Expression of Emotions in Virtual Humans Using Lights, Shadows, Composition and Filters

Celso de Melo and Ana Paiva

IST-Technical University of Lisbon and INESC-ID,
Av. Prof. Cavaco Silva, Taguspark,
2780-990 Porto Salvo, Portugal
cdemelo@gaips.inesc-id.pt, ana.paiva@inesc-id.pt

Abstract. Artists use words, lines, shapes, color, sound and their bodies to express emotions. Virtual humans use postures, gestures, face and voice to express emotions. Why are they limiting themselves to the body? The digital medium affords the expression of emotions using lights, camera, sound and the pixels in the screen itself. Thus, leveraging on accumulated knowledge from the arts, this work proposes a model for the expression of emotions in virtual humans which goes beyond embodiment and explores lights, shadows, composition and filters to convey emotions. First, the model integrates the OCC emotion model for emotion synthesis. Second, the model defines a pixel-based lighting model which supports extensive expressive control of lights and shadows. Third, the model explores the visual arts techniques of composition in layers and filtering to manipulate the virtual human pixels themselves. Finally, the model introduces a markup language to define mappings between emotional states and multimodal expression.

Keywords: Expression of Emotions, Virtual Humans, Expression in the Arts, Light Expression, Screen Expression.

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 Collingwood's passage [1], the artist wishes to express an emotion. He feels 'anger'. But not just any anger. This is a unique anger. Thus, he seeks, in whichever medium need be, to understand the feeling. When he succeeds, he has expressed it using words, sounds, lines, colors or textures. So, we observe, first, artists express emotions through art. In fact, for many, emotions are important for appreciating and attributing value to the arts [2]. Second, in order to express the full complexity of their feelings, artists need flexible media.

Digital technology is a flexible medium for the expression of emotions. In particular, four expression channels can be distinguished: camera [3], lights [4][5], sound [6], and screen [5][7]. Their manipulation for the purpose of expression of

A. Paiva, R. Prada, and R.W. Picard (Eds.): ACII 2007, LNCS 4738, pp. 549–560, 2007.

emotions is inspired, in many ways, in the arts [8] such as theatre, cinema, photography, painting and music.

Virtual humans, which are embodied characters inhabiting virtual worlds [9], introduce yet a new expression channel. Effectively, with embodiment comes the possibility of simulating the kind of bodily expression we see in humans. Such is the case for the expression of emotions through postures, facial expression and voice modulation. But, of course, the expression of emotions in virtual humans need not be limited to the body.

This work proposes to go beyond embodiment and explore the expression of emotions in virtual humans using lights, shadows, composition and filters. The focus lies in two expression channels: light and screen. Regarding the first, a pixel-based lighting model is defined which supports control of lights and shadows. Regarding the second, acknowledging that, at a meta level, virtual worlds and virtual humans are no more than pixels in a screen, composition and filtering from the visual arts is explored. Finally, with respect to emotion synthesis, the Ortony, Clore and Collins [10] cognitive appraisal emotion theory is integrated.

The rest of the paper is organized as follows. Section 2 provides background on expression of emotions in the arts, in the digital medium and in virtual humans, detailing how it can be achieved using lights, shadows, composition and filters. Section 3 presents the virtual human emotion expression model using lights, shadows, composition and filters. Section 4 describes our results with the model. Finally, section 5 draws conclusions and discusses future work.

2 Background

2.1 Arts and Emotions

There are several conceptions about what expression in the arts is. First, it relates to beauty as the creative expression of beauty in nature [11]. Second, it relates to culture as the expression of the values of any given society [12]. Third, it relates to individuality as the expression of the artists' liberties and creativity [13]. Finally, it relates to emotions as the expression of the artists' feelings [2].

In fact, many acknowledge the importance of emotions for appreciating and attributing value to the arts. From the perspective of the creator, expression in the arts is seen as a way of understanding and coming to terms with what he is experiencing affectively [1]. 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 [14].

This work proposes a solution to the problem of expressing emotions in virtual humans which capitalizes on the accumulated knowledge from the arts about expression of emotions. Precisely, the work introduces a model which supports expression of emotions using lights, shadows, composition and filters.

2.2 The Digital Medium and Emotions

Digital technology is a flexible medium for the expression of emotions. Here, four expression channels can be used to express emotions: *camera*, *lights*, *sound*, and *screen*. The camera [3] 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. Lights [4][5] 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: light type, placement and angle; shadow softness and falloff; color properties such as hue, brightness and saturation. Sound [6][7] 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. Finally, the screen [5][7] 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. This work shall focus on the light and screen expression channels.

2.3 Virtual Humans and Emotions

Virtual humans are embodied characters which inhabit virtual worlds [9]. First, virtual humans look like humans. Thus, research draws on computer graphics for models to control the body and face. Second, virtual humans 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. With respect to emotion synthesis, several cognitive appraisal theories of emotion have been explored, the Ortony, Clore and Collins (OCC) [10] being one of the most commonly used. With respect to emotion expression, research tends to focus on conveying emotions through synchronized and integrated gesture [15], facial [16] and vocal [17] 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 [18] 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 [19] propose a system which adds expressiveness to existent motion data based on the effort and shape parameters of a dance movement observation technique called Laban Movement Analysis. Hartmann [20] draws from psychology six parameters for gesture modification: overall activation, spatial extent, temporal extent, fluidity, power and repetition. Finally, closer to this work, de Melo [21] 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 and did not explore screen expression.

2.4 Lighting and Emotions

This work explores *lighting* to express virtual humans' emotions. Lighting is the deliberate control of light to achieve expressive goals. The functions of lighting include [4][5][7]: illumination; modeling; focus; visual continuity; aesthetics; and, expression of affective states. This work focuses on the latter two. These are achieved through manipulation of the following elements of light [4][5][7]: (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 [22] is convenient as these are, in Western culture, regularly manipulated to convey emotions [23]; (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. The functions of shadow include [4][5][7]: defining spatial relationships; modeling; contextualizing; revealing and concealing parts of the scene; aesthetics; and, expression of affective states. This work focuses on the latter two. These are achieved through manipulation of the following elements of shadow: (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 [4][5][7]: decorate; reveal or conceal parts of the scene; simulate the dynamics of indoors and outdoors light sources; change focus; and, relevant for this work, change the mood or atmosphere of the scene.

Digital lighting introduces new possibilities for expressive control of light. Lighting becomes more akin to painting as the artist can, besides controlling the elements of light and shadow, explicitly manipulate the final image. Furthermore, the artist is free to *cheat*, for instance, creating more light sources than the ones motivated by the scene. Specific elements of digital lighting relevant for the expression of emotions include [5]: (a) *composition and effects*, which explicitly manipulate the pixels of the final image. Subsection 2.5 overviews this element; (b) *separate light components*, which refers to independent manipulation of ambient, diffuse and specular components of light; (c)*selective lights*, which constraint illumination to a subset of the objects; (d) *shadow color*, which defines shadow color. If set to white, shadow is omitted; (e) *shadows-only lights*, which don't illuminate the objects but, create shadows.

This work proposes a model which supports the elements of light and shadow as well as transition mechanisms. Specialized techniques for characters and scenes which build on these elements, such as the three-point lighting technique [4], are only implicitly supported.

2.5 Composition and Emotions

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 [5]. 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 [7]. Shaders replace parts of the traditional pipeline with programmable units [24]. Vertex shaders modify vertex data such as position, color, normal and texture coordinates. Pixel shaders modify pixel data such as color and depth. Filtering has many advantages: it has constant performance independently of the scene complexity; it can be very expressive due to the variety of available filters [25]; and, it is scalable as several filters can be concatenated.

3 The Model

This work proposes a model for the expression of emotions in virtual humans based on the light and screen expression channels. Regarding the former, a pixel-based lighting model is defined which supports control of lights and shadows. Regarding the latter, composition and filtering is supported. Regarding emotion synthesis, the Ortony, Clore and Collins [10] (OCC) model is used. Finally, a markup language – Expressive Markup Language (EML) – is proposed to control multimodal expression. The model also supports bodily expression [26], however, this will not be addressed here. Figure 1 summarizes the model.

3.1 Light Expression Channel

Light expression relies on a local pixel-based lighting model. The model supports multiple sources, three light types and shadows using the shadow map technique [24]. 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 [22]; (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, ∞ beyond a distance of *attnMax*; (f)

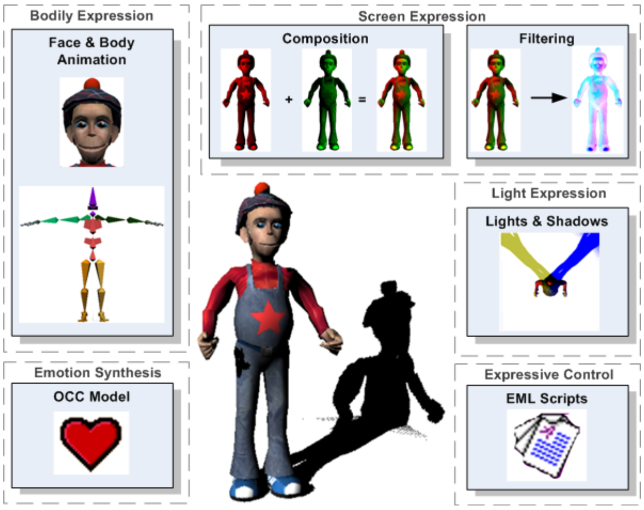

Fig. 1. The virtual human model supports bodily, light and screen expression. EML scripts control expression and the OCC model is used for emotion synthesis.

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; (i) *shadow map size*, which defines the size of the texture used for the shadow map. The greater the size, the more precise is the shadow. Finally, sophisticated lighting transitions (subsection 2.4), such as accelerations and decelerations, are supported based on parametric cubic curve interpolation [24] of parameters.

The lighting model's equation is defined as follows:

$$\mathbf{f}_{tot} = \sum_{l \in lights} lerp(\mathbf{f}^l_{ds}\beta^l, \mathbf{f}^l_{ds} \otimes \mathbf{c}_\beta, 1-\beta^l) + \mathbf{f}^l_a + \mathbf{m}_e \tag{1}$$

Where: $\mathbf{f}_{tot}$ is the pixel final color; *lerp* is the linear interpolator, used here to simulate light softness; $\mathbf{f}^l_{ds}$ is the diffuse and specular contribution from light l; β^l is the shadow factor for light l; $\mathbf{c}_\beta$ is the shadow color; $\mathbf{f}^l_a$ is the ambient contribution from light l; $\mathbf{m}_e$ is the material's emissive component.

The diffuse and specular contribution, $\mathbf{f}_{ds}$, is defined as follows:

$$\mathbf{f}_{ds} = spot \times attn(\mathbf{f}_d + \mathbf{f}_s) \tag{2}$$

Where: *spot* defines a spotlight cone [24]; *attn* defines a distance-based attenuation [5]; $\mathbf{f}_d$ is the diffuse contribution; $\mathbf{f}_s$ is the specular contribution.

The diffuse component, $\mathbf{f}_d$, is defined as follows:

$$\mathbf{f}_d = \mathbf{i}_d \otimes \mathbf{c}_d \otimes \max(\mathbf{n} \cdot \mathbf{l}, 0)\mathbf{m}_d \tag{3}$$

Where: $\mathbf{i}_d$ is the diffuse intensity; $\mathbf{c}_d$ is the diffuse light color; $\mathbf{n}$ is the normal; $\mathbf{l}$ is the light vector; $\mathbf{m}_d$ is the diffuse material component.

The specular component, $\mathbf{f}_s$, is defined as follows:

$$\mathbf{f}_s = \mathbf{i}_s \otimes \mathbf{c}_s \otimes \max(\mathbf{n} \cdot \mathbf{h}, 0)^{shi}\mathbf{m}_s \quad (4)$$

Where: $\mathbf{i}_s$ is the specular intensity; $\mathbf{c}_s$ is the specular light color; $\mathbf{n}$ is the normal; $\mathbf{h}$ is the Blinn-Phong [24] half vector; *shi* is the shininess exponent; $\mathbf{m}_s$ is the specular material component.

The ambient component, $\mathbf{f}_a$, is defined as follows:

$$\mathbf{f}_a = \mathbf{i}_a \otimes \mathbf{c}_a \otimes \mathbf{m}_a \quad (5)$$

Where: $\mathbf{i}_a$ is the ambient intensity; $\mathbf{c}_a$ is the ambient light color; $\mathbf{m}_a$ is the ambient material component.

The shadow factor, β, which lies in the range $[0, 1]$ determines the amount of shadow per-pixel. It is set to 0 if the pixel is in shadow and to 1 if it is fully light. Values within the range $]0, 1[$ occur in the transition between light and dark areas. The algorithm to calculate β corresponds to a variation of the shadow map technique [24] as described in [25].

3.2 Screen Expression Channel

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. Innumerable filters have already been developed [25]. Some have been used in digital art to express emotions [7]. This work explores a subset of such filters. However, the focus here is not in presenting a basic set of filters but, to call attention to its expressiveness. For illustration purposes, some of the filters explored are: (a) the *contrast filter*, Fig.2-(b), which controls virtual human contrast and can be used to simulate exposure effects [4]; (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 [23]; (d) the *grayscale filter*, Fig.3-(c), which totally desaturates the virtual human, rendering him in grayscale, thus, forcing the audience to immerse emotionally in the action [7]. Filters can be concatenated to create compound effects and, its parameters interpolated using parametric cubic curve interpolation [24].

Composition refers to the process of [5]: 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

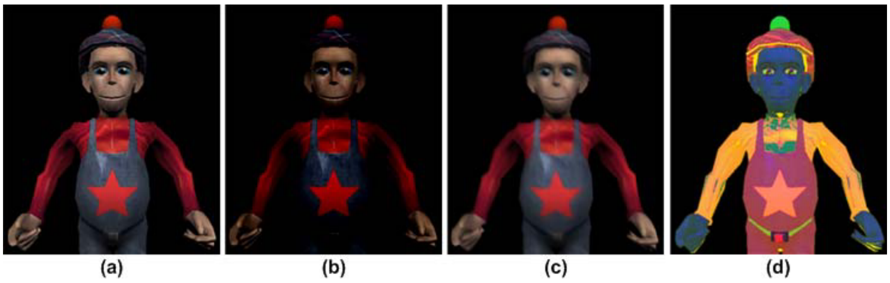

Fig. 2. Filtering manipulates the virtual human pixels. In (a) no filter is applied. In (b) the contrast filter is used to reduce contrast and create a more mysterious and harsh look [4]. In (c) the motion blur is used to convey nervousness [7]. In (d) the style filter, which is less concerned with photorealism, conveys an energetic look [23].

these lights; (e) *Defines a lighting mask*, which defines which components of the associated lights apply to the objects.

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. This work uses a standard blending equation [24]:

$$\mathbf{p} = \odot(\mathbf{p}_s \otimes \mathbf{b}_s, \mathbf{p}_d \otimes \mathbf{b}_d) \tag{6}$$

Where: $\mathbf{p}$ is the output pixel, i.e., an RGBA value; $\odot$ is one of the following operations: addition, subtraction, max or min; $\mathbf{p}_s$ is the source layer pixel; $\mathbf{b}_s$ is the source blend factor, which can be one of the following: zero, one, the source color, the destination color, the source alpha or the destination alpha; $\mathbf{p}_d$ is the destination layer pixel; $\mathbf{b}_d$ is the destination blend factor.

3.3 Emotion Synthesis

Virtual human emotion synthesis is based on the Ortony, Clore and Collins (OCC) model [10]. 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 [27], represented by an inverse exponential function. Emotion reinforcement is, so as to simulate the saturation effect [27], 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; converges to zero linearly in time; reinforces with emotion eliciting; if positive, increases the elicited emotions' potential, if negative, decreases it. Further details about the implementation of this emotion synthesis model can be found in [21].

3.4 Expressive Control

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 [28]. The *body*, *gesticulation*, *voice* and *face* modules control bodily expression [26]. 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 filter lists associated with composition layers. 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: $\{emotionConditions\}* \rightarrow \{bodyAction \mid lightAction \mid screenAction \mid emotionAction\}*$. Emotional conditions – *emotionConditions* – evaluate mood, arousal or active emotions' intensity or valence. Expressive actions – *bodyAction*, *lightAction* and *screenAction* – refer to body, light or screen actions as defined by its respective modules. Emotion actions – *emotionAction* – elicit further emotions.

4 Results

Consider a scenario were a virtual human expresses emotions using the proposed model. Initially, the emotional state is neutral and no emotion is expressed through any of the expression channels. The light setup follows the standard *three-point lighting* technique [4]. The *key light*, which defines the main illumination angle, is placed to the virtual human's right and above. The *fill light*, which softens and extends the key light, is placed to the virtual human's left and above. A point light is chosen for the key light and a directional light for the fill light. Both irradiate white light. Furthermore, in accord with standard practices [5], the key light is set with a higher intensity than the fill light and, only the key light is set to cast shadows. Besides being softened by the fill light, shadows are set to have a soft falloff. Finally, with respect to screen expression, two layers are defined, one for each light source. Layer combination is set to simple addition. No filtering is applied. Fig.3-(a) presents the virtual human in the initial state.

Suppose that, for some reason, the OCC anger emotion is elicited. As the emotion intensity increases, the key light color fades into an intense red, as it is known that, in Western culture, warm colors tend to be associated with excitement and danger [5][7]. Furthermore, the camera approaches so as to increase drama [3]. Finally, with respect to bodily expression, the face is set to express anger. Fig.3-(b) presents the virtual human in the anger state.

To conclude the scenario, consider that it is revealed to the virtual human that the cause of his anger is more serious than anticipated and, furthermore,

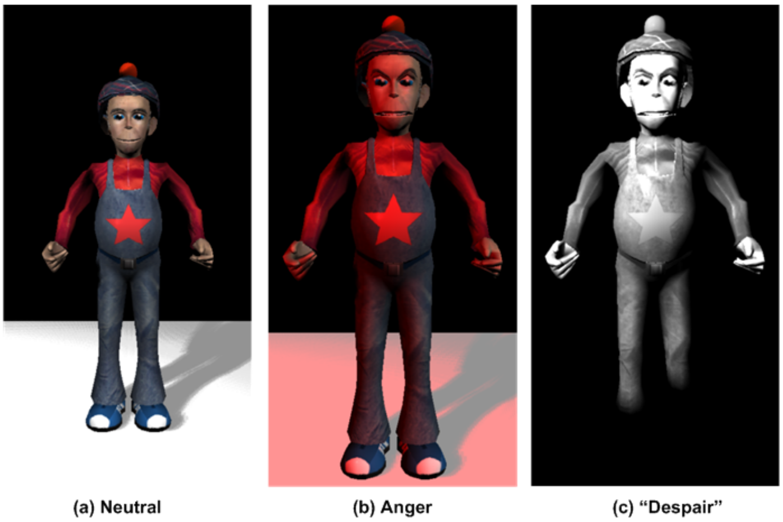

Fig. 3. The virtual human expressing emotions using light and screen expression. In (a) no emotions are being expressed. In (b) anger is expressed by changing the lights' color to a bright red. In (c) anger, fear and distress are being expressed by total desaturation of the scene and an overexposure effect.

that there is nothing he can do to prevent it. In this case, besides anger intensifying, the virtual human starts to feel sad and afraid (thus, eliciting the OCC sadness and fear emotions). Now, suddenly, the key light shadows harden, its falloff increases and its intensity sharply increases creating an overexposure effect [4] which overwhelms the virtual human with light; the fill light layer is faded out, thus reducing overall scene brightness and increasing contrast; finally, the grayscale filter is applied to the scene, thus, removing all chromatic color. The scene is now a perfect example of the photographic *chiaroscuro* lighting type which emphasizes contrasts, thus, increasing overall dramatic effect [4]. Furthermore, the absence of chromatic color is said to, according to the desaturation theory, "induce the audience to participate in the event, to look *into* rather than merely *at* it" [7]. Fig.3-(c) presents the virtual human in this "desperate" complex affective state.

5 Conclusions and Future Work

This work proposes a model for the expression of emotions in virtual humans using lights, shadows, composition and filters. Regarding light expression, a pixel-based lighting model is defined which provides several control parameters. Parameter interpolation based on parametric cubic curves supports sophisticated lighting transitions. Regarding screen expression, filtering and composition are explored. 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 objects are separated into layers, which are subjected to independent lighting constraints and filters, before being combined to generate the final image. Regarding emotion synthesis, the OCC emotion model is integrated. Finally, the model also proposes a markup language to control multimodal expression as well as define rules mapping emotional states to light and screen expression.

Regarding future work, light expression could be enhanced if lighting techniques were explored. For instance, the *chiaroscuro*, *flat* and *silhouette* lighting styles are known to convey specific moods [4][7]. Regarding screen expression, more filters could be explored. Many are already available [25]. However, first, their affective impact should be clarified; second, only filters suitable for real-time execution should be explored, since timing constraints on virtual humans are very tight. Regarding control of multimodal expression, this work proposes a language which supports definition of rules. There are, in fact, several rules and guidelines for effective artistic expression [7]. However, the expression of emotions in the arts is essentially a creative endeavor and artists are known to break these rules regularly [8]. Thus, a model which relies on rules for the expression of emotions is likely not to be sufficient. 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. Finally, the digital medium has immense potential and further expression modalities should be explored. Two obvious extensions to this work include exploring the camera and sound expression channels of which much knowledge already exists in the arts [3][6][7].

References

1. Collingwood, R.: Art and Craft. In: Feagin, S., Maynard, P. (eds.) Aesthetics, pp. 215–220. The Oxford University Press, New York, USA (1997)
2. Bell, C.: The Metaphysical Hypothesis. In: Feagin, S., Maynard, P. (eds.) Aesthetics, pp. 158–159. The Oxford University Press, New York, USA (1997)
3. Arijon, D.: Grammar of the Film Language. Silman-James Press, California, USA (1976)
4. Millerson, G.: Lighting for Television and film, 3rd edn. Focal Press, Oxford, UK (1999)
5. Birn, J.: [digital] Lighting and Rendering, 2nd edn. New Riders, California, USA (2006)
6. Juslin, P., Sloboda, J.: Music and Emotion: theory and research. Oxford University Press, New York, USA (2001)
7. Zettl, H.: Sight Sound Motion: Applied Media Aesthetics. Wadsworth Publishing, California, USA (2004)
8. Sayre, H.: A World of Art, 5th edn. Prentice-Hall, Englewood Cliffs (2007)
9. Gratch, J., Rickel, J., Andre, E., Badler, N., Cassell, J., Petajan, E.: Creating Interactive Virtual Humans: Some Assembly Required. IEEE Intelligent Systems 17(4), 54–63 (2002)

10. Orthony, A.: The Cognitive Structure of Emotions. Cambridge University Press, New York, USA (1990)
11. Batteux, A.: The Fine Arts Reduced to a Single Principle. In: Feagin, S., Maynard, P. (eds.) Aesthetics, pp. 102–104. The Oxford University Press, New York, USA (1997)
12. Clifford, G.: Art as a Cultural System. In: Feagin, S., Maynard, P. (eds.) Aesthetics, pp. 109–118. The Oxford University Press, New York, USA (1997)
13. Kant, I.: Art and Genius. In: Feagin, S., Maynard, P. (eds.) Aesthetics, pp. 180–192. The Oxford University Press, New York, USA (1997)
14. Elliot, R.: Aesthetic Theory and the Experience of Art. In: Feagin, S., Maynard, P. (eds.) Aesthetics, pp. 278–288. The Oxford University Press, New York, USA (1997)
15. Cassell, J.: Nudge, Nudge, Wink, Wink: Elements of Face-to-Face Conversation for Embodied Conversational Agents. In: Cassell, J., Sullivan, J., Prevost, S., Churchill, E. (eds.) Embodied Conversational Agents, pp. 1–27. The MIT Press, Massachusetts, USA (2000)
16. Noh, J., Neumann, U.: A Survey of Facial Modeling and Animation Techniques. USC Technical Report 99-705 (1998)
17. Schroder, M.: Speech and emotion research: an overview of research frameworks and a dimensional approach to emotional speech synthesis. PhD thesis. vol. 7 of Phonus, Research Report of the Institute of Phonetics, Saarland University (2004)
18. Amaya, K., Bruderlin, A., Calvert, T.: Emotion from motion. In: Proceedings Graphics Interface'96, pp. 222–229 (1996)
19. Chi, D., Costa, M., Zhao, L., Badler, N.: The EMOTE model for effort and shape. In: Proceedings of SIGGRAPH 2000, pp. 173–182 (2000)
20. Hartmann, B., Mancini, A., Pelachaud, C.: Implementing Expressive Gesture Synthesis for Embodied Conversational Agents. In: Gesture Workshop. LNCS (LNAI). Springer, Heidelberg (2005)
21. de Melo, C., Paiva, A.: Environment Expression: Expressing Emotions through Cameras, Lights and Music. In: Tao, J., Tan, T., Picard, R.W. (eds.) ACII 2005. LNCS, vol. 3784, pp. 715–722. Springer, Heidelberg (2005)
22. Hunt, R.: The Reproduction of Colour, 6th edn. John Wiley & Sons, Ltd., Sussex, UK (2004)
23. Fraser, T., Banks, A.: Designer's Color Manual: The Complete Guide to Color Theory and Application. Chronicle Books, California, USA (2004)
24. Moller, R., Haines, E.: Real-Time Rendering, 2nd edn. AK Peters Ltd., Massachusetts, USA (2002)
25. St-Laurent, S.: Shaders for Game Programmers and Artists. Thomson Course Technology, Massachusetts, USA (2004)
26. de Melo, C., Paiva, A.: Multimodal Expression in Virtual Humans. Computer Animation and Virtual Worlds Journal. 17(3), 239–248 (2006)
27. Picard, R.: Affective Computing. The MIT Press, Massachusetts, USA (1997)
28. SMIL: Synchronized Multimedia Integration Language (SMIL) (2007), http://www.w3.org/AudioVideo/