

Sistema automático de decisión del plano de cámara en videojuegos y entornos virtuales

Carrillo de Albornoz, Jorge: jcalbornoz@fdi.ucm.es, DISIA
Garmendia, Luis: lgarmend@fdi.ucm.es, DISIA
Gervás, Pablo: pgervas@sip.ucm.es, DISIA
Universidad Complutense de Madrid

RESUMEN

El creciente aumento de mundos virtuales, tanto en entornos de entretenimiento como de aprendizaje, supone un auge en el estudio de modelos de automatización de las tareas de gestión de estas plataformas.

En este artículo se propone un módulo para la gestión de cámaras acorde a técnicas cinematográficas utilizando razonamiento aproximado y lógica borrosa.

Atendiendo a parámetros típicos de los guiones de películas, como el dramatismo de una escena, el aspecto narrativo de la misma o el tipo de esta, se decide el plano correcto a partir de unas reglas borrosas que contienen predicados con imprecisión y obteniendo un resultado más suave y de forma automática.

Palabras claves:

Conjuntos borrosos, razonamiento aproximado, plano de cámara, entorno virtual.

1. INTRODUCCIÓN

Actualmente el desarrollo de videojuegos no solo consiste en un conjunto de pruebas presentadas en un entorno vistoso para el usuario, si no que cada vez son más las empresas que utilizan recursos cinematográficos para envolver sus títulos. Para ello se contratan guionistas que desarrollen la historia, actores para el doblaje de los personajes, incluso directores de fotografía que ayuden a diseñar las diferentes pantallas, todo ello enmarcado en un creciente aumento de la potencia de los sistemas gráficos.

Por otro lado se esta abriendo un gran campo de investigación en el estudio de sistemas de aprendizaje mediante entornos virtuales. Cada día son más los workshops en los que se muestran sistemas que enseñan divirtiéndose, para lo cual se usa frecuentemente mundos en 3D.

En la creación de mundos virtuales, elementos como la iluminación, la colocación de los personajes, los movimientos de los mismo, la gestión de las cámaras, etc, se realizan a mano en el proceso de construcción del entorno. Esta tarea tediosa es necesaria para obtener toda la capacidad y expresividad de estas aplicaciones. Sistemas capaces de automatizar estos procesos permitirían un mayor auge en los mismos, haciendo más fácil su creación y mantenimiento.

La teoría de conjuntos borrosos introducida por Lotfi Zadeh [7], que permite modelar información con predicados lingüísticos que contienen imprecisión y modificadores lingüísticos como ‘poco’, ‘mucho’, ‘más o menos’, etc. Las técnicas de inferencia y razonamiento aproximado tienen un gran fundamento teórico que generalizan los métodos de inferencia tradicionales, y permiten pasar con gran facilidad de un mundo abstracto a un mundo geométrico para decidir el posicionamiento de las cámaras en los motores gráficos de películas y videojuegos.

El conocimiento de los expertos en técnicas cinematográficas y entornos virtuales, descrito con lenguaje natural, contiene gran cantidad de términos imprecisos, modelables con conjuntos borrosos y reglas de inferencia borrosas.

Tradicionalmente ha habido dos líneas de investigación en el campo de la automatización de cámaras en entornos virtuales, la línea cinematográfica y la orientada a videojuegos. Los primeros pasos en la primera estuvieron encaminados en la

generación de secuencias de transición suave acorde con la geometría, y el estudio de los grados de libertad en la cámaras virtuales [5]. Posteriormente se han intentado buscar heurísticas y creado lenguajes para el movimiento de cámaras acorde con la composición geométrica de los objetos, para lo que se han usado en los inicios maquinas de estado [4] y más recientemente algoritmos genéticos y redes neuronales.

Actualmente se esta comenzando a estudiar sistemas que basan la elección de las cámaras no solo en aspectos puramente geométricos, si no que se utiliza el contexto de la historia para elegir el mejor plano en cada caso [1]. No solo tratando la selección de escenas, sino intentando mejorar la generación de cámaras acorde con la historia, la línea de tiempo y los elementos de la escena [2].

En el mundo de los videojuegos, en los inicios se siguió la línea de transiciones suaves a través de los objetos geométricos. Posteriormente se ha estudiado sistemas de cámara de primera persona, o incluso que permiten un control total por parte del usuario de la cámara. Pero normalmente se ha dejado de lado el carácter narrativo de la historia del mismo.

Atendiendo a esto, se considera importante el estudio de estos sistemas y se presenta en este artículo una aproximación mediante técnicas de inteligencia artificial y lógica borrosa, para la modelización de los parámetros que entran en juego en la selección de cámaras atendiendo a técnicas cinematográficas, un sistema de gestión de cámaras para entornos virtuales.

El sistema ha sido diseñado mediante el motor de inferencia de lógica borrosa Xfuzzy [11], que permite la definición de conjuntos borrosos que representan variables lingüísticas y la generación de reglas de inferencia basadas en ellas. La aplicación decide que plano es el más adecuado para el conjunto de entradas, que describen el porcentaje de narrativa, dramatismo o descripción que tiene la escena en el guión de la historia. El tipo de escena que esta sucediendo, como dialogo o persecución, y el sentimiento que se quiere transmitir, como soledad, superioridad, etc. La salida se ha modulado como un conjunto borroso sobre la distancia.

2 INTRODUCCION A LAS TECNICAS CINEMATOGRAFICAS

VARIABLES como la distancia del plano, el ángulo, los movimientos de la cámara, las técnicas de luces o el sonido, hacen de la elección de una cámara la correcta o incorrecta. Las diferentes configuraciones de estos parámetros pueden expresar una misma escena desde muy diversos puntos de vista. Así una secuencia de una mujer llorando con poca iluminación, un plano cercano y una música triste de fondo, refleja un sentimiento de tristeza, mientras que la misma escena con un plano de conjunto, una luz diurna y una música mas alegre expone felicidad.

La distancia de la cámara al objeto es una de las variables que permite expresar una misma secuencia de diferentes formas. Como dijo Hitchcock, *“El tamaño de un objeto en una escena es igual a la importancia de este en la historia.”*. Atendiendo a esto, en esta primera aproximación a un sistema de control de cámaras mediante lógica borrosa, centraremos el estudio en esta componente.

Existen diversas aproximaciones de los planos según los diferentes autores, así en [6] se expone una clasificación de 10 diferentes planos atendiendo a la importancia de los objetos, el sentimiento que se quiera expresar, el dramatismo de la escena, etc. Hawkins en [3] realiza una definición de 13 planos diferentes para su sistema multiagente de cámaras en videojuegos. En este trabajo se ha hecho una síntesis de las diversas posibilidades centrándose en las 10 que se describen a continuación y que serán usadas en el sistema.

1. **Gran Plano General:** carácter descriptivo o dramático (soledad, fatalidad, impotencia Omnipotencia). Se usa mucho para acciones masivas como batallas, etc. En la figura 1 se muestra un ejemplo.
2. **Plano General:** valor narrativo, incluso un poco dramático. Encuadre de una persona que ocupa un tercio del encuadre.
3. **Plano de Conjunto:** Encuadra a un grupo de personas. Tiene carácter descriptivo, narrativo y un poco dramático.
4. **Plano Figura:** valor narrativo, potencia el valor dramático. Encuadre de la persona.

5. **Plano medio Largo:** valor mas dramático, pero también puede ser narrativo. Encuadre de persona de pantorrilla para arriba.
6. **Plano Americano:** valor narrativo y dramático. Encuadre de una persona de rodillas para arriba. En la figura 2 se muestra un ejemplo.
7. **Plano medio:** Encuadre de una persona de cintura para arriba. Valor dramático, pero también un poco narrativo.
8. **Plano medio corto:** Encuadre de una persona de busto para arriba. Valor dramático.
9. **Primer plano:** valor expresivo, psicológico, dramático. La cabeza de una persona. En la figura 3 se muestra un ejemplo.
10. **Primerísimo Plano:** Dientes y boca, aunque puede ser sobre objetos y es descriptivo.



Figura 1: Ejemplo Gran Plano General.



Figura 2: Ejemplo Plano Americano.

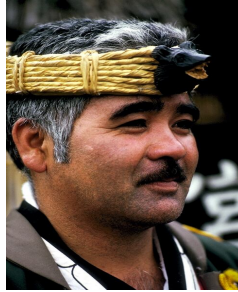


Figura 3: Ejemplo Primer Plano.

Katz expone en [9] que la elección entre los diferentes tipos de planos se basa en la cantidad de dramatismo, narrativa o descripción que posee una escena, en la iluminación y técnicas de sonido. Concretamente la distancia del plano depende de los 3 primeros factores, así como del tipo de escena o el sentimiento que se quiera transmitir. De esta manera un plano cuanto más cercano es expresa por regla general mucho más dramatismo, mientras que un plano muy alejado se usa para describir la escena. El aspecto narrativo se suele mostrar mediante planos de conjunto que abarcan el hilo de lo que esta sucediendo en la historia.

3. UTILIZACION DE LA LÓGICA BORROSA EN EL SISTEMA

Tradicionalmente la lógica ha sido la forma de representar los principios formales y normativos del razonamiento. Con el surgimiento de la lógica borrosa se puede aplicar un razonamiento más parecido al humano, en el que la incertidumbre forma parte del conocimiento y sobre el que se toman decisiones conociendo la información parcialmente. De esta forma se nos permite representar conocimiento común mediante etiquetas lingüísticas, que no se restringen a elementos cuantitativos, mediante un lenguaje matemático a través de la teoría de conjuntos difusos que generaliza la teoría de conjuntos clásica.

Los conjuntos clásicos asignan a cada elemento el grado de pertenencia 1 o 0, según pertenezca o no. En la teoría de conjuntos borrosos [7] un elemento en el universo X , tiene un grado de pertenencia representado en el intervalo $[0,1]$. Mediante los operadores t -Norma y t -Conorma se pueden definir las operaciones de unión, intersección y negación que generalizan a las de la lógica clásica. Con la regla

composicional de inferencia [10] se pueden definir reglas de inferencia mediante los operadores borrosos de implicación.

La lógica borrosa nos permite fácilmente representar parámetros que se definen mediante etiquetas lingüísticas como se ha dicho. De esta manera variables como cuanto de dramática o narrativa es una escena son fácilmente etiquetadas y representadas mediante la lógica fuzzy. También es fácilmente representable la salida del sistema, uno de los diferentes planos mostrados en el apartado anterior, ya que se puede representar mediante una variable borrosa en función de la distancia de los diferentes planos, donde el “Gran Plano General” representa el punto mayor del dominio y el “Primerísimo Plano” representa el menor valor en el dominio.

En esta primera aproximación y atendiendo a las técnicas cinematográficas, a los diferentes planos sintetizados y a toda la documentación consultada, se ha optado por tener 5 entradas:

- El dramatismo de la escena.
- La cantidad de descripción.
- La cantidad de narración.
- El tipo de escena: normal, dialogo, masiva o movimiento.
- El sentimiento a transmitir: neutral, soledad o impotencia.

Las tres primeras variables han sido modeladas como tipos borrosos mediante seis etiquetas lingüísticas, mientras que los dos últimos han sido modelados como entradas que pueden tomar el valor booleano verdadero o falso. En estas dos últimas variables se ha tomado un grupo representativo de las posibles tipos de escenas y sentimientos, ya que englobar todas las posibilidades escapa del ámbito de este artículo, se muestra un ejemplo y se deja para posibles futuros trabajos su ampliación.

4 ARQUITECTURA Y DESARROLLO DEL SISTEMA

El sistema esta compuesto por dos módulos principales que separan la parte grafica del sistema con la parte lógica. El modulo gráfico esta implementado en C++ utilizando el motor gráfico Ogre [8], sobre el que se ha implementado una aplicación que ejecuta historias ya preestablecidas. El sistema lógico esta implementado en Java utilizando la herramienta XFuzzy [11], que permite realizar el diseño de un sistema en

lógica borrosa y exportarlo a Java. La conexión entre ambos módulos se realiza con un puente mediante JNI (Java Natural Interface).

4.1. MODULO GRÁFICO

Esta parte de la aplicación ha sido implementada en C++ usando para ello el motor gráfico Ogre, como se ha dicho anteriormente. Este modulo es el encargado de generar el mundo virtual donde se contara la historia, sobre la que se generaran las cámaras automáticamente, la cual es fija y esta predefinida. Esto quiere decir que los movimientos, diálogos y acciones de los personajes son preestablecidos y generados previamente.

Para ello esta parte consta de un gestor de escenas encargado de gestionar todos los elementos del mundo. Además se ha implementado un sistema de acciones que en cada ciclo del motor comprueba si hay alguna acción que realizar, y si así fuera, la ejecuta. Los diferentes tipos de acciones van desde lanzar una animación, mover un personaje, los diálogos, etc. De esta manera esta aplicación puede generar cualquier historia que se desee mediante la carga de los elementos del mundo y la generación de sus acciones correspondientes.

4.1. MODULO LÓGICO

Esta parte de la aplicación ha sido diseñada y realizada con la herramienta XFuzzy, que permite realizar el diseño de un sistema borroso y genera automáticamente el código del modulo en varios lenguajes de programación, en este caso en Java.

Como se ha comentado el sistema posee 5 entradas, de las cuales 3 han sido modeladas como conjuntos borrosos y otras dos han sido modeladas como valores booleanos. Para poder hacer esta representación en la herramienta XFuzzy se han implementado 10 entradas en el sistema:

- Componente Dramática: variable sobre el tipo Dramatico que representa el % del parámetro, mediante 6 etiquetas lingüísticas.
- Componente Narrativa: variable sobre el tipo Narrativo que representa el % del parámetro, mediante 6 etiquetas lingüísticas.

- Componente Descriptiva: variable sobre el tipo Descriptivo que representa el % del parámetro, mediante 6 etiquetas lingüísticas.
- Sentimiento neutral: variable sobre el tipo Neutral, definida mediante una función singleton que representan el valor 1 o 0.
- Sentimiento de soledad: variable sobre el tipo Soledad, definida mediante una función singleton que representan el valor 1 o 0.
- Sentimiento de impotencia: variable sobre el tipo Impotencia, definida mediante una función singleton que representan el valor 1 o 0.
- Tipo de escena dialogo: variable sobre el tipo Dialogo, definida mediante una función singleton que representan el valor 1 o 0.
- Tipo de escena normal: variable sobre el tipo Normal, definida mediante una función singleton que representan el valor 1 o 0.
- Tipo de escena movimiento: variable sobre el tipo Movimiento, definida mediante una función singleton que representan el valor 1 o 0.
- Tipo de escena masiva: variable sobre el tipo Masiva, definida mediante una función singleton que representan el valor 1 o 0.

La salida del sistema ha sido diseñada mediante un tipo que implementa una función sobre la distancia del plano, representado los distintos tipos de planos descritos anteriormente.

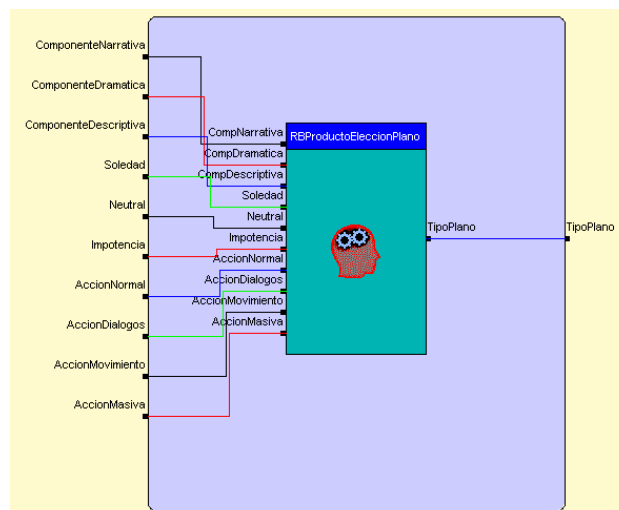


Figura 4: Motor de inferencia del sistema.

Con estas entradas se ha generado un motor de inferencia que mediante las diferentes configuraciones de las entradas del sistema infiere la salida mas adecuada. Para poder realizar esto se han definido 10 reglas de inferencia. Las reglas de inferencia en la herramienta XFuzzy son del tipo IF THEN, con operadores booleanos AND, OR, NOT, etc. Estos operadores según la familia lógica que escojamos producirán una salida diferente para los mismos valores de la entrada.

En la figura 5 se muestra parte de las reglas de inferencia del sistema. Como se puede ver en la reglas intervienen tanto las variables booleanas como las borrosas, formando el conjunto de reglas que inferirán la salida.

En este sistema se han implementado las 3 familias más importantes de lógicas borrosas: la de Zadeh, la del producto y la de Lukasiewicz. Mediante un parámetro de configuración se puede elegir entre una de las 3.

Este sistema ha sido exportado a Java y se ha creado una estructura similar a la estructura de los guiones, con secuencias que contienen escenas y estas a su vez planos., permitiendo una sincronización temporal con la parte grafica y asignando a cada plano los valores específicos para que el motor pueda decidir el plano mas acorde.

| Rule | | | Premise | | Conclusion |
|------|-----|----|----------------------------------------------------------------------------------------------------------------------------------------------|----|------------------------------|
| 0 | 1.0 | if | (AccionMasiva == SI (CompDescriptiva == Todo & CompDramatica <= MuyPoco & AccionNormal == SI & Neutral == SI) (CompDramatica == ... | -> | TipoPlano = GranPlanoGeneral |
| 1 | 1.0 | if | (AccionNormal == SI & Neutral == SI & CompDramatica <= MuyPoco & (CompNarrativa == Todo CompDescriptiva == Mucho) & AccionMasiva ... | -> | TipoPlano = PlanoGeneral |
| 2 | 1.0 | if | (AccionNormal == SI & Neutral == SI & CompNarrativa == Medio & CompDramatica <= MuyPoco & CompDescriptiva == Medio & AccionMasiva = ... | -> | TipoPlano = PlanoConjunto |
| 3 | 1.0 | if | ((AccionNormal == SI AccionMovimiento == SI) & CompNarrativa == Mucho & CompDramatica <= MuyPoco & CompDescriptiva <= Medio & Ac... | -> | TipoPlano = PlanoFigura |
| 4 | 1.0 | if | ((AccionNormal == SI AccionMovimiento == SI) & CompNarrativa == Bastante & CompDramatica <= MuyPoco & CompDescriptiva <= Poco & A... | -> | TipoPlano = PlanoMedioLargo |
| 5 | 1.0 | if | ((AccionNormal == SI AccionMovimiento == SI AccionDialogos == SI) & CompNarrativa <= Medio & CompDramatica == Poco & CompDescripti... | -> | TipoPlano = PlanoAmericano |
| 6 | 1.0 | if | ((AccionNormal == SI AccionDialogos == SI) & CompNarrativa <= Poco & CompDramatica == Medio & CompDescriptiva <= MuyPoco & Accion... | -> | TipoPlano = PlanoMedio |
| 7 | 1.0 | if | ((AccionNormal == SI AccionDialogos == SI) & CompNarrativa <= MuyPoco & CompDramatica == Bastante & CompDescriptiva <= MuyPoco & ... | -> | TipoPlano = PlanoMedioCorto |
| 8 | 1.0 | if | ((AccionNormal == SI AccionDialogos == SI) & CompNarrativa <= MuyPoco & CompDramatica == Mucho & CompDescriptiva <= MuyPoco & A... | -> | TipoPlano = PrimerPlano |
| 9 | 1.0 | if | (AccionNormal == SI & CompNarrativa <= MuyPoco & ((CompDramatica == Todo & CompDescriptiva <= MuyPoco) (CompDramatica == Todo ... | -> | TipoPlano = PrimerisimoPlano |

Figura 5: Reglas de inferencia.

Se ofrece un ejemplo de la aplicación que utiliza estas reglas para decidir el tipo de plano a partir de las entradas descritas por un experto y reflejadas por las reglas de inferencia de la figura 5 en un entorno virtual. El resultado resulta en un movimiento de cámara suave, que enfatiza los momentos más dramáticos y que no requiera intervención manual. Puede observarse un video de ejemplo, cuyo plano es decidido con este sistema en

<http://www.fdi.ucm.es/profesor/lgarmend/SC/Programas/camaraAutomatica.avi>

5 CONCLUSIONES

Se ha presentado un sistema capaz de decidir que plano es el más adecuado acorde a técnicas cinematográficas. Para ello se ha utilizado las posibilidades que brindan los sistemas de lógicas borrosas, ya que permiten una fácil adaptación de parámetros como “la cantidad de narrativa de una escena” o “la cantidad de dramatismo de la misma”. Mediante etiquetas lingüísticas se han modelado estos parámetros, evitando así la pérdida de información típica de los conjuntos clásicos.

Debido a la necesidad de tratar casos especiales como el sentimiento de soledad, que conlleva mucho dramatismo y es un ejemplo típico de un plano cercano pero que se representa con un plano muy alejado, la complejidad del sistema y de las reglas de inferencia se ha multiplicado al tener que introducir variables booleanas para diferenciarlos.

Las lógicas borrosas que han funcionado bien son la de Zadeh [7] y la del producto, sin observarse diferencias notables en la decisión del plano de cámara. La lógica de Lukasiewicz no consigue un comportamiento tan suave como las anteriores, pues cuando los antecedentes de las reglas tienen varias conjunciones, el grado de pertenencia de la premisa tiende a ser uno, y cuando hay varias disjunciones tiende a ser cero, por lo que su resultado es parecido al obtenido usando reglas de inferencia clásicas. El uso del operador de implicación de Mandami [10] para la implementación de las reglas de inferencia ha dado buenos resultados. Un trabajo futuro será utilizar diferentes operadores de implicación, como las residuadas de t-normas, S-implicaciones, QM-implicaciones, etc., comparar sus resultados y encontrar otros operadores que optimicen la suavidad y el comportamiento del plano de cámara.

Mediante este sistema no sería necesario la tediosa tarea de colocación de las cámaras en cada momento, simplemente sería necesario describir cada escena y automáticamente se inferiría el plano más adecuado. Este es uno de los aspectos que dotan de toda la expresividad a un tiro de cámara. Una posible mejora del sistema sería otro motor de inferencia para el ángulo de la cámara, que se podría modelar con las mismas entradas.

Otro aspecto conseguido en este trabajo es la automatización mediante técnicas lingüísticas de la asignación de valores a las distintas entradas en cada plano.

4. REFERENCIAS BIBLIOGRÁFICAS

- [1] ARNAV JHALA AND R MICHAEL YOUNG. A Discourse Planning Approach to Cinematic Camera Control for Narratives in Virtual Environments, 2005.
- [2] ARNAV JHALA AND R MICHAEL YOUNG. Representational Requirements for a Plan Based Approach to Automated Camera Control, 2006.
- [3] BRIAN HAWKINS. Real-Time Cinematography for Games. Charles River Media, 2005.
- [4] CHRISTIANSON DAVID, ANDERSON SEAN, HE LI-WEI, SALESIN DAVID, WELD DANIEL, COHEN MICHAEL (1996). Declarative Camera Control for Automatic Cinematography, 1996
- [5] DRUCKER STEVEN, ZELTER DAVID. Intelligent Camera Control in a Virtual Environment, 1997.
- [6] GERRALD MILLERSON. Tecnicas de Realización y Producción en Televisión, Instituto Oficial de Radio Y Televisión, 1983.
- [7] L.A. ZADEH. Fuzzy sets. Inform. and Control 8, 338–353, 1965.
- [8] Ogre. <http://www.ogre3d.org/>
- [9] STEVEN D. KATZ. Curso de técnicas cinematográficas. Dirección. 2, La planificación de secuencias, Plot, 2000.
- [10] TRILLAS E. AND L. VALVERDE (1985). On mode and implication in approximate reasoning. Approximate reasoning in expert systems. Eds. M. M. Gupta. North-Holland. pp. 157-166.
- [11] Xfuzzy. <http://www.imse.cnm.es>