and from earlier generations. The top 20 parameter configurations are shown in Table 1. Looking at the results, it is possible to observe that the optimal configuration is: $p = 50$, $r = 0.70$, $m = 0.00$, $e = 0.10$ and $SM = tournamentSelection$. Notice that, in this case, the mutation rate is 0 and the 19 best results use tournament selection.

5.2 Evolution Analysis

Having determined the optimal parameters for the proposed fitness functions, the genetic algorithm was run for 50 iterations. A graph showing the evolution of the population value, average fitness and highest fitness is shown in Fig.6. Looking at the graph, it is possible to see that the value and fitness increase quickly up until approximately the 25th iteration. After this point, the highest fitness is already close to 1.0, however, it is still slowly improving. The actual values for the top 20 generations, with respect to population value, are shown in Table 2. Looking at the results it is clear that the populations with the best value occur mostly in later iterations. In fact, the 44th generation was evaluated as having the best value of 2.3324, with its best hypothesis having a fitness of 0.9005. Table 3 details the top 20 hypotheses, with respect to fitness, for the 44th population.

Visual inspection of the evolving hypotheses shows that the hypotheses were accurately reflecting the fitness functions. This evolution is clearly shown in Fig.7, which shows five of the initial populations' hypotheses as well as five of the hypotheses in the 44th generation. The first population, which was randomly generated, has high variance and most hypotheses have relatively low fitness. Evolution, which in this case relied mostly on the crossover operation, successively combined hypotheses which best reflected the fitness functions. In the 44th generation, which had the best population value, most hypotheses illuminate the character with a red color, from below and with opaque shadows. Even so, the hypotheses still differ from each other, perhaps, reflecting the novelty function.

The highest fit hypothesis belongs to the 48th generation and had a fitness of 0.9181. Visually it is similar to the one at the bottom right in Fig.7. The exact hypothesis' parameters are: (a) *diffuse color* (RGB) = (0.99, 0.08, 0.234); (b) *direction* (XY) = (−43.20°, 25.65°); (c) K_d= 2.32; (d) K_s = 1.30; (e) *shadow opacity* = 0.54. However, the importance of the best hypothesis should be deemphasized in favour of a set of highly fit hypotheses. This set is likely to be more useful because, aside from reflecting the emotion in virtue of high fitness, it allows for variety in conveying the emotion. This set could be constructed using the gallery in the long-term memory (subsection 4.4).

Overall, the model seems to be able to evolve appropriate population of hypotheses which reflect the fitness functions. However, the success of the mappings hinges on the quality of the critics' feedback. If it is clear that the final hypotheses

in Fig.7 reflect the fitness functions, it is not that these functions actually reflect anger. The selection of appropriate fitness functions as well as accommodating feedback from human critics is the subject of our future work.

6 Conclusions and Future Work

Drawing on accumulated knowledge from the arts, this work proposes a virtual human model for the expression of emotions which goes beyond the body and uses lights, shadows, composition and filters. Lighting expression relies on a pixel-based lighting model which provides many of the light and shadow parameters regularly used in the visual arts. Screen expression explores filtering and composition. Filtering consists of rendering the scene to a temporary texture, manipulating it using shaders and, then, presenting it to the user. Filters can be concatenated to generate a combined effect. In composition, aspects of the scene are separated into layers, which are independently manipulated, before being combined to generate the final image. Regarding emotion synthesis, the Ortony, Clore and Collins emotion model is integrated.

To learn mappings from affective states to multimodal expression, an evolutionary approach is proposed based on genetic algorithms. Hypotheses encode a three-point lighting configuration and a set of filters which are to be applied to the virtual human's skin body groups. The initial population is either created randomly or loaded from long-term memory into working memory. One population is maintained for each affective state. Generation of alternatives is achieved through the crossover and mutation operators. Selection of alternatives is supervised by a critic ensemble composed of both human and artificial critics. The latter consist of a set of fitness functions which should be inspired on art theory. Several parameters influence the genetic algorithm search: population size, crossover rate, mutation rate, elitism rate, selection method, fitness functions and fitness function weights. Finally, a gallery is maintained with very high fitness hypotheses. These can, later, be fed into the current population to promote the generation of new highly fit offspring.

This work has demonstrated the evolutionary model for the case of learning how to express anger using lighting expression. By visual inspection it is clear that the evolution is reflecting the fitness functions. However, the following questions arise: Do the fitness functions reflect the intended affective states? What about the fitness functions' weights? These issues need to be addressed in the near future. First, the results should be confronted with people's intuitions about the expression of emotions, perhaps in the form of inquiries. Furthermore, it seems clear that the art literature is insufficient to provide a comprehensive set of fitness functions. Thus, second, human critics should be brought into the evolution loop. Both regular people and artists should be involved. In this setting, the system could be made to learn new fitness functions from human feedback and the existent fitness func-

tions' weights could be updated, perhaps using a reinforcement learning mechanism. It would, of course, be interesting to expand the mappings to the six basic emotions (Ekman 1999) – anger, disgust, fear, joy, sadness and surprise – and, furthermore, explore more complex affective states. The gallery could also be used to feed supervised learning algorithms to generate models which explain highly fit hypotheses. These models could, then, feed a self-critic which would, in tandem with the artificial and human critics, influence the selection process. Finally, an obvious extension to this work is exploring the camera and sound expression channels of which much knowledge already exists in the arts (Sayre 2007).

References

Alton J (1949) Painting with Light. Macmillan Co., New York, USA

Amaya K, Bruderlin A, Calvert T (1996) Emotion from motion. Proceedings of Graphics Interface'96, 222-229

Angeline P (1996) Evolving fractal movies. Proceedings of the First Annual Conference in Genetic Programming, 503-511

Arijon D (1976) Grammar of Film Language. Hastings House, New York, USA

Averill J, Nunley E, Tassinary L (1995) Voyages of the Heart: Living an Emotionally Creative Life. Contemporary Psychology, 40(6):530

Batteux C (1969) Les Beaux Arts Réduits à un meme Principe. Slatkine Reprints, Genève, Switzerland

Birn J (2006) [digital] Lighting and Rendering – 2nd edn. New Riders, London, UK

Block B (2001) The Visual Story: Seeing the Structure of Film, TV, and New Media. Focal Press, Boston, USA

Blumberg B, Galyean T (1995) Multi-Level Direction of Autonomous Creatures for Real-Time Virtual Environments. Proceedings of SIGGRAPH'95, 30(3): 173-182

Cassell J (2000) Nudge, nudge, wink, wink: Elements of face-to-face conversation for embodied conversational agents. In: Cassell J, Sullivan J and Churchill E, (eds) Embodied Conversational Agents, 1-27, MIT Press, Massachusetts, USA

Chi D, Costa M, Zhao L et al (2000) The EMOTE model for effort and shape. Proceedings of SIGGRAPH 2000, 173-182

Collingwood R (1938) The Principles of Art. Clarendon Press, Oxford, UK

de Melo C, Paiva A (2005) Environment Expression: Expressing Emotions through Cameras, Lights and Music. Proceedings of Affective Computing Intelligent Interaction (ACII'05), 715-722

de Melo C, Paiva A (2006a) Multimodal Expression in Virtual Humans. Computer Animation and Virtual Worlds Journal, 17(3):215-220

de Melo C, Paiva A (2006b) A Story about Gesticulation Expression. Proceedings of Intelligent Virtual Agents (IVA'06), 270-281

de Melo C, Paiva A (2007). The Expression of Emotions in Virtual Humans using Lights, Shadows, Composition and Filters. Proceedings of Affective Computing Intelligent Interaction (ACII'07), 546-557

Dissanayake E (1988) What is Art for?. University of Washington Press, Seattle, USA

Ekman P (1999). Facial Expressions. In: Dalgleish T and Power M, (eds) Handbook of Cognition and Emotion, John Wiley & Sons, New York, USA

Elliot R (1966) Aesthetic Theory and the Experience of Art paper. Proceedings of the Aristotelian Society, NS 67, III-26

Fraser T, Banks A (2004) Designer's Color Manual: the Complete Guide to Color Theory and Application. Chronicle Books, San Francisco, USA

Geertz C (1976) Art as a Cultural System. Modern Language Notes, 91(6):1473-1499

Gombrich E (1960) Art and Illusion; a Study in the Psychology of Pictorial Representation. Pantheon Books, New York, USA

Gratch J, Rickel J, Andre E et al (2002). Creating Interactive Virtual Humans: Some Assembly Required. IEEE Intelligent Systems, 17(4):54–63

Gross L, Ward L (2007) Digital Moviemaking – 6th edn. Thomson/Wadsworth, Belmont, USA

Hartmann B, Mancini A, Pelachaud C (2005) Implementing Expressive Gesture Synthesis for Embodied Conversational Agents. Proceedings of Gesture Workshop, 173-182

Hunt R (2004) The Reproduction of Colour. John Wiley & Sons, West Sussex, UK

Juslin P, Sloboda J (2001) Music and Emotion: Theory and Research. Oxford University Press, New York, USA

Kant I, Bernard J (1951) Critique of judgment. Hafner Pub. Co., New York, USA

Katz S (1991) Film Directing Shot by Shot: Visualizing from Concept to Screen. Focal Press, Studio City, USA
Keltner D, Ekman P, Gonzaga G et al (2003) Facial Expression of Emotion. In: Davidson R, Scherer K and Goldsmith J (eds) Handbook of Affective Sciences, 415-433, Oxford University Press, New York, USA
Lewis M (2001) Aesthetic Evolutionary Design with Data Flow Networks. PhD thesis, Ohio State University
Machado F (2006) Inteligencia Artificial e Arte. PhD thesis, Universidade de Coimbra
Malkiewicz K, Gryboski B (1986) Film Lighting: Talks with Hollywood's Cinematographers and Gaffers. Prentice Hall Press, New York, USA
Malkiewicz K, Mullen D (2005) Cinematography: a Guide for Filmmakers and Film Teachers. Simon & Schuster, New York, USA
Mesquita B (2003) Emotions as Dynamic Cultural Phenomena. In: Davidson R, Scherer K and Goldsmith J (eds) Handbook of Affective Sciences, 871-890, Oxford University Press, New York, USA
Millerson G (1999) Lighting for Television and Film – 3rd edn. Focal Press, Oxford, USA
Mitchell M (1998) An Introduction to Genetic Algorithms. MIT Press, Massachusetts, USA
Moller T, Haines E (2002) Real-Time Rendering – 2nd edn. AK Peters, Massachusetts, USA
Noh J, Neumann U (1998) A survey of facial modelling and animation techniques. Technical report, USC Technical Report 99-705
Oatley K (2003) Creative Expression and Communication of Emotions in the Visual and Narrative Arts. In: Davidson R, Scherer K and Goldsmith J (eds) Handbook of Affective Sciences, 481-502, Oxford University Press, New York, USA
Ortony A, Clore G, Collins A (1988) The Cognitive Structure of Emotions. Cambridge University Press, Cambridge, USA
Perlin K, Goldberg A (1996) Improv: A System for Scripting Interactive Actors. Virtual Worlds in Proceedings of SIGGRAPH'96, 205-216
Picard R (1997) Affective Computing. MIT Press, Massachusetts, USA
Sayre H (2007) A World of Art – 5th edn. Prentice Hall, New Jersey, USA
Schroder M (2004) Speech and emotion research: an overview of research frameworks and a dimensional approach to emotional speech synthesis. PhD thesis, Institute of Phonetics, Saarland University
Sims K (1991) Artificial evolution for computer graphics. ACM Computer Graphics, 25:319-328
St-Laurent S (2004) Shaders for Game Programmers and Artists. Thomson/Course Technology, Massachusetts, USA
Todd S, Latham W (1992) Evolutionary Art and Computers. Academic Press, San Diego, USA
Ventrella J (1995) Disney meets Darwin-the evolution of funny animated figures. IEEE Proceedings of Computer Animation, 19-21
World L (1996) Aesthetic selection: The evolutionary art of Steven Rooke. IEEE Computer Graphics and Applications, 16(1): 4-5
Zettl H (2008) Sight, Sound, Motion: Applied Media Aesthetics – 5th edn. Thomson/Wadsworth, Belmont, USA

Fig. 1. The virtual human model

Fig. 2. Filtering manipulates the virtual human pixels. In (a) no filter is applied. In (b) the contrast filter is used to reduce contrast and create a more mysterious and harsh look. In (c) the motion blur is used to convey nervousness. In (d) the style filter, which is less concerned with photorealism, conveys an energetic look

Fig. 3. The evolutionary model for expression of emotions in virtual humans

Fig. 4. Hypotheses encoding

Fig. 5. The crossover and mutation genetic operators

Fig. 6. Value and fitness evolution of a 50-iteration run with optimal parameters

Fig. 7. The initial and 44^{th} generation of a 50-iteration run using optimal parameters

Table 1. Top 20 parameter configurations

R	(p, r, m, e, SM)	HF	AF	V	AV
1	(50, 0.70, 0.00, 0.10, TS)	0.86	0.59	3.01	2.51
2	(25, 0.70, 0.00, 0.00, TS)	0.84	0.59	2.97	2.47
3	(50, 0.60, 0.00, 0.10, TS)	0.85	0.59	2.97	2.47
4	(50, 0.60, 0.00, 0.00, TS)	0.85	0.59	2.97	2.47
5	(25, 0.60, 0.00, 0.00, TS)	0.85	0.58	2.96	2.45
6	(50, 0.70, 0.10, 0.10, TS)	0.86	0.57	2.95	2.45
7	(50, 0.70, 0.00, 0.00, TS)	0.84	0.60	2.95	2.48
8	(25, 0.70, 0.00, 0.10, TS)	0.85	0.57	2.91	2.43
9	(25, 0.70, 0.20, 0.00, TS)	0.84	0.58	2.90	2.42
10	(25, 0.70, 0.10, 0.00, TS)	0.83	0.58	2.89	2.40
11	(50, 0.70, 0.10, 0.00, TS)	0.83	0.58	2.88	2.44
12	(50, 0.70, 0.20, 0.00, TS)	0.83	0.58	2.87	2.44
13	(50, 0.70, 0.20, 0.10, TS)	0.84	0.56	2.87	2.40
14	(50, 0.60, 0.20, 0.10, TS)	0.85	0.55	2.86	2.40
15	(50, 0.60, 0.10, 0.00, TS)	0.84	0.55	2.86	2.37
16	(25, 0.60, 0.20, 0.00, TS)	0.83	0.56	2.85	2.38
17	(25, 0.60, 0.10, 0.00, TS)	0.82	0.56	2.83	2.38
18	(25, 0.70, 0.10, 0.10, TS)	0.82	0.55	2.83	2.36
19	(50, 0.60, 0.20, 0.00, TS)	0.83	0.56	2.82	2.38
20	(50, 0.60, 0.10, 0.00, RW)	0.83	0.54	2.82	2.36

R - rank; SM - selection method; TS - tournament selection; RW - roulette wheel; HF - highest fitness; AF - average fitness; V - value; AV - average value

Table 2. Top 20 generations of a 50-iteration run using optimal parameters

G	PHF	PAF	PV
44	0.9005	0.7857	2.3324
48	0.9181	0.7492	2.2947
46	0.9106	0.756	2.2933
34	0.9083	0.7338	2.2441
26	0.8941	0.7353	2.2188
49	0.9106	0.7136	2.1884
33	0.9061	0.7124	2.1771
36	0.9138	0.6914	2.1704
42	0.9095	0.6944	2.1677
23	0.9113	0.6921	2.1667
17	0.8958	0.7005	2.1525
31	0.9129	0.6808	2.1473
37	0.8983	0.711	2.1386
35	0.9118	0.6814	2.1263
25	0.9038	0.6886	2.1249
16	0.903	0.6884	2.1229
19	0.9114	0.6759	2.1146
38	0.8903	0.6961	2.1129
43	0.9171	0.6784	2.1112
18	0.9001	0.6704	2.101

G - generation; PHF - population highest fitness; PAF - population average fitness; PV - population value

Table 3. Top 20 hypotheses from the 44th generation population of the 50-iteration run using optimal parameters

HF	HC	HD
0.9001	1	0.02
0.8979	1	0.02
0.8974	1	0.02
0.8974	1	0.02
0.8974	1	0.02
0.8890	1	0.02
0.8512	1	0.02
0.8512	1	0.02
0.8512	1	0.02
0.8512	1	0.02
0.8512	1	0.02
0.8485	2	0.04
0.8485	1	0.02
0.8485	1	0.02
0.8485	1	0.02
0.8460	1	0.02
0.8460	1	0.02
0.8460	1	0.02

HF - hypotheses fitness; HC - number of times the hypothesis occurs in the population; HD – hypothesis dominance, i.e., the ratio of HC over population size