Evolving Expression of Emotions through Color in Virtual Humans using Genetic Algorithms

Celso M. de Melo[1] and Jonathan Gratch[1]

[1] Institute for Creative Technologies, University of Southern California,
13274 Fiji Way, Marina Del Rey, CA 90292, USA
demelo@usc.edu, gratch@ict.usc.edu

Abstract. For centuries artists have been exploring the formal elements of art (lines, space, mass, light, color, sound, etc.) to express emotions. This paper takes this insight to explore new forms of expression for virtual humans which go beyond the usual bodily, facial and vocal expression channels. In particular, the paper focuses on how to use color to influence the perception of emotions in virtual humans. First, a lighting model and filters are used to manipulate color. Next, an evolutionary model, based on genetic algorithms, is developed to learn novel associations between emotions and color. An experiment is then conducted where non-experts evolve mappings for joy and sadness, without being aware that genetic algorithms are used. In a second experiment, the mappings are analyzed with respect to its features and how general they are. Results indicate that the average fitness increases with each new generation, thus suggesting that people are succeeding in creating novel and useful mappings for the emotions. Moreover, the results show consistent differences between the evolved images of joy and the evolved images of sadness.

1 Motivation

Virtual humans are embodied agents which inhabit virtual worlds and act and look like humans [1]. Inspiring on the human face-to-face conversation paradigm, virtual humans are capable of expressing themselves using verbal and non-verbal modalities in an integrated and synchronized fashion. In order to further increase believability, naturalness and efficiency of communication, virtual humans have been endowed with models of emotions. In particular, research on expression of emotions has tended to focus on the modalities people use in daily interaction: gesture, face and voice. In contrast, this work explores a new form of expression which capitalizes on accumulated knowledge from the visual arts and goes beyond the usual bodily, facial and vocal forms of expression.

In fact, artists have been exploring for centuries the idea that it is possible to perceive emotions in line, space, mass, light, color, texture, pattern, sound and motion [2]. In a simpler conception, art is seen as the expression of the artist's feelings [3, 4]. However, John Hospers [5] refined this view by noting that the work of art need not reflect the emotions of its creator but can be said to possess emotional properties in its own right. Thus, first, the creator manipulates the formal elements of art (line, space, mass, light, color, texture, pattern, sound and motion) to convey felt or imagined

emotions. Then, the audience relies on analogies with the internal and external manifestations of emotions they experienced in the past to interpret the work of art. This work takes this insight and explores color to manipulate the perception of emotions in virtual humans.

Color has been widely manipulated by artists in the visual arts to convey emotion [2, 6]. Color is the result of interpretation in the brain of the perception of light in the human eye. Thus, the manipulation of light in the visual arts, called *lighting*, has always been a natural way of achieving specific effects with color [7, 8]. In this work, color is manipulated using a lighting model. Moreover, color can also be looked at as an abstract property of a scene and manipulated explicitly with no particular concern for the physics of light. This has been explored in abstract painting [2] and, more recently, in the visual media [9]. The work presented in this paper also explores this form of manipulation and uses *filters* to achieve such color effects. Filters do post-processing of pixels in a rendered image according to user-defined programs [10].

Having defined the expression modality the following question ensues: How to find *novel* mappings of emotions into color which are *useful* both for the individual and society (i.e., that generalize beyond the individual)? A first difficulty is that perception of emotion in color is influenced by biological, individual and cultural factors [2, 6]. Secondly, looking at the literature on lighting, it is possible to find general principles on how to convey moods or atmosphere [7, 8, 11, 12] but, these aren't sufficient to differentiate between emotions and usually do not reflect the character's mood but the narrative (such as the climax, for instance). The literature on filters is far scarcer and tends to focus on technical aspects or typical uses rather than on its affective properties [8, 9, 13]. Therefore, this work pursues an approach which is not dependent on the existent literature and tries, instead, to learn such mappings directly from people.

Moreover, the interest here is in learning intuitions about expression of emotion through color from non-experts. This is in contrast to previous approaches which attempt to learn the affective properties of lighting from artists [15, 16] or the existent literature [17, 18]. Effectively, being able to learn from non-experts is a necessity when new forms of expression are being explored. As noted above, this is specially the case with respect to finding expertise on the affective properties of filters. Furthermore, this will later facilitate extending the proposed system to other elements of art. Therefore, the system needs to be responsible for generating the alternatives, which a non-expert is unlikely to be proficient in doing, and the user should only be responsible for evaluating them (as to how well they convey the emotion).

An evolutionary approach, which relies on genetic algorithms, is used to learn mappings between emotions and color. The focus is on joy and sadness and whether the approach is applicable to other emotions is a topic of future work. Genetic algorithms [14] are appropriate for several reasons. The clear separation between generation and evaluation of alternatives is convenient. Alternatives can be generated using biologically inspired operators – mutation, crossover, etc. Evaluation, in turn, relies on feedback from people. Finally, the expression space defined by lighting and filters is very large and genetic algorithms deal well with intractable search spaces.

The rest of the paper is organized as follows: Section 2 describes the lighting and filters model used to manipulate color; Section 3 describes the evolutionary model used to learn the mappings of emotions into color; Section 4 describes two

experiments which were conducted to define and understand the mappings of joy and sadness; finally, Section 5 discusses the results and draws conclusions.

2 The Expression Model

The lighting model defines local pixel-level illumination of the virtual human. Among the supported parameters, the following are used in this work: (a) *type*, defines whether the light source is directional, point or spotlight; (b) *direction*, defines the illumination angle; (c) *ambient*, *diffuse* and *specular colors*, define the light color for each component. Color can be defined in either RGB (red, green, blue) or HSB (hue, saturation, brightness) spaces; *ambient*, *diffuse* and *specular intensity*, define a value which is multiplied with the respective component color. Setting the value to 0 disables the component. Filters are used to post-process the pixels of the illuminated rendered image of the virtual human. Several filters are available in the literature [19] and this work uses the following subset: the *color* filter, Fig.1-(b) and (c), sets the virtual human's color to convey a stylized look such as black & white, sepia or inverted colors; the *HSB* filter, Fig.1-(d) and (e), manipulates the virtual human's hue, saturation or brightness. Filters can also be concatenated to create compound effects. Further details about the expression model can be found elsewhere [20].
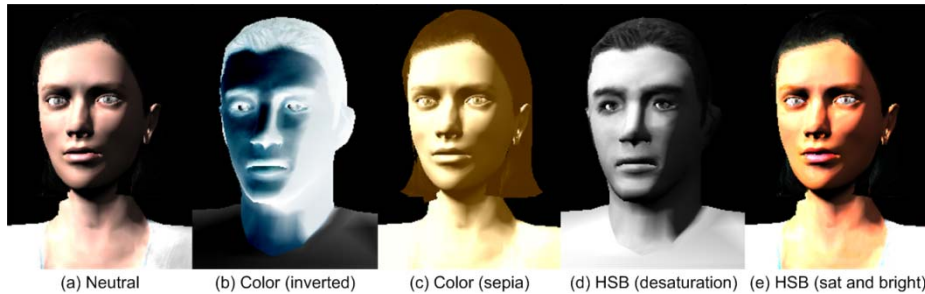

(a) Neutral   (b) Color (inverted)   (c) Color (sepia)   (d) HSB (desaturation)   (e) HSB (sat and bright)

Fig. 1. Filters used to post-process the rendered image of the illuminated virtual human. No filter is applied in (a). The color filter is used to invert the colors in (b) and create the sepia look in (c).The HSB filter is used to reduce saturation in (d) and to increase the saturation and brightness in (e). Both virtual humans used in this work are shown.

3 The Evolutionary Model

Building on the expression model, the evolutionary model uses genetic algorithms to evolve, for a certain emotion, a *population* of *hypotheses*, which define specific configurations of lighting and filters parameters. Evolution is guided by feedback from the user as to how well each hypothesis conveys the intended emotion. The *fitness function*, in this case, is the subjective criteria of the user.

At the core lies a standard implementation of the genetic algorithm [14]. The algorithm is characterized by the following parameters: (a) stopping criteria to end the algorithm, i.e., the maximum number of iterations; (b) the size of the population, p, to

be maintained; (c) the selection method, *sm*, to select probabilistically among the hypotheses in a population when applying the genetic operations. Two methods are supported: *roulette wheel*, which selects a hypothesis according to the ratio of its fitness to the sum of all hypotheses' fitness; *tournament selection*, which selects with probability *p'* the most fit among two hypotheses selected using roulette wheel; (d) the crossover rate, *r*, which defines the percentage of the population subjected to crossover; (e) the mutation rate, *m*, which defines the percentage of the population subjected to mutation; (f) the elitism rate, *e*, which defines the percentage of the population which propagates unchanged to the next generation. The rationale behind elitism is to avoid losing the best hypotheses from the previous population in the new population [14].

The algorithm begins by setting up the initial population with random hypotheses. Thereafter, the algorithm enters a loop, evolving populations, until the stopping criterion is met. In each iteration, first, *(1-r)p* percent of the population is selected for the next generation; second, *r*p/2* pairs of hypotheses are selected for crossover and the offspring are added to the next generation; third, *m* percent of the population is randomly mutated; fourth, *e* percent of the hypotheses is carried over unchanged to the next generation. Evaluation is based on feedback from the user.

The hypothesis is structured according to the lighting and filter parameters. Lighting uses the common three-point configuration [7, 8] which defines a primary key light and a secondary fill light. The backlight is not used in this work. Both lights are modeled as directional lights and are characterized by the following parameters: (a) *direction*, corresponds to a bi-dimensional floating-point vector defining angles about the x and y axis with respect to the camera-character direction. The angles are kept in the range [-75.0°, 75.0°] as these correspond to good illumination angles [5]; (b) diffuse color, corresponds to a RGB vector; (c) K_d, defines the diffuse color intensity in the range [*0.0, 5.0*]; (d) K_s, defines the specular color intensity in the range [*0.0, 3.0*]. The HSB and color filters are also applied to the virtual human. Thus, four more parameters are defined: (a) *HSB.hue*, *HSB.saturation* and *HSB.brightness*, define the HSB filter's hue (in the range [*0.0, 10.0*]), saturation (in the range [*0.0, 5.0*]) and brightness (in the range [*0.5, 3.0*]); (b) *color.style*, which defines whether to apply the black & white, sepia or inverted colors style for the color filter. Both filters can be applied simultaneously. Further details on the evolutionary model can be found in another article [21].

4   Results

4.1   Learning the Mappings

In a first experiment, non-experts evolve mappings for joy and sadness. The experiment is designed so that subjects are unaware that genetic algorithms are being used. They are asked to classify five 'sets' (i.e., populations) of 'alternatives' (i.e., hypotheses) for the expression of each emotion. Classification of alternatives goes from *0.0* ('the image does not express the emotion at all' or low fitness) to *1.0* ('the image perfectly expresses the emotion' or high fitness). The sets are presented in

succession, being the first generated randomly and the succeeding ones evolved by the genetic algorithm. The experiment is automated in software. The user can save the session and continue at any time. A random name is given to the session so as to preserve anonymity. The parameters for the genetic algorithm are: $p = 30$, $sm =$ *tournament selection*, $r = 0.70$, $m = 0.15$ and $e = 0.10$. Two virtual humans are used: a male and a female. The rationale for using multiple virtual humans is to minimize geometry effects in the analysis of the results (e.g., the illusion of a smile under certain lighting conditions even though no smile is generated). Participants are evenly distributed among virtual humans. The virtual human assumes the anatomical position and Perlin noise and blinking is applied. No gesture, facial or vocal expression is used throughout the whole experiment. Transition between hypotheses is instantaneous. The camera is fixed and frames the upper body of the virtual human.

The study was conducted in person at the University of Southern California campus and related institutions. Thirty subjects were recruited. Average age was 26.7 years, 46.7% were male, mostly having superior education (93.3% college level or above) in diverse fields. All subjects were recruited in the United States, even though having diverse origins (North America: 50.0%; Asia: 20%; Europe: 20%; South America: 6.7%). Average survey time was around 20 minutes.

The evolution of the average population fitness for joy and sadness is shown in Fig.2. Fourteen (out of possible thirty) of the highest fit hypotheses, one per subject, for joy and sadness are shown in Figures 3 and 4, respectively.
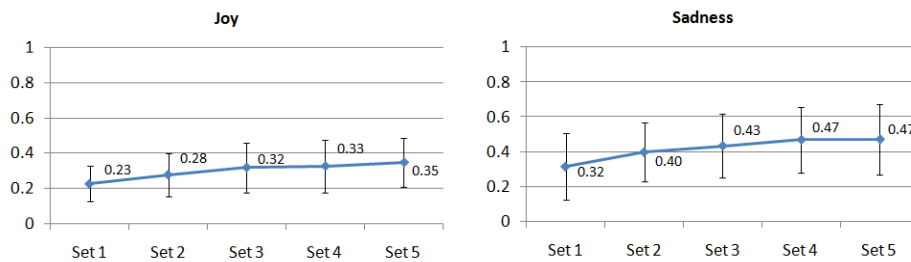

Fig. 2. Average fitness per set (with standard deviations) for joy and sadness.

4.2 Understanding the Mappings

The goals of a second experiment are to understand: (a) what features differentiate the mappings evolved in the first experiment; (b) how general the mappings are. Regarding the first goal, features refer to characteristics in the image generated by the respective hypothesis. The idea, then, is to differentiate the best images for joy and sadness using these features. These images are the union of, for each emotion, for each subject in the first study, the one with the highest classification. Thus, in total, 60 images are used: the 30 best for joy, one per subject; the 30 best for sadness, one per subject. Now, if the first experiment already provided a measure of value for the individuals, the second goal seeks to assess how valuable are the mappings beyond the individuals that generated them. The idea is to understand if there are common

patterns in the mappings evolved by each individual and how do these mappings relate to the existent literature. The existent literature is used here as a standard which represents knowledge which already has been shown to be of value to the field.

Fig. 3. Fourteen of the highest fit hypotheses for joy. Each hypothesis is from a different subject.
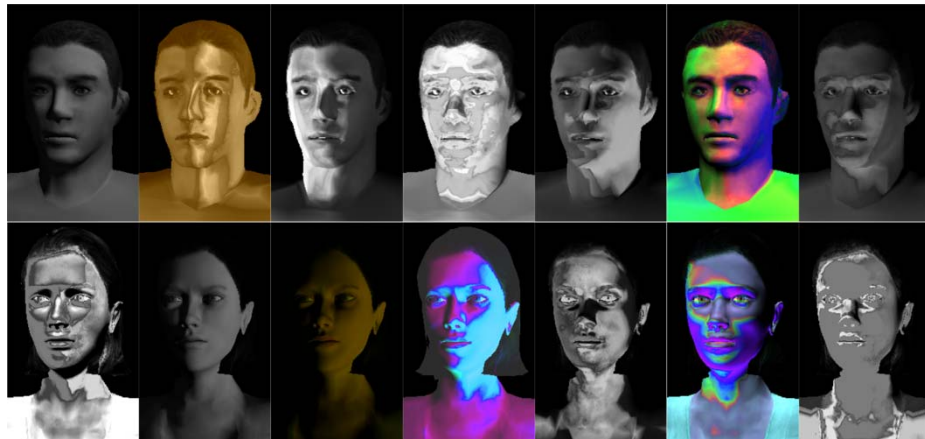

Fig. 4. Fourteen of the highest fit hypotheses for sadness. Each hypothesis is from a different subject.

Three features were chosen from the literature that measure properties of the pixels in the images generated by the hypotheses: brightness, saturation and number of colors. The *brightness* of an image is defined, in the range [*0.0, 1.0*], as the average brightness of the pixels. The brightness of a pixel is the subjective perception of luminance in the pixel's color. The *saturation* of an image is defined, in the range [*0.0, 1.0*], as the average saturation of the pixels. Saturation of a pixel refers to the intensity of the pixel's color. Standard formulas are used to calculate brightness and saturation [22]. Finally, the *number of colors* of an image is defined to be the number of different colors in the pixels. However, the maximum number of colors was

reduced by rounding the RGB components to one decimal place. Intuitively, this means the feature is only interested in relatively large differences in color.

Having calculated the feature values, the dependent t test was used to compare means between joy and sadness hypotheses with respect to each feature. The results are shown in Table 1.

Table 1. Dependent t test statistics ($df=29$) for difference in means between the joy and sadness images with respect to brightness (BRIG), saturation (SAT) and number of colors (NCOL).

	Brightness*	Saturation*	Number of Colors *
Mean Diff.	0.12	0.25	199.23
Std. Deviation	0.15	0.29	326.14
Std. Err. Mean	0.03	0.05	59.55
95% CI Lower	0.06	0.14	77.45
95% CI Upper	0.17	0.35	321.02
t	4.26	4.70	3.35
Sig. (2-tailed)	0.00	0.00	0.00

* Significant difference, p<0.05

The results in Table 1 show that:
- The average brightness in joy images (*M=0.36, SE=0.02*) is higher than in sadness (*M=0.24, SE=0.02, t(29)=0.00, p<0.05, r=0.62*);
- The average saturation in joy images (*M=0.44, SE=0.04*) is higher than in sadness (*M=0.19, SE=0.04, t(29)=0.00, p<0.05, r=0.66*);
- The average number of colors in joy images (*M=302.20, SE=374.46*) is higher than in sadness (*M=102.97, SE=29.93, t(29)=0.00, p<0.05, r=0.53*).

Finally, to assess how general the mappings are, supervised learning techniques were used to learn models that differentiate images of joy and sadness. In particular, decision trees [23] were used to classify the 60 images with respect to the three features. The J48 implementation of decision trees in Weka [24] was used with default parameters and 10-fold cross-validation. The resulting tree correctly classifies 47 (78.3%) of the images and is shown in Fig.5. Further details on this and the previous experiment can be found in another paper [25].

```
NCOLORS <= 26: sadness (23.0/3.0)
NCOL ORS> 26
|  BRIGHTNESS<= 0.302
|  |  SATURATION <= 0.413: sadness (7.0)
|  |  SATURATION > 0.413: joy (10.0/2.0)
|  BRIGHTNESS > 0.302: joy (20.0/1.0)
```

Fig. 5. Decision tree that distinguishes joy from sadness.

5 Discussion

This paper proposes to use accumulated knowledge from the arts to explore new forms of expression of emotions which go beyond the usual bodily, facial and vocal channels in virtual humans. In particular, the work focuses on how to convey emotion through one formal element of art: color. Color is manipulated using a sophisticated lighting model and filters. The paper further proposes an evolutionary approach, based on genetic algorithms, to learn novel and useful mappings of emotion into color. The model starts with a random set of hypotheses - i.e. configurations of lighting and filters - and, then, uses genetic algorithms to evolve new populations of hypotheses according to feedback provided by non-experts.

In a first experiment, subjects are asked to evolve mappings for joy and sadness using the evolutionary model. Subjects successively classify five sets of hypotheses, for each emotion, without being informed that a genetic algorithm is being used to generate the sets. The results show that the average set fitness for both emotions is monotonically increasing with each succeeding set (Fig.2). This suggests that: (a) subjects are succeeding in finding a novel mapping for the expression of emotions through color; (b) the genetic algorithm is succeeding in providing more useful hypotheses with each successive generation. The fact that subjects are unaware that an evolutionary approach is being used allows us to exclude the possibility that they are classifying later hypotheses better just because that is what is expected of them in an evolutionary approach. Nevertheless, the results also show that the average fitness of the fifth and final set is well below the perfect score of 1.0. This might be explained for two reasons: (a) too few sets are being asked to be evolved. This, then, would have been an experimental constraint which existed to limited survey time and not a fundamental limit on the expressiveness of color; (b) no gesture, facial or vocal expression is used. Effectively, these channels have already been shown to play an important role on the expression of emotions in virtual humans [1] and this paper is not arguing otherwise.

A second experiment analyzes which features characterize the mappings for joy and sadness. Three features were drawn from the literature: brightness, saturation, and number of colors. The results show consistency between the mappings evolved by different subjects. In particular, the results show that images of joy tend to be brighter, more saturated and have more colors than images of sadness (Table 1 and Fig.5). This suggests that the mappings also reflect values which are shared among the individuals and, therefore, that the mappings have the potential to generalize beyond the individuals that created them. Moreover, these results are in line with the lighting literature [7, 8, 11, 12]. This provides further support that the mappings reflect values which generalize beyond the individuals. Finally, the fact that it was possible to learn, using 10-fold cross-validation, a decision tree model which explains the data with a relatively high success rate, also suggests that there is the potential for generalizing beyond the particular examples that were used to learn the decision tree. In summary, if the first experiment suggested that the proposed evolutionary approach is capable of producing novel mappings that are useful at least for the individual, the second experiment suggests that those mappings are also useful for society.

Regarding future work, it would be interesting to explore whether the evolutionary approach generalizes to more emotions. From our experience and the feedback from

subjects, we believe this might be so for some, but not all, emotions. Finally, color is but one of the many elements that have been widely explored in the arts. Other elements include: line, space, mass, texture, shape, pattern, sound, motion, etc. It should, therefore, be worth exploring whether the proposed approach also generalizes to these other formal elements in the visual arts [2].

Acknowledgments

References

1. Gratch, J., Rickel, J., Andre, E., Badler, N., Cassell, J., Petajan, E.: Creating Interactive Virtual Humans: Some Assembly Required. IEEE Intelligent Systems. IEEE Computer Society, 17(4): 54–63 (2002)
2. Sayre, H.: A World of Art, fifth ed. New Jersey, USA: Prentice Hall (2007)
3. Collingwood, R.. The Principles of Art. Oxford, UK: Clarendon Press (1938)
4. Tolstoy, L.: What is Art? Oxford, USA: Oxford University Press (1955)
5. Hospers, J.: Aesthetics, problems of. The Encyclopedia of Philosophy, P. Edwards. New York, USA: Ed. Macmillan Publishing Co (1967)
6. Fraser, T., Banks, A.: Designer's Color Manual: The Complete Guide to Color Theory and Application. San Francisco, USA: Chronicle Books (2004)
7. Millerson, G.: Lighting for Television and film, 3rd ed. Oxford, UK: Focal Press (1999)
8. Birn, J.: [digital] Lighting and Rendering, 2nd ed. California, USA: New Riders (2006)
9. Zettl, H.: Sight, Sound, Motion: Applied Media Aesthetics. Belmont: Thomson/Wadsworth (2008)
10. Akenine-Moller, T., Haines, E., Hoffman, N.: Real-Time Rendering. Wellesley, USA: AK Peters Ltd (2009)
11. Brown, B.: Motion Picture and Video Lighting, 2nd ed. Burlington, USA: Elsevier Inc. (2007)
12. Alton J.: Painting with Light. New York, USA: Macmillan Co. (1949)
13. Gross, L., Ward, L.: Digital Moviemaking, sixth ed. Belmont, USA: Thomson/Wadsworth (2007)
14. Mitchell, M.: An Introduction to Genetic Algorithms, 3rd ed. Massachusetts, USA: MIT Press. 1998
15. Patow, G., Pueyo, X.. A survey of inverse rendering problems. Comp. Graphics Forum 22(4):663–687 (2003)
16. Pelaccini, F., Battaglia, F., Morley, R., Finkelstein, A.. Lighting with paint. ACM Transactions on Graphics 26(2) (2007)
17. El-Nasr, M., Horswill, I.: Automatic lighting design for interactive entertainment. ACM Computers in Entertainment 2(2) (2004)

18. Tomlinson, B., Blumberg, B., Nain, D.: Expressive autonomous cinematography for interactive virtual environments. Proceedings of the Fourth International Conference on Autonomous Agents (2000)
19. St-Laurent, S.: Shaders for Game Programmers and Artists. Massachusetts, USA: Thomson/Course Technology (2004)
20. de Melo, C., Paiva, A.: Expression of Emotions in Virtual Humans using Lights, Shadows, Composition and Filters. Proceedings of Affective Computing Intelligent Interaction (ACII07) (2007)
21. de Melo, C., Paiva, A.: Evolutionary Expression of Emotions in Virtual Humans Using Lights and Pixels. J. tao, T.N. Tan (eds.): Affective Information Processing, pp.313–336, Springer Science+Business Media LLC (2008)
22. Hunt, R.: The Reproduction of Colour, Sixth ed. Hoboken, USA: John Wiley & Sons (2004)
23. Quinlan, J.: Induction on decision trees. Machine Learning, 1(1): 813–106 (1986)
24. Witten, I., Frank, E.: Data Mining: Practical machine learning tools and techniques, 2nd ed. San Francisco, USA: Morgan Kaufmann (2005)
25. de Melo, C., Gratch, J.: The Effect of Color on Expression of Joy and Sadness in Virtual Humans. Affective Computing and Intelligent Interaction 2009 Conference (ACII09) (2009)