DAVID SANZ KIRBIS
FRANCISCO JAVIER SANMARTÍN PIQUER

Facultad de Bellas Artes. Laboratorio de Luz.
Departamentos de Pintura y Escultura.
Universidad Politécnica de Valencia.
dasankir@upv.es
frasanpi@pin.upv.es
www.cvcinema.blogs.upv.es

Aplicación de las técnicas de visión artificial en el campo del audiovisual

vol 9 / Dic.2013 197-207 pp Recibido: 07-09-2013 - revisado 15-09-2013 - aceptado: 30-10-2013

Arte y políticas de identidad

ISSN edición impresa: 1889-979X. ISSN edición web (http://revistas.um.es/api): 1989-8452

APPLICATION OF COMPUTER VISION TECHNIQUES IN AUDIOVISUAL

ABSTRACT

In this case study about audiovisual language innovative computer vision, are used to search for other artistic applications of cinematographic resources, both in the acquisition and the editing. Proposed here is that the tools and computer vision techniques can be used to generate new audiovisual languages in the field of interactive cinema. To prove this, a number of consecutive steps were taken, that have allowed to lead a progressive research based on a study of artistic references. From this study we have synthesized a number of key concepts identified both in the artistic pieces and in the critical debates referenced. With these concepts, a series of experiments have been conducted prior to the development of the prototypes finally integrated into the system exposed to the public after completion. A number of conclusions have been extracted as evaluation of the overall results of the study. As a result: firstly, certain relationships between sound and image were obtained, that are unique in the use of resources such as the change-rate level, the sound-image interdependence; and secondly, it has been demonstrated that partially depositing creative responsibility of audiovisual on an automatic device can provide new aesthetic experiences to the viewer.

Keywords
Computer Vision, Experimental cinema, installation art, interactive art, Vjing.

RESUMEN

En este estudio práctico sobre lenguaje audiovisual se utilizan sistemas de visión artificial para buscar otras aplicaciones artísticas de los recursos cinematográficos tanto en la adquisición como en el montaje. Se propone que las herramientas y técnicas de visión artificial pueden ser utilizadas para generar nuevos lenguajes audiovisuales en el campo del cine interactivo. Para demostrarlo se siguen una serie de pasos consecutivos que han permitido llevar una investigación progresiva sobre la base de un estudio de referentes artísticos. A partir de este estudio se han sintetizado una serie de conceptos clave identificados tanto en las obras artísticas como en los debates críticos referenciados. Con estos conceptos se han elaborado una serie de experimentos previos al desarrollo de los prototipos que componen el sistema expuesto al público. Se extrajeron una serie de conclusiones a modo de evaluación de los resultados globales del estudio. Como resultado, por una parte se han obtenido relaciones entre sonido e imagen que son singulares en el empleo de recursos como el ritmo de cambio de plano, la interdependencia sonido- imagen; por otra parte se ha demostrado que depositar parcialmente la responsabilidad creativa de audiovisuales en un dispositivo automático puede proporcionar nuevas experiencias estéticas al espectador.

Palabras Clave
Visión por computadora, Cine Experimental, instalación, interactivos, Vjing.

1 INTRODUCCIÓN

Este estudio forma parte de la investigación "APLICACIÓN DE LAS TÉCNICAS DE VISIÓN ARTIFICIAL EN REALIZACIONES AUDIOVISUALES"[1], y supone una aportación a los trabajos que se están realizando en la actualidad sobre nuevas estrategias de lenguaje audiovisual propiciadas por las tecnologías de la imagen. Esta aportación es posible gracias al empleo de sistemas innovadores de hardware y software de visión artificial a la hora de buscar otras aplicaciones artísticas de los recursos cinematográficos tanto en la adquisición (encuadre, movimientos, fuera de campo, etc.) como en el montaje (fundidos, cortes, elipsis, metáforas, etc.). Las investigaciones que se desarrollan en el área de visión artificial están dando lugar a importantes aplicaciones en el campo del cine y el arte multimedia. Un ejemplo es el software de visión artificial y efectos especiales por el que la empresa "23D", spin-of de la Universidad de Oxford, ganó un Emmy en 2002. Entre estas investigaciones se encuentran las desarrolladas por artistas tan importantes en el panorama internacional como Lars Von Trier o Lev Manovich. Éstos estudian cómo las tecnologías de la información pueden usarse para cambiar profundamente la forma en que se genera el discurso audiovisual. Una de las acciones diferenciales es la inclusión de software de automatización en los procesos de decisión que determinan los parámetros de los procesos de adquisición, montaje y proyección del material fílmico.

2 HIPÓTESIS Y METODOLOGÍA

La hipótesis que aquí se propone es que las herramientas y técnicas de visión artificial pueden ser utilizadas para generar nuevos lenguajes audiovisuales en el campo del cine interactivo. La trascendencia de esta afirmación queda patente si se visualizan sus implicaciones potenciales, como, por ejemplo, un plató de cine donde las labores de producción (dirección, manejo de cámara, luces, micrófonos, etc.) y posproducción (selección, montaje, etalonaje, etc.) son llevadas a cabo, en tiempo real, por un sistema informático inteligente. Para demostrar esta hipótesis hemos tenido que cubrir los siguientes objetivos en la investigación: Recopilar y clasificar información acerca de los avances tecnológicos de los medios de expresión audiovisual y sus repercusiones en la transformación de los propios lenguajes audiovisuales, relacionando

Figura 1. Presentación de CVcinema en la Facultad de BB.AA. de la Universidad de Málaga. 11 de octubre de 2012. *Imagen propiedad de los autores del artículo.*

los parámetros involucrados en cada avance con las técnicas utilizadas y el contexto artístico y social del momento. Estudiar la relación ojo humano – ojo máquina que se deduce de la información recopilada en el punto anterior, extrayendo conclusiones que permitan proyectar una aplicación de las herramientas de visión artificial al campo audiovisual con el objetivo de abrir nuevos horizontes en la expresión audiovisual. Diseñar experimentos en los que se usen las herramientas de visión artificial tanto para generar material audiovisual como para componer dicho material en busca de nuevos discursos.

Tanto de las pruebas de campo de los estudios experimentales como de la exposición del prototipo final se extrajeron una serie de conclusiones a modo de evaluación de los resultados globales del estudio.

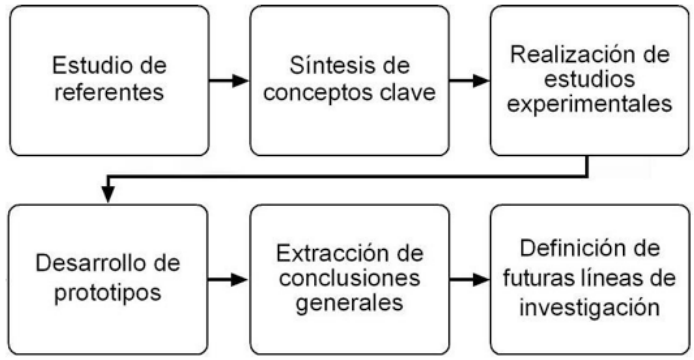

Figura 2. Esquema de la metodología empleada. *Imagen propiedad de los autores del artículo.*

3 RESULTADOS

De entre todas las posibles configuraciones del sistema, hemos desarrollado dos prototipos.

Prototipo 1
En este experimento hemos aplicado al campo audiovisual varios algoritmos de visión artificial ampliamente usados en áreas como la gestión de la calidad de productos industriales, la videovigilancia o la robótica, entre otros. Además hemos combinado dos de estos algoritmos para verificar la efectividad de aplicarlos a la vez. Por último hemos aplicado también una variante elaborada más novedosa gracias a los avances de la investigación en este campo en los últimos años. El desarrollo del prototipo consiste en el diseño y la programación de un sistema informático que recibe como input un flujo de imágenes de entrada (ya sea en vivo, por medio de una cámara de vídeo de alta definición o en diferido, desde archivos de vídeo almacenados previamente), realiza un procesado de los fotogramas de entrada como imágenes independientes (aplicando los algoritmos y el mapeo establecidos), y produce como output un flujo de imágenes de salida. El mapeo o correspondencia entre la entrada y la salida consiste en un reencuadre a modo de mirada, similar al que haría un director de cine al seleccionar la orientación de la cámara y el tamaño de plano para fijar la atención en un elemento o zona del escenario. Básicamente se trata de determinar qué porción de universo visible se mostrará a la salida del sistema, dentro de las posibilidades técnicas disponibles. Este mapeo parte de

la reflexión de que, en sus campos de aplicación habitual, las técnicas de visión empleadas "fijan su atención" sobre determinadas zonas de la imagen analizada, según su "objeto de interés" (brillo, movimiento, etc.). La combinación de técnicas de visión artificial que utilizamos producen información acerca de diferentes zonas de la imagen que captan. Por ejemplo, en el caso de los algoritmos utilizados para la detección de movimiento, obtenemos una lista de datos que describen zonas o blobs de la imagen que se han movido. Esta información se clasifica generalmente por cantidad de movimiento y por tamaño de la zona en movimiento. Esto permite realizar varios tipos de mapeo o correspondencia entre la información proporcionada por los algoritmos de visión artificial y la imagen de salida. Podemos, por ejemplo, establecer que se encuadre la zona con mayor cantidad de movimiento o la mayor zona en movimiento. Además, al obtener datos de más de una zona/ blob, también aparece la posibilidad de utilizar los datos de varios blobs en conjunto para establecer el mapeo, por ejemplo, priorizando el encuadre

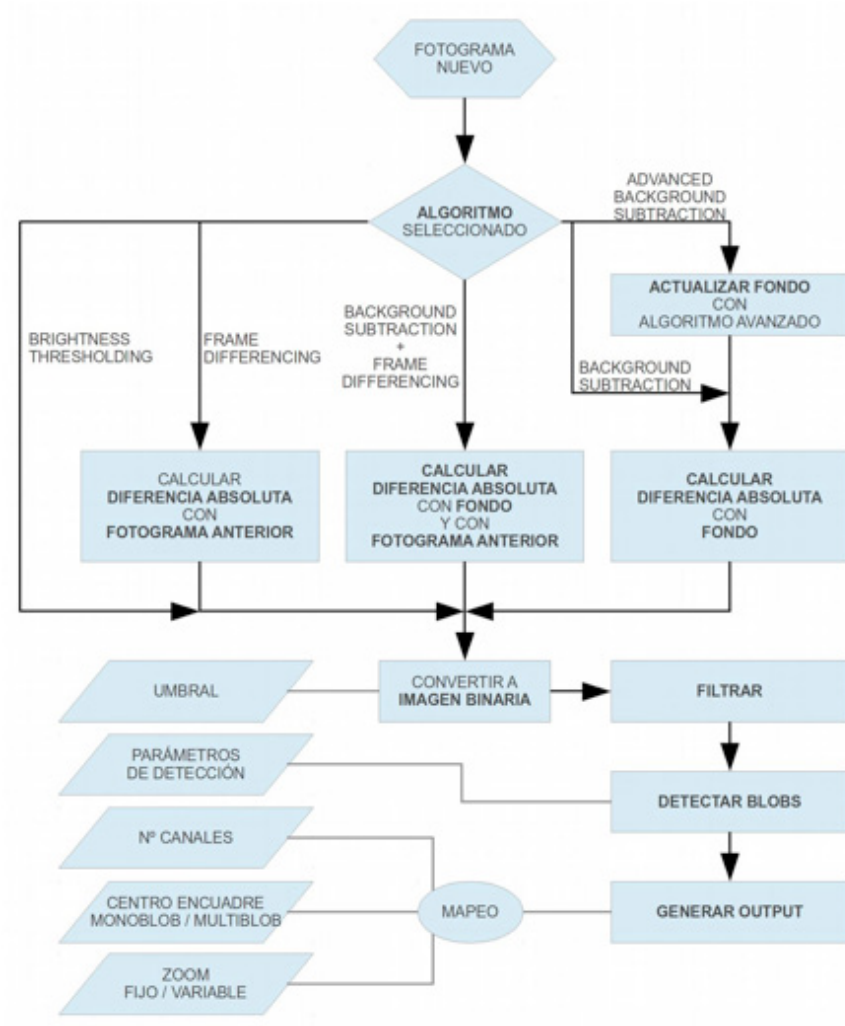

Figura 3. Diagrama de flujo del prototipo 1. *Imagen propiedad de los autores del artículo.*
En total con los cinco algoritmos implementados y con las opciones de mapeo establecidas se generan un total de 30 posibles combinaciones que generan otras tantas salidas.

en el centro de un grupo de blobs o generando un canal de salida para cada uno de ellos. Otra de las opciones de las que disponemos es la de utilizar datos como el tamaño de los blobs para realizar el reencuadre, de manera que se genere una especie de zoom variable al cambiar el tamaño de la zona; o por el contrario podemos escoger un tamaño zoom fijo y que solamente cambie el punto de la imagen de entrada en la que de centrará el encuadre de la imagen de salida. En nuestro prototipo 1 hemos experimentado con diferentes mapeos para evaluar los resultados que se producen. Dado que los resultados del análisis varían con cada cambio en los fotogramas de entrada, el vídeo producido por el sistema a la salida experimenta un reencuadre cambiante, generando a los ojos del espectador efectos de movimiento aparente de cámara (panorámicas y zoom), de manera análoga a los que haría un operador de cámara, o efectos de montaje en corte, de manera análoga al trabajo de un montador de cine. La estrategia de realizar un reencuadre partiendo de una imagen de entrada con encuadre fijo permite comenzar a experimentar de manera inmediata, al evitar el desarrollo o la programación de un sistema mecánico de movimiento de cámara y los posibles problemas de implementación que pudieran surgir. Resulta evidente que las imágenes de entrada serán más favorables cuanto mayor campo de visión ofrezcan, o sea, tomadas con objetivos de focales cortas, ya que un campo amplio equivale a un rango de movimiento mayor en los reencuadres. También resulta conveniente partir de una imagen con la mayor resolución posible, pues la forma de reencuadrar una zona de la imagen es por medio de la selección de un subconjunto de píxeles, cuya posterior ampliación va en detrimento de la calidad de la imagen de salida. Los pasos comunes para obtener la imagen de salida a partir de la de entrada son:

1) Aplicación de la técnica propia de cada algoritmo.
2) Aplicación de los filtros genéricos necesarios para favorecer la detección de las zonas de interés, concretamente los filtros de umbral, erosión y dilatación.
3) Aplicación de la herramienta de detección de blobs de OpenCV.
4) Recorte de la imagen original de entrada en base al mapeo seleccionado.

Prototipo 2

El lanzamiento al mercado, a finales de 2010, del sensor 3D para videojuegos “Kinect” por parte de Microsoft, y el posterior descifrado de su flujo de datos por la comunidad de desarrolladores independientes ha supuesto una revolución en el mundo de la visión artificial. No es que no existiesen con anterioridad métodos y dispositivos capaces de proporcionar mediciones tridimensionales de un espacio, sino que cualquier alternativa resultaba o demasiado lenta e imprecisa como para permitir un uso eficaz, o demasiado costosa como para que la mayoría de grupos de investigación invirtiesen en ellas. Ahora, por el contrario, el sensor se puede adquirir en cualquier tienda de electrónica a bajo coste y proporciona datos 3D en tiempo real a 60 fotogramas por segundo, con una precisión de milímetros en un rango de 0,5 m a 4,0 m. Desde el punto de vista de un sistema informático el incremento de información que supone la sustitución de una cámara convencional por una 3D es, valga la comparación, como el cambio de visión que pueda experimentar una persona con falta de visión cuando se pone unas gafas. Antes cabía la posibilidad de detectar la presencia de un rostro en posición frontal ante la cámara y obtener su posición (x,y) en el cuadro. Ahora podemos detectar la postura y la posición de una persona en una habitación, los gestos que realiza, a donde mira y cómo interactúa con los objetos del entorno. Como es normal, ante tal elenco de posibilidades a bajo coste, nuestro equipo de investigación decidió incluir sin dilación este sensor en el repertorio de dispositivos a utilizar en la parte experimental del proyecto.

Pasar de dos dimensiones a tres puede resultar fácil para un humano adulto que lleva entrenando sus sentidos desde su nacimiento. Desde la posición del investigador en arte electrónico supone un reciclado completo de conocimientos. Al igual que cada nueva línea de investigación se apoya en la anterior, la programación de un sistema informático en el contexto open source se apoya en las funcionalidades puestas a disposición pública gracias a los generosos aportes de otros investigadores. Sin embargo, cuando aparecen nuevos dispositivos o técnicas la bolsa de utilidades de programación es escasa, es necesario aprender a analizar y utilizar los datos obtenidos por uno mismo. En este sentido, nuestro equipo ha realizado un considerable esfuerzo para ponerse al día y estar lo más cerca del estado del arte, dentro de nuestras posibilidades y en el contexto en el que nos movemos. Uno de los principales experimentos desarrollados es el resultado de la extensión lógica al mundo 3D de las técnicas empleadas en el análisis 2D de flujos de vídeo expuestas anteriormente. Al igual que con el algoritmo de sustracción de fondo aplicado a imágenes 2D, se trata de mantener localizada la posición de las figuras que se introducen en la escena y que se diferencian del fondo. Para ello hemos adaptado a datos en tres dimensiones la sustracción de fondo (background subtraction) y la detección de blobs para analizar la nube de puntos obtenida con el sensor Kinect. Una de las principales ventajas de utilizar datos en 3D es que se evitan los problemas que ocurrían en el análisis 2D cuando coincidía el color de la figura con el del fondo, dificultando o haciendo imposible la discriminación. Ahora no se comparan colores sino las tres coordenadas XYZ de lo puntos en el espacio de las superficies que capta el sensor. Una vez detectadas las posiciones 3D de interés, es posible utilizarlas para controlar dispositivos físicos, como por ejemplo hacer que la iluminación de un foco motorizado siga constantemente a una persona en concreto, o mover cámaras robóticas para encuadrar a esa persona.

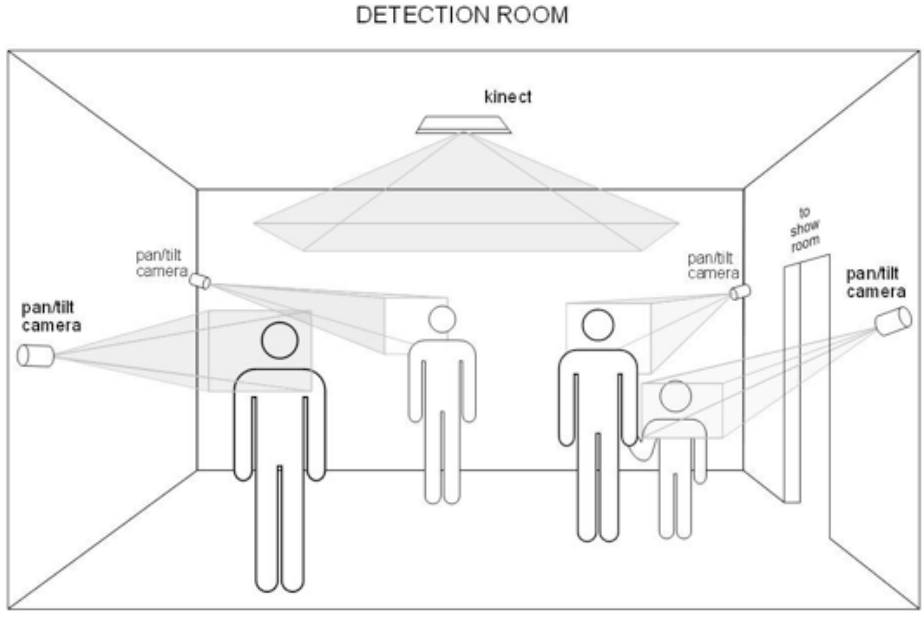

Figura 4. Esquema de disposición de los dispositivos de entrada y sensor para el Prototipo 2. Usando un sensor 3D para localizar un punto de interés en el espacio podemos hacer que otros dispositivos como cámaras o focos apunten a esa posición. *Imagen propiedad de los autores del artículo.*

Como hemos dicho, para la detección hemos aplicado la técnica de Background Subtraction a las tres dimensiones. Después de discriminar los datos correspondientes a los elementos del fondo de la escena, agrupamos los puntos restantes por proximidad, identificando las diferentes masas de puntos y sus características físicas. Para ello hemos desarrollado la aplicación de detección de masas o clusters de puntos 3D adyacentes a partir de un algoritmo genérico de extracción euclídea de clusters[2]. La decisión de implementar nosotros mismos el algoritmo en lugar de utilizar una implementación genérica disponible en la biblioteca PCL tiene dos propósitos. En primer lugar, tener independencia de la biblioteca PCL que en el momento de desarrollo del prototipo no era fácilmente accesible desde openFrameworks (en la actualidad existe un addon que facilita esta tarea). En segundo lugar optimizar la detección aprovechando que usamos un sensor que produce una nube de puntos 3D ordenada en una matriz de dos dimensiones x-y, de manera que la búsqueda de puntos adyacentes se puede restringir a estas dimensiones, mejorando el rendimiento de la aplicación (en el momento de desarrollo del prototipo el algoritmo de búsqueda de blobs de PCL no hace distinción entre nubes de puntos ordenadas y desordenadas).

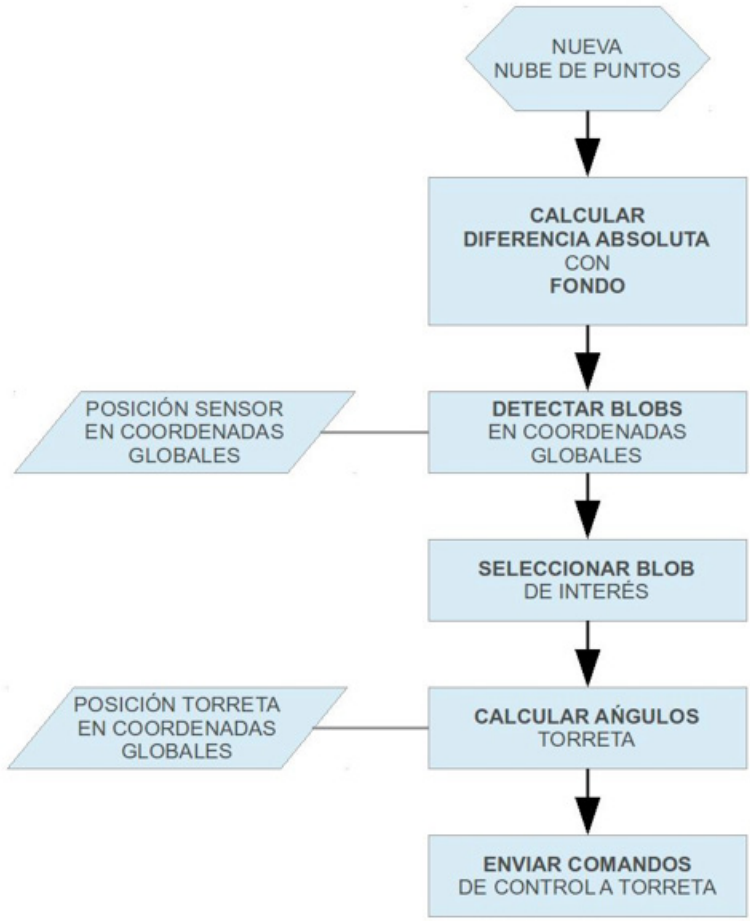

Figura 5. Diagrama de flujo de la aplicación del Prototipo 2. Imagen propiedad de los autores del artículo.

4 CONCLUSIONES

El estudio y los experimentos realizados han dado lugar a una serie de resultados y conclusiones que se resumen a continuación.

- Relaciones entre sonido e imagen
En las pruebas de campo con los prototipos desarrollados se ha observado que se generan varios tipos de relaciones entre las imágenes y los sonidos (ya sean estos más o menos musicales o aquellas más o menos abstractas).

-En los conciertos en vivo realizados en directo, el sistema actúa como complemento a la actuación de un grupo musical, a modo de videojockey. Durante las distintas actuaciones que este tipo de relaciones son bastante sutiles, llegando a producir una especie de sinestesia en la que se intuye que el fluir de las imágenes y del sonido tiene cierta relación, pero sin llegar a los niveles estructurados del montaje rítmico.

-En la exposición CVCinema, durante el concierto realizado para la inauguración, en algunos momentos la sincronización entre imagen y sonido es perceptible, pudiéndose relacionar los cambios que se producen en la pantalla con cambios en el sonido como golpes de percusión, y viendo un cierto paralelismo entre el ritmo de cambio de la imágenes y el nivel de cambio del sonido. Durante el resto de días de la exposición la mayor parte del tiempo los espectadores no se apercibían de la relación entre el sonido y la imagen, ya que mientras el micrófono no captase variaciones importantes, la mezcla de imágenes mostraba cambios suaves.

-En el cortometraje “At one's fingertips” la relación entre el sonido y la imagen es opuesta al caso anterior. El audio es dependiente de la imagen, siendo una parte del mismo generado a partir de ella, mientras que otra parte es independiente de la imagen y se regula por medio de otros parámetros.

-Creatividad computacional
Siguiendo una búsqueda de conocimiento artístico, realizada con un planteamiento que encaja en las categorías de creatividad formuladas por Boden (1977), hemos incorporado algunos de los resultados del sistema en audiovisuales en las pruebas de campo, ampliando el espacio conceptual del cine experimental en caso y de los visuales en realizaciones de conciertos en directo en otro caso, lo cual entraría dentro de la categoría de creatividad exploratoria. Al juntar disciplinas habitualmente separadas como la visión artificial y los audiovisuales estamos propiciando la creatividad combinacional. Un análisis de lo que el sistema desarrollado considera “interesante”, comparado con lo que los humanos solemos valorar estéticamente puede llevarnos a identificar nuevos parámetros estéticos positivos (en el caso de que ambos, computador y humano, coincidan) o negativos (en el caso de que no coincidan), con la posible transformación del espacio conceptual estético de los audiovisuales (creatividad transformacional).

-Difusión de resultados
Los experimentos de la investigación desarrollados han tenido proyección pública en diferentes formatos y contextos. La primera prueba pública de utilización del sistema se materializó en el cortometraje de cine experimental “at one's fingertips”, exhibido en mayo de 2012 en el

Experimental Film Festival Portland 2 y en el Montreal Underground Film Festival 3, mientras que el sistema "CvCinema" se instaló con todos los prototipos desarrollados en la sala de exposiciones de la Facultad de BB.AA. de la Universidad de Málaga y se expuso del 11 al 31 de octubre de 2012. Aparte de los festivales en los que se proyectó el cortometraje "At one's fingertips" y de la instalación "CvCinema" expuesta en Málaga, durante el desarrollo de esta investigación se han hecho públicos tanto los resultados parciales de los experimentos y prototipos como el resultado final materializado en el sistema CvCinema, en una serie de eventos y/o medios de comunicación.

Bibliografía

Abel, R., Ed. (2005). *Encyclopedia of Early Cinema*, Routledge.

Bazin, A. (2001). *¿Qué es el cine?,* Ed. Rialp.

Boden, M. A. (1977). *Artificial Intelligence and Natural Man*, Hassocks, Sussex: The Harvester Press.

Cheroux, C. (2009). *Breve historia del error fotográfico*. Ediciones Ve. México.

Chion, M. (1993). *La audiovisión*. Paidos Comunicación. Barcelona.

Colton, S., y Wiggings, G. (2012). *Computational Creativity: The Final Frontier?, Frontiers in Artificial Intelligence and Applications, Volumen 242*. ECAI.

Eisenstein, S. (2005). *La forma del cine*. Siglo XXI Editores.

Fuller, B. (1969). *Operating Manual for Spaceship Earth.* Carbondale, Ill.: Southern Illinois University Press.

Iglesias, R. (2012). *La robótica como experimentación artística, tesis doctoral, Facultad de BB.AA., Universidad de Barcelona.*

Isaac, A. (1998). *Cuentos Completos II de Isaac Asimov*. Ediciones B. Santiago de Chile.

Jefferson, G. (1949) *The mind of mechanical man. British Medical Journal.*

Koetsier, T. (2001). *On the prehistory of programmable machines: musical automata, looms, calculators, Mechanism and Machine Theory 36.*

Majid Al-Rifaie, M., y **Bishop, M.** (2012) *Weak vs. Strong Computational Creativity, 5th AISB Symposium on Computing and Philosophy, University of Birmingham, UK*.

Manovich, L. (2001). *The Language of New Media*. MIT Press.

Y**oungblood, G.** (1970). *Expanded Cinema*. P. Dutton & Co., Inc., New York.

NOTAS

1. Esta investigación, financiada por la Universitat Politècnica de València, pertenece a la línea de investigación "Tracking Video" del grupo de investigación Laboratorio de Luz (www.laboluz.org) de la Facultad de BB.AA de la UPV, y ha contado con un investigador contratado dentro del programa nacional de becas FPU del Ministerio Español de Educación, Cultura y Deporte.

2. Euclidean Cluster Extraction. Recurso en línea, disponible a 19 de noviembre de 2012. http://www.pointclouds.org/documentation/tutorials/cluster_extraction.php