Director of Photography and Music Director for Interactive Storytelling

Edirlei E. Soares de Lima[1] Cesar T. Pozzer[2] Bruno Feijó[1]
Angelo E. M. Ciarlini[3] Antonio L. Furtado[1]

[1]PUC-Rio, Depto. de Informática, Brasil
[2]UFSM, Depto. de Eletrônica e Computação, Brasil
[3]UNIRIO, Depto. de Informática Aplicada, Brasil

Figure 1: Scenes of our storytelling system with visual effects generated by the Director of Photography agent.

Abstract

The way emotions are expressed in a film has great impact on the viewer's understanding of the narrative. Over the years, filmmakers developed several techniques to enhance the perception of these emotions, such as the photography and the audio editing of the scenes. This paper proposes two cinematography-inspired autonomous agents designed to better express the emotional aspects of interactive storytelling environments. Both the director of photography and music director use support vector machines trained with a cinematography-knowledge dataset to create and manipulate the audio and visual parameters of a runtime dramatization engine, thus increasing the immersion of the viewers in the story.

Keywords: Interactive Storytelling, Cinematography, Support Vector Machine

Authors' contact:
[1]{elima,bfeijo,furtado}@inf.puc-rio.br
[2]{pozzer}@inf.ufsm.br
[3]angelo.ciarlini@uniriotec.br

1. Introduction

In recent years, with the growth of film-based games and game-based films being released, the convergence between games and films already could be foreseen. Recently it became more evident that newer games started to use many cinematography concepts. Game developers and film producers begin to work together.

In the center of this convergence a new kind of media arises. Interactive storytelling is a new medium of digital entertainment where authors, audience, and virtual agents engage in a collaborative experience. While in conventional games stories are essentially used to create challenges for the player, in interactive storytelling stories are expected to surprise and entertain the spectators. As a consequence, the quality of the stories in terms of coherence and dramatic content must be regarded as a prime concern.

Interactive storytelling is closer to a film than a game. At dramatization level many concepts of cinematography are needed to create an interesting content. However, apply concepts of cinematography in an interactive storytelling system is not an easy task. Usually the storyline of a dramatization results from the real-time interaction between the audience and the virtual actors; in some system these actors are also autonomous agents that incorporate a deliberative behavior. So, it is very complicated to create pre-defined scripts, as usually happens in games, to apply cinematography concepts on the control of the camera, actors and environments.

The cinematography concepts are used to enrich the dramatization in a way to tell an interesting and involving story. The objective is to express to the viewers the real emotions and feelings involved in the scenes. The emotion of a scene can be expressed not only through the dramatization of the actors, but also through the audio and the visual aspects. In a film creation the visual aspect and the audio are so important that exist a specific director for each of these functions; the director of photography is responsible for the quality of the photography and the cinematic look of the film. Using his/her knowledge of lighting, lenses, cameras, and film emulsions, creates the appropriate mood, atmosphere, and visual style of each shot to evoke the emotions that the scene must to express [LoBrutto 2002]. The same is done by the music director, however, using music and sounds to express the scenes emotions.

In this work we propose two autonomous agents: the director of photography and the music director for interactive storytelling. The role of the director of photography is to manipulate the visual aspect of the scenes to affect the way how the viewers perceive them and help the viewers to feel the real emotions of the scenes. The music director has the same task; however it uses and manipulates the music in the dramatization to express the scene emotions and present the story in an interesting and involving way.

The paper is organized as follows. Section 2 presents related works on cinematography concepts applied to games and interactive storytelling. Section 3 presents the director of photography and the music director. Section 4 presents the cinematography architecture of our storytelling dramatization system, and section 5 brings a detailed look at the agent's implementation. In section 6, we analyze the performance and accuracy results to demonstrate the efficiency of our approach. Finally, in section 7 we present the concluding remarks.

2. Related Work

Many works have already been done with the objective of applying concepts of cinematography in games. The basic principle of camera positioning employing cinematography knowledge in form of idioms was first explored by Christianson et al. [1996]. These idioms encapsulate the combined knowledge of several personal roles in a traditional filming set and are widely used in research involving camera systems. However, film idioms are only able to solve the problem of direct manipulation of the virtual camera.

Hawkins [2004] proposes a cinematography architecture where the system is divided into different modules or agents. These entities represent the various roles people play in a movie set. Basically the architecture is composed by three logical layers: directors, editors and cinematographers, comparing these with their counterparts in a real movie set, where the director is responsible to propose the film idioms. The editor is in charge of choosing the shots, while the cinematographer takes care of direct camera positioning.

In general, most games can involve the player in the mood of the story, however this process is pre-defined; usually the story of a game follows a pre-defined script and even when a player action can affect the storyline all upcoming events are known and are provided by the script. Only recently the real dynamic drama started to be explored; one example is the game Left 4 Dead. The game uses a dynamic artificial intelligence system for game dramatics, pacing and difficulty. One of the functions of the system is to create the mood and tension with emotional cues, such as visual effects, dynamic music, and character communication, based upon on the player's situation, status, skill and location [Left4Dead, 2010]. However, the system focuses mainly on the control of the game pacing and difficulty, placing enemies in varying positions and numbers in the scenarios.

In interactive storytelling, evolve the viewers in story mood is even more important, but also more difficult. In interactive storytelling the storylines are unknown, so it is very difficult to create pre-defined scripts to manipulate correctly the music or create other visual effects to express the emotion and mood of the scenes. A related work is presented by Melo and Paiva [2005]. The authors propose an approach to express the emotions of a dramatization using the environment – however the work is focused mainly on dialogs and do not consider the cinematography knowledge to express the emotions.

The use of light changes to influence the viewer's psychological and aesthetic perception of 3D scenes was first explored by El-Nasr [2005]. The author presents ELE (Expressive Lighting Engine), an intelligent lighting system for virtual environments that adjusts lights for aesthetic and communicative purposes. The system uses a computational model of dynamic lighting to achieve specific authorial goals, such as emotion evocation and direction of a user's visual attention. The system models tension using simple patterns of increasing contrast in terms of saturation, lightness, or warm and cool colors. The model doesn't include a method for encoding complex lighting patterns as a function of time or narrative parameters.

The use of music in interactive entertainment applications was explored by Casella and Paiva [2001]. The authors describe MAgentA, a system that automatically composes background music choosing an appropriate composition algorithm, among several others stored in a database and associated to a particular emotional state. The system doesn't consider the emotional state of the actors or players; only the emotional state of the environment is used.

3. Director of Photography and Music Director

In a film creation, the director of photography interprets the director vision of the scenes, extracts from it the emotions that each scene must to express, and represents these emotions in a sequence of shots. For this task the director of photography has a set of visual tricks that are used to pass to the viewer certain emotions when they watch the scenes. These visual effects can enhance the story in many different ways, providing the director of photography with multiply options to assist the director in telling an involving story [LoBrutto 2002]. The music director has the same task, however, working after the film shoot. Using his/her knowledge of music creates the mood and atmosphere of each scene based on the emotions and feelings that the scenes must to express. Music is an integral part of the film as it helps to connect the emotional content with the events on the screen.

Films are linear sequences of scenes, even when the story do not is linear. In a film each scene is perfectly planned; the director of photography and the music director know exactly what will happen and the emotions involved in the scenes. Most storytelling systems are not simple sequences of scripted scenes; each time a new storyline is generated and something different can happens. So it is extremely difficult to simulate a director of photography and a music director in a storytelling environment.

3.1 Expressing Emotions on Images

There are several ways to express emotions on images. Most of them are related to the light intensity and color. The light is an important element to enhance the scenes emotions. Scenes with a lot of darkness and shadows increase the impact of emotions such fear and foreboding. Scenes with bright light increase the sense of well-being. Mood is another vital ingredient provided by light. Bright light, like a sunny day, tends to make us feel cheerful, relaxed. If the light is harsh, as it is on a bright hazy day, the mood will tend to be stark and unfriendly. Soft light, like mist, tends to make us nostalgic, wistful, and dreamy. Darkness, like night or an approaching storm, makes us feel worried, frightened, and serious [O'Brien and Sibley 1995].

Color is other important element to express emotions trough images. Color can communicate time and place, define characters, and establish emotion, mood, atmosphere, and a psychological sensibility. Warm colors tend to represent tenderness and humanity. Cool colors represent cold, lack of emotion, and distant feelings. Hot colors represent sexuality, anger, and passion. A monochromatic palette is a limited range of colors that can establish a colorless world, sameness, masked emotion, or a sense of simplicity [LoBrutto 2002].

Camera visual effects also can be used to express specific emotions. Film grain can be used to increase the emotion of fear in dark and shadowy environments. Depth of field can be used to bring the viewers attention to a subject in the scene, increase the scene aesthetic and also bring up the emotions of the subject. The vignette effect when used in dark indoor scenes can pass to the viewer a claustrophobic sensation. Visual effects also can help the director to place the viewers in the story time. A misty blur border around the edge of the screen can represent a flashback scene.

3.1 Expressing Emotions by Music

Music also is a powerful tool to express emotions. In a film the music can change the feel of a scene, bring out the emotions as well enhancing an audience's reaction. Several characteristics have been suggested that might influence the emotion of music. According to Gabrielsson and Lindstrom [2001], major keys and rapid tempos cause happiness, whereas minor keys and slow tempos cause sadness, and rapid tempos together with dissonance cause fear. The choice of instrumentation, whether soothing or obnoxious, will have an effect. Music can set the stage and place spectators in a different world, a different country, or a different time. Music is primarily designed to create a certain atmosphere or feeling for the scenes. It can create a dark and mysterious world, adding tension and desperation to reinforce the seriousness of a situation.

Music can express emotions and feelings so successfully because it works beneath our conscious level. It can cue us as to how to respond to the film or to a particular scene of the film without taking up additional screen time or space. [Miller 1988]. As music can enhance a scene, it can also ruin a scene. Incorrect type of music during a particular scene can nullify the emotions expressed by the actors.

4. Cinematography Architecture

The dramatization module of our storytelling system called LOGTELL [Ciarlini et al. 2005] (figure 2) is based on the architecture proposed by Hawkins [2004]. At higher level we have the Plot Manager who is responsible for the generation of the story scenes based on the user interaction. These scenes are sent to the director agent that interprets the events to execute actions on the environment or send actions to the actors. Finally, the director passes the events to the Editor and Cameraman agents. The Cameraman is responsible to place a set of cameras on the scene based on cinematography rules and suggests these camera angles to the Editor agent. The Editor agent [Lima et al. 2009] incorporates cinematography knowledge to enhance the emotions of the scene choosing adequate camera angle based on the angles suggested by the Cameraman. Finally the Cameraman activates the selected camera to film the scene.

The story of our dramatization is generated by a non-linear prolog planner called IPG [Ciarlini 1999][Furtado and Ciarlini 1999]. The stories are presented as a sequence of scenes and the scenes are composed by a flow of events. Our model assumes cuts between the scenes - this means that not all actor’s steps are dramatized; only the important events are showed to the viewers, as actually happens in films.

The role of the new agents proposed in this work (director of photography and the music director) is to interpret the feelings and emotions of the story events and to create the appropriated mood and atmosphere of each scene.

The scenes and events generated by the Plot Manager do not provide emotional information to the actors. However, to create compelling and interesting stories, feelings and emotions are of great importance. Without this information it would not be possible to establish the real emotion that a scene must to express. To solve this problem our dramatization architecture incorporates a dynamic network structure to represent the actor's emotions and relations.

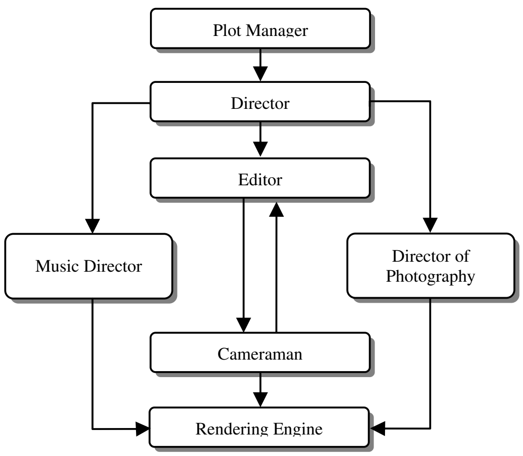

Figure 2: System Architecture.

3.1 Emotions and Relations Network

An emotion is a mental and physiological state associated with a wide variety of feelings, thoughts, and behavior. Emotions are subjective experiences, often associated with mood, temperament, personality, and disposition [Glowsticking 2008]. Social relations are related to the relationships between people and their social world. It assumes that people are engaged in and shaped by multiple relationships, events and influences. Emotions and relations are constantly changing.

To represent the emotions and relations of the actors our system incorporate a structure called "Emotions and Relations Network". In the structure each actor is represented as a node and the relation through a connection between the nodes (Figure 3). The emotional state of an actor i is given by intensity levels of its basic emotions $e_i^k(t) \in [-1,+1]$ and its affective relations $affection_{ij}(t) \in [-1,+1]$ with the other actors. The relations are directed and are not necessarily symmetric:

$$\exists\, i, j\ relation_{i,j}(t) \neq relation_{j,i}(t)$$

By the use of intensity levels in the emotional attributes it is possible to represent a range between two basic emotions, for example, positive values of $e_i^1(t) = happiness_i(t)$ represent happiness, while negative values represent sadness (the opposite emotion).

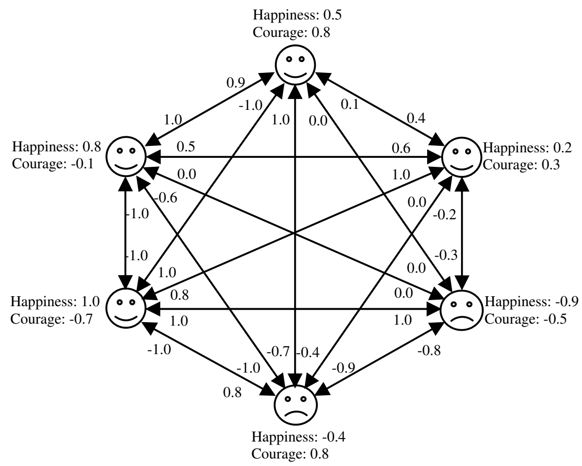

Figure 3: Emotion and Relations Network.

The network is dynamic and updated in real time during dramatization when some event occurs. Each event/action affects the emotions and relationships in different ways. The network has a fixed topology but a variable set of values along the axis of time. In this work, we are using only two basic emotions (happiness/sadness and courage/fear) and the relation between the actors. Even though simplified, this model of emotions and relations fulfills all the needs of our test scenario.

5. Implementation

The agents proposed in this work must be autonomous and intelligent enough to make their own decisions, know when and how manipulate the visual and audio effects to create interesting and coherent scenes. To represent the agent's knowledge we use a support vector machine (SVM) trained to classify the emotions involved in the scenes. SVM, proposed by Vapnik [1995], is a method for pattern recognition of general purpose and are specialized for small sample sets. It consists of a supervised learning method that tries to find the biggest margin to separate different classes of data. Kernel functions are employed to efficiently map input data, which may not be linearly separable, to a high dimensional feature space where linear methods can then be applied. SVM has already been used in our system by the Editor agent to select camera shots in real-time [Lima et al. 2009]. In recent years, SVM have been found to be remarkably effective in many real-world applications such as in systems for detecting microcalcifications in medical images [El-Naqa 2002], automatic hierarchical document categorization [Cai

2004], spam categorization [Drucker 1999], among others.

5.1 Support Vector Machine

The essence of the SVM method is the construction of an optimal hyperplane, which can separate data from opposite classes using the biggest possible margin. Figure 4 shows an optimal hyperplane separating the squares class from the stars class. As can be seen, there can be many others hyperplanes which can separate the two classes, however the SVM use the best solution, the optimal hyperplane.

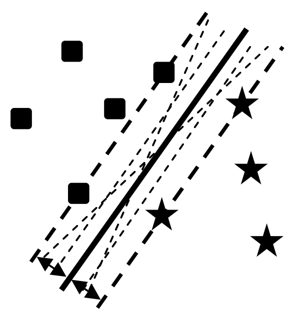

Figure 4: Optimal hyperplane separating two classes.

Suppose the training data $(x_1, y_1), \cdots, (x_l, y_l)$, where each sample $x_i \in R^n$ belongs to a class $y_i \in \{-1,+1\}$. The boundary hyperplane can be as follows:

$$\omega \cdot x + b = 0$$

and the separate margins as:

$$\omega \cdot x + b = +1$$
$$\omega \cdot x + b = -1$$

where, ω is a weight vector; b is a bias; and x is a point in the space R^n.

This set of vectors is separated by the optimal hyperplane if and only if it is separated without error and the distance between the closest vector and the hyperplane is maximal. The separating hyperplane can be described in the following form:

$$\begin{cases} \omega \cdot x_i + v \geq +1, & if \quad y_i = +1 \\ \omega \cdot x_i + v \leq -1, & if \quad y_i = -1 \end{cases}$$

or equivalently:

$$y_i(\omega \cdot x_i + b) \geq 1, \qquad i = 1, \ldots, l$$

The optimal hyperplane is the one that satisfies the conditions and minimizes the function: $\frac{1}{2}\|\omega\|^2$

Vapnik [1995] has shown that, to perform this minimization, we must maximize the following function with respect to the variable α_i:

$$W(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j)$$

subject to: $0 \leq \alpha_i$, $i = 1, \ldots, l$ and $\sum_{i=1}^{l} a_i y_i$

Those $x_i s$ with $0 < \alpha_i$ are termed Support Vectors (SV). The support vectors are located on the separating margins and are usually a small subset of the training data set, denoted by X_{SVM}.

For an unknown vector x_i, its classification corresponds to finding:

$$f(x) = sign\left(\sum_{x_i \in X_{SVM}} \alpha_i y_i (x \cdot x_i) + b \right)$$

where

$$\omega = \sum_{x_i \in X_{SVM}} \alpha_i y_i x_i$$

and the sum is over those nonzero SVs with $0 < \alpha_i$. In other words, this process corresponds to finding which side of the hyperplane the unknown vector belongs.

However, in some cases the classes cannot be linearly separable by a simple hyperplane. To construct the optimal hyperplane in the case when the data is linearly non-separable, SVM uses two methods. First, it allows training errors. Second, it non-linearly transforms the original input space into a higher dimensional feature space by a function $\varphi(x)$. In this higher space, it is possible that the features may be linearly separated.

SVM were originally created for binary pattern classification. For our problem, a multi-class pattern recognition is necessary. To solve this problem, we use the "one-against-one" approach [Knerr 1990] in which classifiers are constructed and each one trains data from two different classes, creating a combination of binary SVMs. In classification use a voting strategy to decide the class of the input pattern. The first use of this strategy on SVM was by Friedman [1996].

5.2 Director of Photography and Music Director Agents

The cinematography-inspired agents proposed in this work analyze and extract features from the actors and environment and then use the SVM to decide which is the best audio and visual effect to enhance the scene emotions. The knowledge of each agent is represented by one SVM. The SVM of the director of photography is trained to solve problems of visual effects classification and the SVM of the music director is trained to solve problems of audio classification.

To use the SVM in our agents we must follow two steps. First, the supervised training process, which is done before the story dramatization, consists in simulating some common scenes and defining the audio and visual effect that better express the desired emotion. The features of these scenes, actors and environment are used to teach the SVM how to proceed in this situation in order to detect similar situations in the future. The second step is the prediction process that is done in real-time during the dramatization by using the knowledge acquired through the training process to predict (classify) an unknown situation.

5.2.1 Training Process

To train our agents, first we need to define the features that are important and can influence the choice of music and visual effects in the dramatization. In our implementation we decided to use 3 features extracted from actors (happiness, courage and affection) and 1 feature extracted from the environment (emotional atmosphere).

These features are extracted from the emotions and relations network. However, the scenes can be composed by several actors, so is necessary to determine the emotion of the entire scene. Melo and Paiva [2005] propose that "the scene emotional state is the average of all the participant characters' emotional states". However this approach is not an appropriate solution. Supposing a scene with two actors with opposite emotional states, the average will be a neutral state, which may be contradictory to real emotion of the scene.

To determine the scene emotion we propose a method based on the filmmakers vision. The scene emotion is established by the emotion intensity and the importance of the actors in the dramatization. In other words, the emotion of the scene is defined by the protagonist actor in the scene with highest emotional intensity. Formalizing the method, let a_{hs} be the set of all happiness/sadness values an actor a may assume along a scene S, and $i \in [0...1]$ the actor importance factor in the dramatization. The function *maxintensity*(*set*) returns the maximum absolute value (considering positive and negative values) of the set. The actor happiness in the scene S can be described as:

$$A_h(S) = maxintensity(a_{hs}) * i$$

Consider a scene composed by a vector of actors *Sa*. The scene happiness can be described as:

$$S_h(Sa) = maxintensity(A_h(Sa_h)) \quad (1)$$

The same formulation applies to the evaluation of the courage/fear of a scene. Considering the a_{cf} as the set of all courage/fear values of an actor along a scene *S*. The actor courage can be described as:

$$A_c(S) = maxintensity(a_{cf}) * i$$

and the scene courage as:

$$S_c(Sa) = maxintensity(A_c(Sa_c)) \quad (2)$$

To determine the scene affectivity we consider the affectivity average of the involved actors in the scene. Considering the a_{cf} as the set values of affectivity of an actor a_1 to an actor a_2 along a scene *S*, the actors affectivity can be described as:

$$A_a(S) = \frac{1}{n}\sum_{i=0}^{n-1} affectivity(a_{cf})$$

and the scene affectivity as:

$$S_a(S) = \frac{1}{n_a}\sum_{i=0}^{n_a-1} affectivity(A_a(Sa_{cf})) \quad (3)$$

The emotional atmosphere is a feature extracted from the location where the scene is happening. The scenario of our dramatization is composed by several regions; each region is divided into several sub-regions and each one has a mood value (between -1 and 1). High values represent happy locations and lower values scary locations.

To define the classes (output) of the SVM we create 5 scene profiles used for both directors (table 1 and table 2). They are labeled by the emotional scene type and are used to describe the audio and visual effects needed to express the desired emotion in the scenes.

Table 1: Director of photography scene profiles

Scene	Visual Effect
Happy Scene	Bright light with white color and high intensity; Distant fog.
Sad Scene	Soft light with gray color and medium intensity; Middle distant fog.
Fear Scene	Dark light with dark gray color and low intensity; Grain film effect in dark scene areas; Dense fog.
Anger Scene	Soft light with red color and medium intensity; Very soft motion blur effect.
Tension Scene	Soft light with dark red color and low intensity.

Table 2: Music director scene profiles

Scene	Audio
Happy Scene	Soundtrack with major keys; rapid tempos; high pitched with large variations.
Sad Scene	Soundtrack with minor keys; slow tempos; narrower range melodies.
Fear Scene	Soundtrack with rapid tempos; dissonance; small pitch variations.
Anger Scene	Soundtrack with minor keys; fast tempos; high pitched.
Tension Scene	Soundtrack with minor keys; ascending melodies; Dissonant harmonies.

After defining the structure of the SVM it is necessary to train effectively the agents. For this task we simulated several scenes and each one was classified according to the vision of a real director of photography and a real music director. Despite the SVM have the same structure, in some cases the vision of a director of photography can be different from the vision of the music director. Therefore, two training databases are created. Thus, each agent has a SVM trained specifically for their task.

The training database is composed of several samples of simulated scenes, each one with the features and the selected profile of audio or visual effects for the simulated scene. These training databases are created once and are used in all future dramatizations.

5.2.2 Predicting Process

With the SVMs trained, the director of photography and music director are able to act as real directors and, based on the previous experience, select in real-time the best audio and visual effect to enhance the scenes emotions.

The prediction process is triggered by the story events. If the agents detect the beginning of a new scene or an event that change the place where the current scene is happening, the directors executes the prediction process. To predict the best audio and visual effect to a scene, each agent executes the following steps:

- Simulate the scene events in the Emotions and Relations Network.
- Extracts the features used by the SVM from the Emotions and Relations Network; these features are the same used to create the training database;
- Applies the extracted features to the SVM;
- Use the SVM output to set the active audio and visual effect emotion profile.

To explain the predicting process we will use an example of a scene generated in our dramatization system. In this scene the knight Brian proposes marriage to the princess Marian, considering that this scene happens after the knight saves the princess from the villain and the previous event increases the princess affection to the knight. The events of this scene are described in table 3.

Table 3: Marriage proposal scene events

Cod	Event
1	Go(Marian, Castle Village)
2	Look(Brian, Marian)
3	Go(Brian, Marian)
4	Talk(Brian, "Good day, my fair lady.")
5	Look(Marian, Brian)
6	Talk(Marian, "Hello, my hero.")
7	Talk(Brian, "How do you find yourself this day?")
8	Talk(Marian, "Fine, and you?")
9	Talk(Brian, "Fine...")
10	Talk(Brian, "If you will answer me a question...")
11	LookAround(Marian)
12	Talk(Brian, "Will you marry me?")
13	Look(Marian, Brian)
14	Talk(Marian, "Yes!")
15	Kiss(Brian, Marian)

The first step to predict the audio and visual effects to example scene is the simulation of the scene events in the Emotions and Relations Network. It is necessary because the emotions and relations are updated by the story events at runtime. However, the agents must to apply the audio and visual effects in the scene before the beginning of the dramatization. Therefore, in the beginning of a new scene the directors use the emotion and relation networks to simulate the scene events in order to generate the set of all emotions and relations along the scene for each actor. With this data set it is possible to extract the features used by the SVM.

The set of happiness/sadness values of the marriage proposal scene is shows in figure 5. With this data set, it is possible to calculate the scene happiness using the equation 1. In our story the Marian importance factor a_i is 0.8 and the Brian is 0.9. This factor is used to ensure that the emotion of the main actors will have more influence over the other actors. In this example, the scene happiness is 0.9.

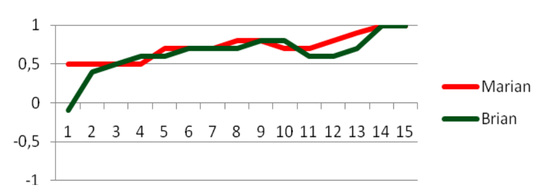

Figure 5: Variation of happiness/sadness in the marriage proposal scene.

The set of courage/fear values is shown in figure 6. To determine the scene courage we apply the set of courage/fear values of the actors to the equation 2. The result is the scene courage as 0.72.

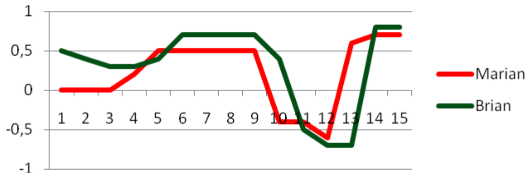

Figure 6: Variation of courage/fear in the marriage proposal scene.

The same is done to determine the scene affection. Appling the set values of affection between the actors (figure 7) to the equation 3 we determine the scene affection as 0.88.

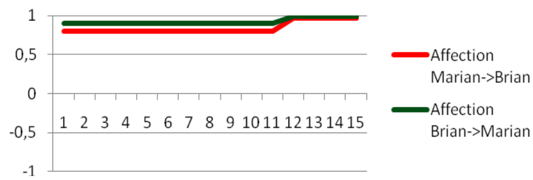

Figure 7: Variation of affection in the marriage proposal scene.

The last feature is the emotional atmosphere. The feature is extracted from the location where the scene in happening. In the example scene the value is 0.8.

With all the features used by the SVM in hands, the agents can classify these features to decide which audio/video profiles should be used. In the example, both directors classify the scene as “Happy Scene”, according tables I and II.

With the emotional profiles selected, the agent’s finally can simulate the audiovisual effects in the dramatization scene. The visual result is showed on figure 8. The other options for this same scene are showed on figure 9. Unfortunately, is not possible to demonstrate the audio result on the paper.

Figure 8: Marriage proposal scene.

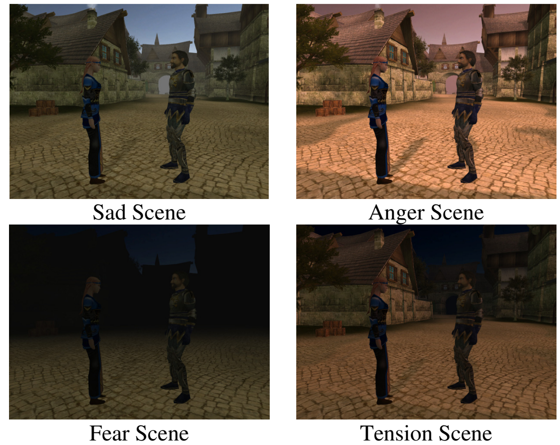

Figure 9: Other visual alternatives for the marriage proposal scene.

5.2.3 Simulating the Audio and Visual Effects in the Dramatization

To simulate the visual effects in the dramatization the Director of Photography agent manipulates the environment light, fog and sky based on the information described by the selected visual profile. However, before change the environment parameters, the agent must decide the way how these parameters will be changed. There are two options: the abrupt and the gradual change. The abrupt way, directly changes the parameters to the desired values. The gradual way, gradually increase/decrease the current environment parameters to the desired values. These two ways to apply a visual effect in a scene are necessary to prevent the director of photography to break the continuity of the scenes. Supposing that a scene ends in a location of the scenario and the next scene begins in the same location. If the Director of Photography selects another visual profile to the next scene and abrupt change the environment parameters, the viewers will realize that something change and will became confused.

To choose the manner how the visual effects will be applied to the scene, the Director of Photography analyses the entire scene to check the location where the scene begins and where it ends. If the scene begins in the same location of the previous scene and a new visual effect was selected, the agent gradually changes the visual parameters of the environment. If the scene begins in a different location, the agent abruptly changes the parameters and shows a black screen for two seconds using a fade-in and fade-out effect.

The soundtracks, different from the visual effects, can be played, and changed between the scenes without breaking the continuity of the narrative. The Music Director agent just uses a fade-in and fade-out effect in the audio channel while change from one track to another to avoid abrupt changes.

6. Results

To test our architecture we run two experiments: first the performance test, to check the necessary time to train the SVM and also the time necessary to predict a new audio and visual effect. The second experiment was the recognition rate, to check the accuracy of the predicted profile.

To test the performance of our proposed solution, first we trained our SVM with a different number of samples and calculate the necessary time to the training process. Next, we use the trained SVMs to predict the audio and visual effects for a sequence of 20 scenes. For each scene, we calculate the necessary time for the prediction process. Table 4 shows the training and recognition performance results with the training set size ranging from 25 to 65 samples, running on a single core of an Intel Core 2 Quad 2.40 GHz CPU. The time for training and recognition is very low and is almost linear to the number of samples.

To test the recognition rate we created 5 training sets with a different number of samples and, for each one, a testing set with half the size of the corresponding training set. The training sets are used to train the SVM and the samples of the current test set are predicted. Correct and wrong predicted audio and visual effects are then computed. Table 4 shows the computed results of this test with the training set size ranging from 25 to 65 samples.

Table 4: Prediction, recognition performance and recognition rate with different training sets

Number of Samples	25	35	45	55	65
Training Time (ms)	199	264	374	451	502
Recognition Time (μs)	30.3	38.6	51.4	60.1	63.9
Recognition Rate	93%	94.6%	95.4%	96%	96.8%

The high accuracy indicates that the result of the classification match (in most cases) with the vision of a real director of photography and music director. However, this doesn't mean that the final result will enhance the stories.

To check the effectiveness of the directors on the quality of audiovisual generated content we test the dramatization system with 20 users that do not knew the story being presented. Half of the users tested the version with the director of photography and music director and the other half test the system without the agents. During the dramatizations, for each user we chose a random scene and interview the users asking the emotion that the current scene is expressing. As result of the test we discover that 90% of the users that tested new the version recognize correctly the emotions of the scenes. On the other hand, only 30% of the users that tested the system without the new agents recognize the emotions. It is important to notice that the system doesn't provide complex animations for the actors; they have only a set of predefined animations for the basic actions and do not express emotions using facial expressions or body movements.

7. Conclusion

In this paper we have presented two cinematography-inspired autonomous agents that use cinematography knowledge to express emotions in storytelling dramatizations. In our tests, our approach has shown that the use of cinematography-inspired agents can effectively improve the stories and express the scenes emotions to the viewers. We demonstrate that the SVM classifier can be effectively used in real-time without high computational coast and ensure a high recognition rate. The supervised learning methodology is useful to train the agents with the knowledge of a real director of photography and music director, ensuring that the agents will follow the cinematography concepts to classify the audio and visual effects to the scenes.

This architecture also can be improved and expanded. In our implementation we decided to use 5 scene profiles to represent the basic scene emotions. These profiles shows to be enough to represent all the possible scene emotions for our purposes. However, for more complex stories, may be necessary to create more emotion profiles, add more features to the SVM, and also improve the emotions and relations network with new attributes. The methodology proposed in this paper can also be adapted to other entertainment applications, such as games and virtual worlds.

Acknowledgements

This work was supported by CAPES/RH-TV-Digital, CAPES/PROCAD, and CNPq. Authors would like to express their gratitude to Laboratório de Computação Aplicada (LaCA) - UFSM.

References

CAI, T., HOFMANN, T., 2004. Hierarchical document categorization with support vector machines. *In: Proceedings of the ACM 13th Conference on Information and Knowledge Management.*

CASELLA, P., PAIVA, A., 2001. MAgentA: an Architecture for Real Time Automatic Composition of Background Music. *In: International Workshop on Intelligent Virtual Agents,* 224-232.

CHRISTIANSON, D.B., ANDERSON, S.E., HE, L., COHEN, M.F., SALESIN, D.H., WELD, D.S., 1996. Declarative Camera Control For Automatic Cinematography. *In: Proceedings of the AAAI '96*. 148-155.

CIARLINI, A., 1999. *Geração interativa de enredos.* PhD thesis, PUC-Rio, Rio de Janeiro.

CIARLINI, A., POZZER, C., FURTADO, A., FEIJÓ, B., 2005. A Logic-Based Tool for Interactive Generation and Dramatization of Stories. *In: Proceedings of the International Conference on Advances in Computer Entertainment Technology.* Valencia, pp.133-140.

DRUCKER, H., WU, D., VAPNIK, V., 1999. Support Vector Machines for Spam Categorization. IEEE Trans. on Neural Networks, Vol 10, Number 5, 1999, pp. 1048-1054.

EL-NASR, M.S., 2005. Intelligent lighting for game environments. Journal of Game Development. 1(2), 17-50.

EL-NAQA, I., YANG, Y., WERNICK, M.N., GALATSANOS, N.P., NISHIKAWA, R.M., 2002. A support vector machine approach for detection of microcalcifications. IEEE Transactions on Medical Imaging, vol. 21, no. 12, pp. 1552-1563.

FRIEDMAN, J., 1996. Another approach to polychotomous classification. Technical report, Department of Statistics, Stanford University.

FURTADO, A., CIARLINI, A., 1999. Operational Characterization of Genre in Literary and Real-life Domains. *In: Proceedings ER'99 Conceptual Modelling Conference.* Paris, France.

GABRIELSSON, A., LINDSTRÖM, E.., 2001. Music and Emotion Theory and Research. Oxford University Press, New York.

GLOWSTICKING, 2010. *Music and Our Emotions*. Available from: http://www.glowsticking.com/articles-andtutorials/culture/75-edm-music/201-music-and-our-emotions.html [Accessed 17 June 2010].

HAWKINS, B., 2004. Real-Time Cinematography for Games (Game Development Series), Charles River Media, Inc., Rockland, MA, USA.

KNERR, S., PERSONNAZ, L., DREYFUS, G., 1990. Single-layer learning revisited: a stepwise procedure for building and training a neural network. Neurocomputing: Algorithms, Architectures and Applications, Springer.

LEFT4DEAD, 2010. *The Director - The Left 4 Dead*. Available from: http://left4dead.wikia.com/wiki/The_Director [Accessed 17 June 2010].

LIMA, E.S., POZZER, C., ORNELLAS, M., CIARLINI, A., FEIJÓ, B., FURTADO, A., 2009. Virtual Cinematography Director for Interactive Storytelling.. *In: Proceedings of the International Conference on Advances in Computer Entertainment Technology.* Greece.

LOBRUTTO, V., 2002. The Filmmaker's Guide To Production Design. Allworth Press, New York.

MELO, C., PAIVA, A., 2005. Environment Expression: Telling Stories Through Cameras, Lights and Music. *In: International Conference on Virtual Storytelling,* 129-132.

MILLER, W., 1988. Screenwriting for Narrative Film and Television, Virgin Books.

O'BRIEN, M., SIBLEY, N., 1995. The Photographic Eye: Learning to See with a Camera, Sterling.

VAPNIK, V., 1995. The Nature of Statistical Learning Theory, Springer, New York.