Relevamiento de herramientas de software destinadas a la captura, análisis y síntesis del gesto corporal

Autor: Tarcisio Lucas Pirotta

Proyecto de investigación "Desarrollo en multimedia del arte bio-generativo y los sistemas de captación del gesto y la emoción humana." – Director Carmelo Saitta - Facultad de Bellas Artes de la Universidad Nacional de La Plata (Argentina) – Diag.78 Nº 680 Ciudad de La Plata Tel. 54-221-432-0532
tarcisiopirotta@hotmail.com

Palabras claves

Captura, movimiento, gesto, software

Resumen

La captura y análisis del gesto corporal se ha convertido últimamente en una importante herramienta para la interacción del hombre con la computadora. El avance en lo estudios sobre HCI (human interface interaction) ha llevado al desarrollo de diversas técnicas y herramientas cada vez mas precisas, que brindan al hombre una manera natural, intuitiva y efectiva de comunicarse con la computadora a través de su gestualidad.

1. Introducción

Abordaremos específicamente el desarrollo de diversos softwares y hardwares destinados al análisis y síntesis del gesto y que permiten el control de sonido y video en tiempo real.

Para abordar esta temática debemos tener en cuenta los elementos que intervienen en la captación. En primer lugar se puede diferenciar entre los dispositivos de entrada, el hardware que permite la captación, y los dispositivos de análisis, el software que habilita la comprensión de los datos y su posterior utilización para el control de imagen y sonido.

2. Hardware

Son los dispositivos físicos que intervienen en la captación de movimiento, generan información que representa las medidas físicas del movimiento capturado y constituyen diversos sistemas según su naturaleza.

Podemos clasificarlos en:

Mecánicos/ Protésicos

Acústicos

Magnéticos

Ópticos

2.1. Sistemas Mecánicos/ Protésicos

Este es uno de los más primitivos métodos para la captura de movimiento en diferentes partes del cuerpo.

Podríamos considerarlo un método ideal sino fuera por la complejidad mecánica que requiere y las limitaciones que esto puede causar al usuario, ya que el método se basa en una serie de prótesis o armaduras que son colocadas sobre el cuerpo. Este método puede utilizar sistemas de detección simples como los interruptores que captan dos estados (on/off) así como también otros más complejos como potenciómetros o sliders colocados en las articulaciones del cuerpo.

Estos dispositivos están interconectados a través de una serie de codificadores radiales y lineales que se conectan a su vez a una interface que los lee simultáneamente. Finalmente a través de una serie de funciones trigonométricas, el movimiento del usuario puede ser analizado.

Por lo general se utiliza para la animación de personajes virtuales.

2.2. Sistemas Acústicos

Este método se basa en la captación del movimiento a través del sonido. Se puede utilizar para determinar la posición del usuario, su movimiento, trayectoria, un determinado gesto corporal. Este método involucra el uso de una triada de audio-receptores. Un arreglo de transmisores es colocado en varias partes del cuerpo del usuario. Estos transmisores son secuencialmente disparados y cada receptor calcula el tiempo que tarda el sonido en desplazarse desde cada transmisor. La distancia calculada es triangulada para determinar un punto en el espacio. Lo cual en sumatoria nos determina un mapeo general de la posición del usuario.

Una de las ventajas de este método es que carece de problemas de oclusión, normalmente asociados a los sistemas de captación ópticos. De todas maneras, hay algunos factores negativos, podríamos decir que los cables pueden ser un estorbo para determinadas acciones o performances. En segundo lugar, el sistema no soporta la suficiente cantidad de transmisores para captar en detalle la personalidad y movimientos sutiles del usuario. En tercer lugar el área de captura, que es limitada por la velocidad del sonido y el numero de transmisores. Y por ultimo el sistema puede a veces ser afectado por reflexiones sonoras.

2.3. Sistemas Magnéticos

Este es un método popular usado sobre todo para capturas en performances.

Involucra el uso de transmisores locales centralizados, y de una serie de receptores los cuales se colocan en varias partes del cuerpo del usuario/performer. Estos receptores son capaces de determinar su relación espacial respecto del transmisor. Cada receptor es conectado a una interface que puede ser sincronizada para prevenir la deformación de la información. El resultado consiste en la posición y orientación de cada receptor.

Como en el caso de la captación por dispositivos mecánicos, este método también es utilizado para la animación de personajes.

Este método comparte junto con el método acústico la ventaja de la falta de problemas de oclusión, aunque también comparte factores negativos como la molestia de los cables, la limitación de los receptores y el área de captura. A demás siendo un sistema magnético, es afectado por elementos contundentes de metal próximos al área de captación.

Actualmente como una variante dentro de los dispositivos magnéticos, podríamos incluir aquellos que utilizan tecnología de radio frecuencia y que emplean el uso de RFID tags (Radio Frequency Identification). El sistema es similar pero el uso de los tags brinda mayor libertad al usuario/performer, si bien debe llevar en alguna parte del cuerpo el receptor (tag) no se necesaria la utilización de cables.

La superficie sensible desarrollada por Proximity Lab (http://www.proximitylab.org/) es un claro ejemplo de esto. El dispositivo consiste en una plataforma modular transitable, en la que se colocan a modo de trama una serie de cables que funcionan como antenas, que captan la ubicación de los RFID tags, las cuales están colocadas en el calzado de los usuarios. A través de una serie de interfaces los datos se envían a una computadora en la cual se analizan y se puede determinar la posición exacta del usuario. Es interesante destacar que permite la captación de varios participantes de manera simultánea. Estos datos permiten el control de sonido y video que se proyecta sobre la plataforma.

Analizando en conjunto los dispositivos de captación descriptos hasta el momento, mecánicos, acústicos y magnéticos, podemos decir que en mayor o menor medida, se tratan en cierto aspecto de métodos invasivos para el usuario, y que en algún punto restringen su libertad de movimiento y acción.

2.4. Sistemas Ópticos

En los últimos años este tipo de sistemas se ha convertido en uno de los más populares por el fácil acceso a los dispositivos que utiliza, las cámaras. La democratización del uso de cámaras de video y webcams, y a su vez la mayor libertad de movimiento que ofrecen al usuario, el cual no debe estar conectado a ningún cable, hacen de los sistemas ópticos los de mayor difusión en los últimos años.

Estos sistemas están limitados por la resolución de las cámaras y la sofisticación del software de captura que utilicen.

2.4.1. Cámaras

Dentro de estos dispositivos podemos hacer dos diferentes tipos de clasificaciones. Una en relación a la fidelidad y calidad de captura, donde diferenciamos claramente dos grandes grupos, las cámaras web, o mas conocidas como webcams, y las cámaras DV. Las primeras son las más difundidas por su bajo costo y fácil acceso, en comparación con las DV, actualmente popularizadas como MiniDV, que tienen una mayor fidelidad y estabilidad en la calidad de la imagen, lo que favorece el proceso de captura, pero un costo mucho más elevado.

Por otro lado la segunda clasificación la podemos establecer de acuerdo al tipo de captura que realicen las cámaras, donde básicamente podemos distinguir entre cámaras de captura de espectro visible, y cámaras infrarrojas.

Las de espectro visible son las estándar, tradicionales, cámaras que tienen un tipo de “visión” similar a la humana en cuanto al rango del espectro cromático que perciben. A diferencia de esto las cámaras infrarrojas, dentro de las cuales también podemos considerar cualquier tipo de cámara de espectro visible a la cual se le incorpore un filtro infrarrojo, son dispositivos que captan solo una porción del espectro, más precisamente el espectro infrarrojo.

Independientemente del tipo de cámara que se utilice, todas en mayor o menor medida permiten realizar captura de movimiento, sin embargo tienen una singular relevancia los dispositivos infrarrojos, ya que permiten realizar el proceso de captura totalmente independiente del proceso de síntesis de imagen, es decir de las proyecciones. Esto lo podemos apreciar sobre todo en instalaciones o performances donde el objeto de la captura esta simultáneamente siendo proyectado y modificado, en especial en el caso de las pantallas sensibles o interfaces táctiles.

Hagamos referencia a dos proyectos que utilizan cámaras infrarrojas para la captura de movimiento en pantallas sensibles, por un lado FTIR Multi-Touch Sensing y The Khronos Project.

Un dato interesante es creciente desarrollo en las técnicas conocidas como Hardware Hacking, relacionadas con la generación casera de interfaces específicas (home-made interfaces). Estas técnicas se basan en la adaptación de los dispositivos con el fin de que sean capaces de realizar tareas para las cuales no fueron específicamente fabricados originalmente.

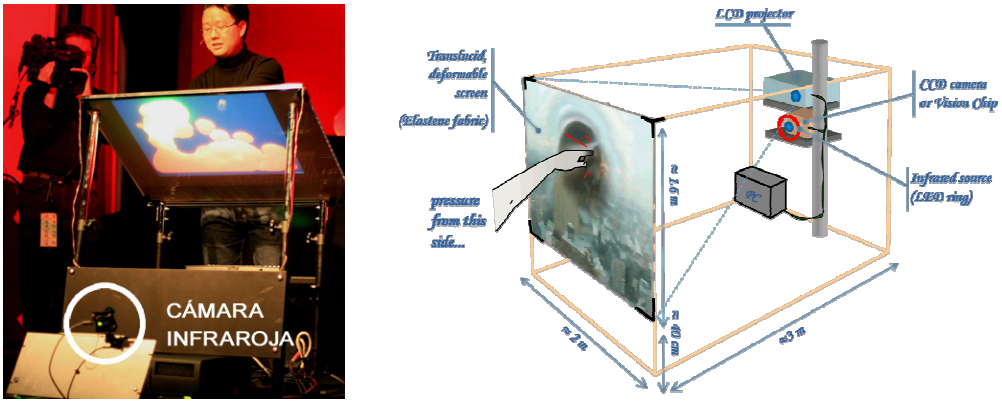

Figura 1: A la izquierda una vista del dispositivos utilizado en FTIR Multi-Touch Sensing interface (http://mrl.nyu.edu/~jhan/ftirtouch/)
A la derecha, esquema del funcionamiento de The Khronos Project (http://www.k2.t.u-tokyo.ac.jp/members/alvaro/Khronos/)

De estas técnicas podemos rescatar una muy sencilla que permite transformar una cámara webcam hogareña de espectro visible, en una cámara infrarroja. La documentación acerca del procedimiento se encuentra disponible en:

http://www.hoagieshouse.com/IR/

http://www.kailashnadh.name/docs/ir_cam/ir_cam.html

Podemos sintetizar que dicho procedimiento se basa en el intercambio de uno de los filtros de la cámara por un film a modo de filtro infrarrojo, el cual puede ser un acetato de radiografías o un negativo fotográfico velado.

Según la cantidad de cámaras utilizadas podemos determinar el tipo de captura. Por un lado hay sistemas más complejos para la captura del usuario en 3d, donde se emplean como mínimo tres cámaras y cuya principal finalidad es la animación y síntesis de personajes virtuales. En este caso, en estos sistemas además de las cámaras se utilizan pequeñas piezas esféricas direccionalmente reflectivas que se colocan sobre el cuerpo y se emplean como marcas. Cada una de las cámaras están sincronizadas y conectadas a la computadora para poder calcular la posición 3d de cada una de la marcas.

Por ultimo están los sistemas que emplean una sola cámara, que capturan el movimiento y gestualidad del usuario desde un solo punto de vista, donde el proceso de captura que realizan las diferentes aplicaciones se basa en el análisis y comparación de imágenes en 2d. Son estos sistemas en los que se basan prácticamente la totalidad de las aplicaciones y herramientas que son objeto de estudio de este trabajo y que abordaremos a continuación.

3. Software

Hemos mencionado que el software es la herramienta que nos permite analizar y comprender los datos obtenidos por alguno de los dispositivos descriptos. Es lo que en realidad concreta la captura como tal y nos habilita a poder utilizar esos datos para el control de imagen y sonido en tiempo real o el control de otros dispositivos como determinados actuadores de índole robótica.

Aquí diferenciamos dos grandes grupos de herramientas, aquellas que emplean programación visual basada en objetos y las que utilizan lenguajes de programación por código orientada a objetos.

3.1. Entornos de programación gráfica

Así podemos denominar a estas herramientas que utilizan la programación visual basada en objetos.

En estas aplicaciones el programa se construye usando objetos gráficos, lo que reduce la necesidad de aprender una sintaxis específica y brinda una manera clara e intuitiva de programar simplemente conectando los objetos entre si.

Dentro de este grupo de herramientas mencionaremos algunas como Eyesweb, Max/Msp o PD.

3.1.1. Eyesweb

http://www.eyesweb.org/

Esta es una plataforma de código abierto, desarrollado por el Laboratorio de Informática Musical de la Universidad de Génova. Se especializa en la captación de diferentes patrones de movimiento y gestualidad del cuerpo humano y la interpretación musical, y esta orientado a la producción de sistemas multimedia interactivos para el análisis de movimientos escénicos en tiempo real y controlar la síntesis de sonido y la ejecución en vivo de instrumentos.

El fuerte del programa es la captación y análisis de movimiento, si bien el Eyesweb tiene objetos propios destinados al control y procesamiento del sonido, estos tienen ciertas limitaciones, podríamos decir que no es la herramienta más adecuada para este fin.

En relación a esto citamos la instalación interactiva Espejo Espectral (http://www.proyecto-biopus.com.ar/instalaciones.html#Espejo) en la que el usuario puede controlar en tiempo real al generación de imágenes a partir de su propia silueta.

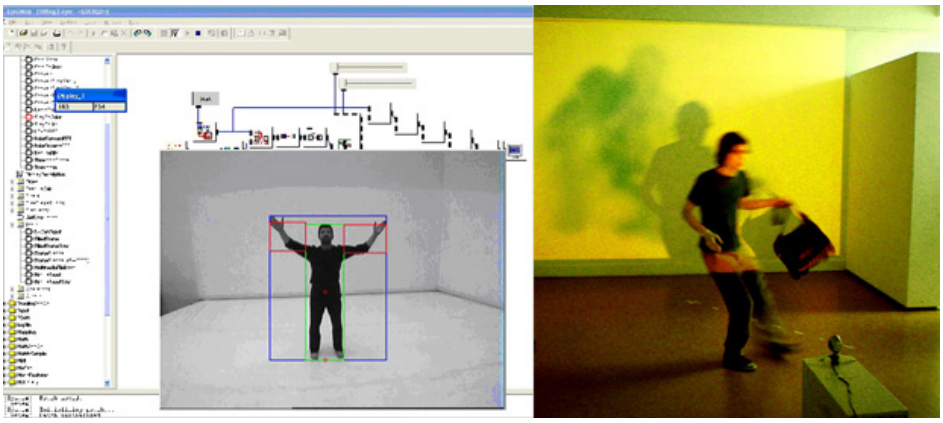

Figura 2: A la izquierda captura de interface de Eyesweb (imagen obtenida en http://www.infomus.dist.unige.it/eywindex.html). A la derecha, fotografía de la instalación Espejo Espectral del grupo Proyecto Biopus (http://www.proyecto-biopus.com.ar)

3.1.2. Max/MSP/Jitter

http://www.cycling74.com/

Max Fue concebido en 1986 como un proyecto para producir música interactiva en el IRCAM (Institut de Recherche et de Coördination Acoustique/Musique) de Paris.

Su autor original fue Miller S. Puckette y Desde entonces la aplicación se expandió hasta incluir procesamiento de audio, con la introducción de MSP, e imagen y video con Jitter.

Justamente Jitter es el módulo de max encargado de la edición de imagen y video en tiempo real, y que con la implementación de determinadas librerías como cv.jit posibilita la captura y análisis de movimiento. También se incorporan librerías GL para la generación de gráficos 3D.

Este módulo tiene más de 200 objetos específicos para la edición de imagen y video. La lógica con la cual fue concebido el programa permite controlar los datos de una manera muy versátil, pudiendo

crear aplicaciones para componer, improvisar y modificar contenidos en tiempo real. Lo que Max hace es transformar todo en un simple flujo de números donde todo se puede conectar con todo.

Max esta basado en el lenguaje C de programación, lo que permite que uno pueda escribir en C sus propios objetos. El lanzamiento de Max 4.5 extiende las capacidades de max, ya q soporta Java y javascript.

Ancestrales nocturnos (http://www.iuna.edu.ar/departamentos/multimedia/observatorio/eventos/

ancestr.htm) es una instalación interactiva sobre el calco de una estela (escultura) de la cultura Cotzumalhuapa, en donde los sonidos y las imágenes parecen emerger de la piedra misma creando una comunicación intima entre el espectador y la obra. Se implementó un sistema de captación de movimiento que, a partir de la posición del espectador, controla las transformaciones sonoras y visuales proyectadas sobre la misma piedra en tiempo real.

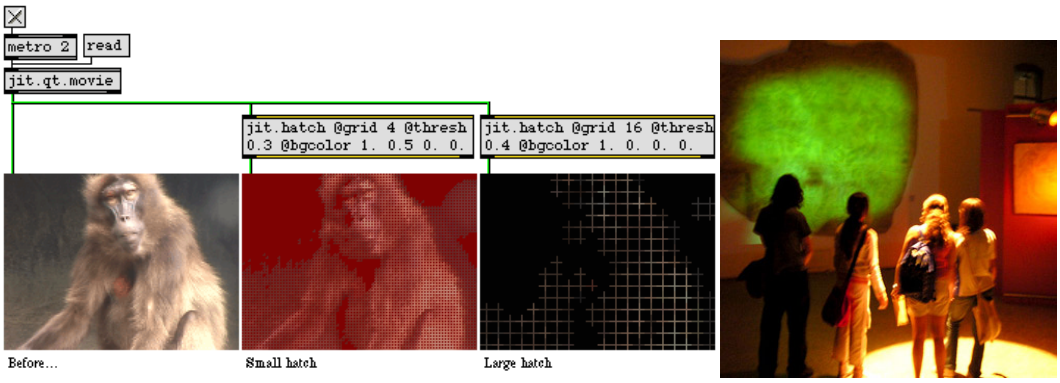

Figura 3: A la izquierda captura de interface de Jitter. A la derecha, fotografía de la instalación Ancestrales Nocturnos de Pablo Cetta y Matías Romero Costas (http://www.iuna.edu.ar/departamentos/multimedia/observatorio/eventos/ancestr.htm)

3.1.3. PD/GEM

http://pd.iem.at/

El autor de esta aplicación también fue Miller S. Puckette, y trabaja de una manera similar a Max/MSP. Podríamos definirlo como la versión free de Max, ya que es de distribución gratuita.

Diferenciándose en objetos específicos, la lógica y alcances de esta herramienta es similar Max.

En el caso del procesamiento de la imagen y la captación de movimiento, el modulo encargado de eso en PD es GEM, original de Mark Danks, que emplea objetos para la captación de movimiento e incorpora librerías como Open-GL para la generación de gráficos 3D.

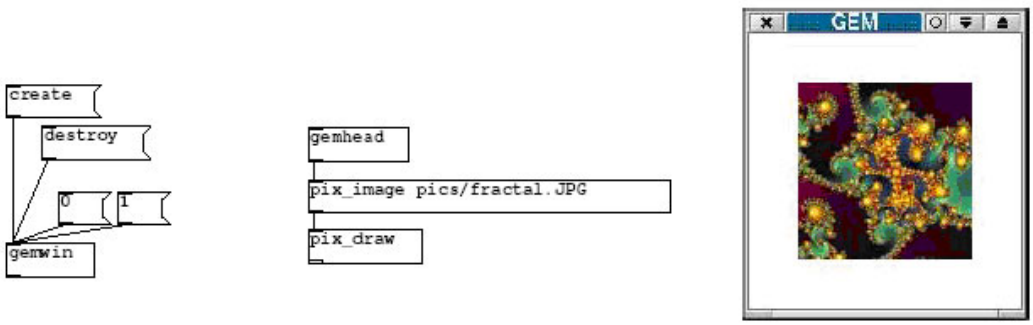

Figura 4: Captura de interface de GEM

3.1.4. Generalidades

Sintéticamente lo que realizan estas herramientas de software es utilizar una cámara, mini DV o webcam, como sensor de movimiento, esto es, captar la imagen, la cual también puede ser un video, procesarla y detectar el movimiento.

Las diversas herramientas permiten emplear filtros lineales y no lineales, realizar cambios de formato y operaciones, como adición, substracción, multiplicación, crop, extracción de zona, extracción de canal y chroma key.

A través de la combinación de diferentes filtros y operaciones es posible lograr la captación del movimiento a partir de una secuencia de imágenes dadas.

En la captación de movimiento se pueden distinguir dos modalidades, una basada en la comparación entre cuadros y otra en la identificación por color.

La primera modalidad utiliza básicamente la sustracción de dos cuadros a modo de comparación para detectar las variaciones en los píxeles y así determinar el movimiento de la imagen. Dentro de los métodos de sustracción podemos hablar de captación por sustracción de fondo, o por delay. En la primera, las imágenes son comparadas en relación a un único cuadro capturado previamente, por ejemplo para terminar la actividad de un ambiente en una instalación se registra con anterioridad una imagen del lugar vacío, es esta captura con la cual se comparan todas las imágenes mientras dure el proceso de captación. Este procedimiento sirve no sólo para la captación de movimiento sino también para la captación de presencia. El caso de la sustracción por delay es similar, sólo que la comparación se realiza entre dos instancias consecutivas, la imagen A se compara con la B, la B con la C…, y así sucesivamente. A diferencia del caso anterior, con la sustracción por delay no es posible la captación de presencia.

Por último la segunda modalidad utiliza el seguimiento por color (color tracking). En este caso se identifican dentro de la imagen determinados colores asociados a los objetos a capturar.

A su vez, las aplicaciones cuentan con ciertos módulos que pueden configurarse para determinar parámetros de movimiento más específicos, como el área, centro, posición, velocidad, captación de la silueta y la sombra, posturas e índices de contracción.

Es posible también mapear espacios, es decir la captación por grilla de regiones y captación de regiones en movimiento.

Con mayor o menor eficiencia, estas herramientas permiten el control de sonido también, sin embargo una característica importante es la de permitir enviar y recibir datos hacia y desde otras aplicaciones, locales o remotas. A demás de los protocolos como el MIDI y TCP/IP, soportan el protocolo OSC (open sound control) que puede conectar diferentes softwares (a través del puerto UDP) y tiene una estructura de flujo diseñada especialmente para la transmisión de datos de tipo vector o matriz, esenciales para enviar los datos de una silueta o de una región a través de la red.

Es a través de los mencionados protocolos de comunicación, que la información obtenida de la captación del gesto que realizan estas aplicaciones pueden ser transmitida a programas más especializados en la síntesis y procesamiento del sonido, como el Max por ejemplo. De esta manera se utilizan estos parámetros para sintetizar y modificar sonidos en tiempo real, esto es, manipular en sonido con la gestualidad corporal por ejemplo.

3.2. Programación secuencial / Librerías

A diferencia de las aplicaciones anteriormente descriptas, el conjunto de aplicaciones y librerías que abordaremos a continuación, están basadas en la programación secuencial, como hemos dicho programación por código orientada a objetos (paradigma Java, C++), aquí ya no hablamos de objetos que podemos interconectar gráficamente, sino de programación por código, con un lenguaje y sintaxis específico, la cual podríamos definir como una serie de instrucciones escritas que se ejecutan en un orden, en secuencia.

Esto permite un mayor control y manejo de los recursos, especialmente si hablamos del procesamiento de la imagen píxel por píxel.

Es este tipo de procedimientos lo que hace a estas herramientas tan potentes y lo que nos permite analizar una imagen, ya sea estática o de video, parte por parte, desde su mínima unidad, esto es píxel por píxel y desglosar su contenido en una serie de datos contenidos en una matriz para luego poder analizarlos y procesarlos con diversos algoritmos.

Justamente este tipo de procedimientos es lo que permite realizar los diversos tipos de captura y análisis de movimiento para posteriormente generar una respuesta expresiva. Aunque desarrollaremos puntualmente con mas detenimiento este tema, mencionaremos como ejemplo de estos procedimientos la captación de la silueta. Se analiza píxel por píxel una imagen y se la compara con valores de nuestras anteriores, pudiendo así determinar cual es su silueta, su contorno, y calcular su tamaño y posición entre otras cosas

Del mismo modo con la programación secuencial es posible simular sistemas complejos y caóticos, generar diversos patrones como los fractales y aplicarlos a la imagen de video en tiempo real a través del mencionado procesamiento píxel por píxel.

Haremos referencia a aplicaciones como Processing o librerías como opencv, reactivision y artoolkit entre otras.

3.2.1. Processing

http://www.processing.org/

Es un proyecto desarrollado por Ben Fry (Broad Institute) y Casey Reas (UCLA Design | Media Arts), un entorno y lenguaje de programación de código abierto orientado a la programación de imágenes, animaciones y sonido en tiempo real, y está diseñado especialmente para realizar aplicaciones multimedia en entornos web e instalaciones.

Pitch Fractal (http://www.proyecto-biopus.com.ar/instalaciones.html#pitch) es una instalación interactiva en donde las personas pueden generar trazos a partir del canto, que a su vez generan figuras fractales. La obra combina algoritmos desarrollados en max/msp y processing.

Es interesante mencionar como este ejemplo nos muestra la versatilidad de estas herramientas en relación a la captación del gesto, ya que en este caso no hablamos de gestualidad visual sino sonora.

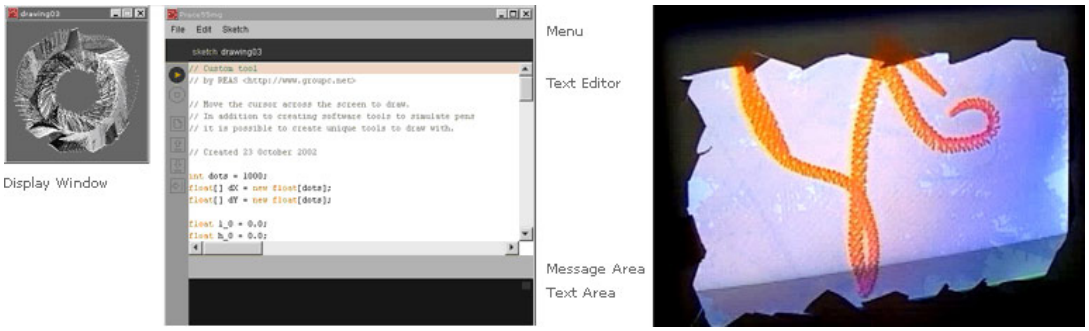

Figura 5: A la izquierda captura de interface de Processing. A la derecha, fotografía de la instalación Pitch Fractal del grupo Proyecto Biopus (http://www.proyecto-biopus.com.ar)

3.2.2. Librerías de Visión Artificial

Estas librerías de visión artificial (CV - computer vision) trabajan con métodos que permiten a la computadora “comprender” las imágenes. Aquí el termino “comprender” se refiere a que información específica es extraída de la imagen para un propósito determinado, controlar un proceso. La imagen digital que se transfiere al sistema de computación artificial generalmente es en escala de grises o

RGB, pero puede ser también 2 o más imágenes simultaneas o, como en el mayor de los casos que estudiaremos, una secuencia de video en tiempo real.

Teniendo en cuenta los métodos utilizados y las tareas específicas que realizan las diferentes librerías y aplicaciones que utilizan visión artificial, distinguimos entre aquellas diseñadas con diferentes funciones que permiten realizar una serie de tareas y las que realizan una tarea específica.

3.2.2.1. Open Source Computer Vision Library

http://www.intel.com/technology/computing/opencv/index.htm

Original de la empresa Intel, se encuentra dentro del primer grupo, y permite realizar diversas tareas tales como la identificación de objetos, segmentación y reconocimiento de una imagen, reconocimiento facial, seguimiento de trayectorias y movimiento, captación del gesto.

3.2.2.2. cv.jit

http://www.iamas.ac.jp/~jovan02/cv/index.html

En este caso hablamos de una serie de herramientas para el desarrollo de aplicaciones de visión artificial, diseñadas específicamente para trabajar con Max/MSP/Jitter. El objetivo de este proyecto es proveer al usuario del programa de una serie de objetos para realizar tareas de reconocimiento y segmentación de la imagen, así como la captación de movimiento. Por otro lado incorpora herramientas educativas que delinean los conceptos básicos de las técnicas de visión artificial.

3.2.2.3. Myron WebCamXtra

http://webcamxtra.sourceforge.net/

Myron es un plugin de código abierto (open source), una serie de librerías compiladas para diferentes plataformas y lenguajes. La versión para Java y Processing se conoce como JMyron y la de Macromedia Director, como WebCamXtra. El objetivo de este proyecto es mantener libre de costos las herramientas y técnicas de la visión artificial para su fácil utilización dentro de la comunidad artística y en la educación. Las tareas que realiza esta aplicación se concentran en la captación de movimiento, el reconocimiento y segmentación de la imagen.

Una de las obras que implementa JMyron en processing es Tango Virus (http://www.proyecto-biopus.com.ar/instalaciones.html#tango), una instalación interactiva que permite al público modificar en tiempo-real un tema de tango.

El público puede bailar el tema de tango que se está escuchando, pero dicho baile se transforma en un comportamiento viral que ataca al tema musical, haciendo que este varíe, quizás al punto de “fallecer”.

Dentro de los diferentes módulos que tiene el sistema para poner en funcionamiento la instalación, el primero de ellos es el encargado en registrar la posición de los usuarios/bailarines y captar el gesto del baile.

Figura 6: A la izquierda esquema del dispositivo de captura de la instalación Tango Virus. A la derecha, fotografía de la instalación Tango Virus del grupo Proyecto Biopus (http://www.proyecto-biopus.com.ar)

3.2.2.4. Librerías de detección facial

La tarea concreta de estas librerías es la detección de rostros humanos dentro del contexto de una determinada imagen, en movimiento a través de una señal digital de video o estática en una imagen digital.

La detección facial, o reconocimiento del rostro humano, es esencial para una interacción inteligente entre el hombre y la computadora.

Las librerías y aplicaciones de reconocimiento facial emplean diversos métodos para identificar y localizar un rostro dentro de una imagen o una secuencia de imágenes y para construir sistemas automatizados que analicen la información de estas imágenes es necesario emplear algoritmos de detección facial altamente efectivos.

El objetivo de estas herramientas es identificar todas aquellas regiones de la imagen dada, ya sea una imagen o una secuencia de video, que contengan un rostro a pesar de sus diferentes posiciones, orientación o condiciones de luz.

3.2.2.4.1. Métodos por conocimiento

Este método incluye una serie de reglas que codifican información de lo que constituye un rostro para el ser humano, lo que el hombre entiende por cara.

Generalmente, las reglas almacenan la relación entre los rasgos faciales. Estos métodos son diseñados principalmente para localización facial, previo proceso de verificación para reducir falsas detecciones.

3.2.2.4.2. Enfoque de rasgos constantes

La función de estos algoritmos es encontrar rasgos estructurales que están presentes aun cuando la postura, punto de vista o condiciones de luz varían.

A diferencia del método anterior, la investigaciones se centran en la identificación de estos rasgos constantes para el reconocimiento facial. Rasgos como cejas, ojos, nariz, boca y cabello son comúnmente extraídos utilizando detectores de límites. Sobre los rasgos extraídos, un modelo estadístico es construido para verificar la existencia de un rostro. Estos métodos son diseñados principalmente par la localización facial.

3.2.2.4.3. Comparación por plantilla

Este método utiliza un patrón de rostro estándar, generalmente frontal, que es parametrizado por una función. Dada una imagen, los valores correlativos del patrón son computados con los del contorno del rostro, ojos nariz y boca independientemente. La existencia de un rostro es determinada por la correlación de esos valores. Este método tiene la ventaja de ser fácil de implementar, pero presenta las dificultades de no ser efectivo si se presentan variaciones de escala, pose o forma. El método se circunscribe a la detección frontal del rostro.

3.2.2.4.4. Métodos por apariencia

A diferencia del método por plantilla, los modelos (o templates) son “aprendidos” a partir de una serie de imágenes de entrenamiento la cuales deben capturar la variaciones representativa de la apariencia facial. Un problema de este método es la dificultad de trasladar el reconocimiento humano a reglas bien definidas. Estos métodos son diseñados principalmente para detección facial, previo proceso de verificación para reducir falsas detecciones.

3.2.2.5. Reconocimiento de patrones

El método empleado para el reconocimiento y seguimiento (tracking) de marcas en imágenes binarias combina la comparación de patrones sobre gráficos de topologías binarias para reconocimiento e identificación, con simples técnicas geométricas para calcular la ubicación y orientación de las marcas.

Los gráficos pueden ser comprendidos como representación de un árbol de nodos, en el cual se representan cada una de las zonas por su de jerarquía, según su orden de inclusión y su posición dentro de la marca. Las regiones negras están contenidas por regiones blancas o viceversa.

El primer nodo representa a la forma que contiene a todas las demás o aquella que no es contenida por ninguna otra por decirlo de otra manera. Mientras que los nodos del último nivel, las “hojas” del árbol, representan la cantidad de formas que no contengan en su interior a ninguna otra forma más que ellas mismas.

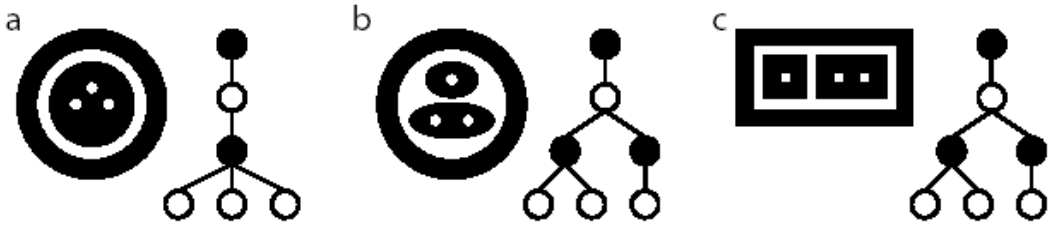

Figura 7: Tres gráficos binarios (a, b, c) utilizados como marcas. A la derecha de cada uno de ellos se presenta el mapa de los nodos a modo de árbol.

El método para el calculo de la ubicación y orientación de las marcas esta vinculado al desarrollo del algoritmo de segmentación, el cual solo guarda los ejes alineados de los cuadrantes (bounding boxes) de cada una de las regiones de la marca, lo cual es efectivo si la región es cuadrada, circular y / o relativamente pequeña.

Las “hojas” son siempre las regiones más pequeñas del árbol, y sus centros son la información espacial más exacta que se tiene acerca de la marca.

Por esto, el algoritmo calcula la ubicación y orientación de la marca como una combinación del cálculo de los centros de los cuadrantes de las “hojas” o nodos del último nivel como se explica en la figura 8.

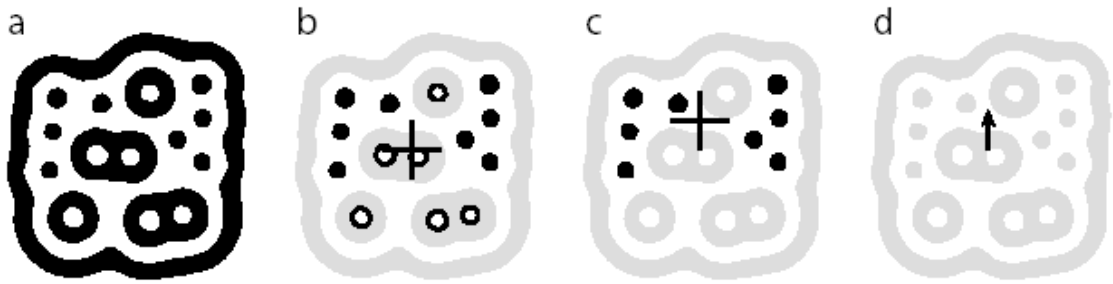

Figura 8: Proceso de ubicación y orientación. Se computa el centro de la forma (a), detectando y promediando los centros de todas las “hojas” (b). El vector trazado desde este centro hacia el punto dado por el promedio de los centros de los nodos negros (c) es usado para calcular la orientación de la forma (d).

.

Cada nodo es identificado para los diversos cálculos en función de su profundidad, se calcula el área que ocupan dentro de sus regiones contenedoras. Esto permite aplicarse a cualquier tipo de diseño, aun en aquellas formas que tienen una sola región blanca y una sola región negra, lo que permite variar las estructuras de las marcas sin cambiar el método con el cual se las reconoce.

3.2.2.5.1. ArtoolKit

http://www.hitl.washington.edu/artoolkit/

ARToolKit es una librería para la construcción de aplicaciones (AR), aplicaciones que implican la superposición de imágenes virtuales con las del mundo real.

Una de las dificultades en el desarrollo de este tipo de aplicaciones es el seguimiento del punto de vista del usuario. Para saber desde que perspectiva se construirá la imagen virtual, la aplicación necesita saber como el usuario esta mirando en el mundo real.

ARToolKit utiliza algoritmos de visión artificial (CV) para resolver este problema, sus librerías de seguimiento de video, calculan en tiempo real la posición y orientación de la cámara en relación a las marcas ubicadas en diferentes objetos. Esto habilita el fácil desarrollo de un amplio rango de aplicaciones AR

Alguno de los proyectos que podemos mencionar es The Mixed Reality Lab (http://www.mixedrealitylab.org/), el cual no solo se especializa en el desarrollo de imágenes virtuales en 3d como vemos en al imagen x, sino que incorpora un novedoso método para la generación de imágenes virtuales 3d a partir de señal de video (ver figura 9).

La incorporación de contenido 3D en vivo dentro de aplicaciones AR se basa en un algoritmo de reconocimiento de la forma a partir de la silueta (shape-from-silhouette algorithm). El objeto o sujeto a capturar es monitoreado por 15 cámaras, el sistema genera una vista original, que puede ser generada a una velocidad interactiva, y que es incluida dentro del espacio de AR. El objetivo principal de este proyecto es producir un modelo humano tridimensional real que pueda ser utilizado en entornos virtuales.

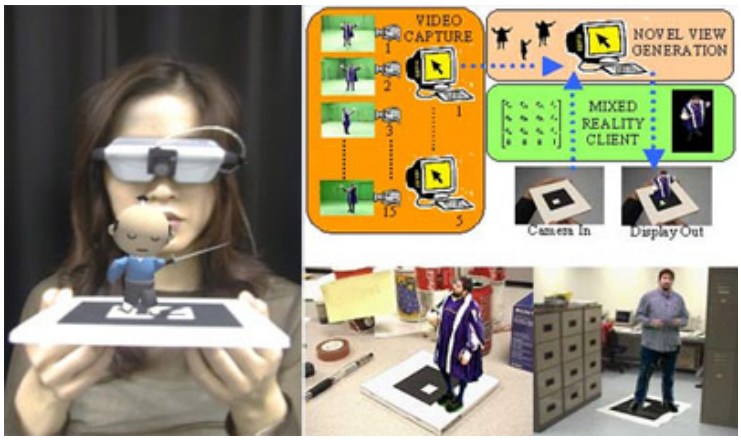

**Figura 9: A la izquierda imagen obtenida de http://www.hitl.washington.edu/artoolkit/
A la derecha esquema e imágenes del dispositivo utilizado en 3D Live desarrollado por The Mixed Reality Lab (http://www.mixedrealitylab.org/research/LIVE/LIVE_webpage/research-LIVE-infor.htm)**

3.2.2.5.2. reacTIVision

Es un software de código abierto, un entorno para el reconocimiento rápido y efectivo de patrones en tiempo real.

Se diseñó principalmente como una herramienta para la creación de mesas tangibles a modo de interfaces. Basada en otras librerías, esta aplicación fue desarrollada por Martin Kaltenbrunner, en el Music Technology Group, en Barcelona, como parte del proyecto de reacTable, un instrumento musical que emplea las mesas tangibles anteriormente mencionadas.

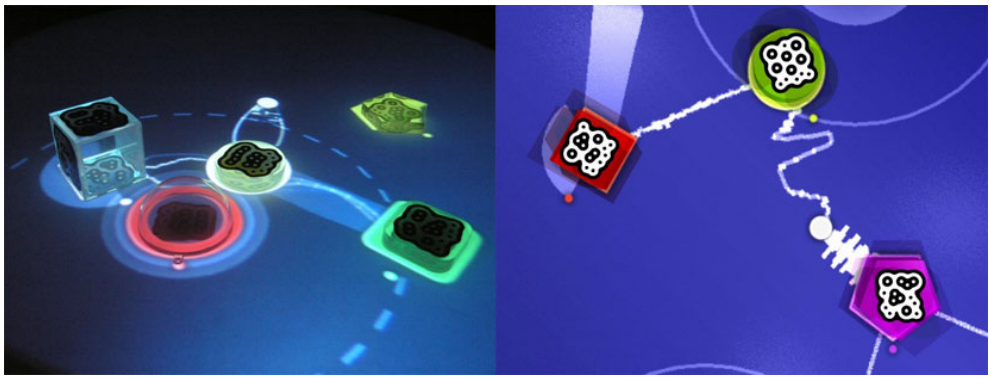

Figura 10: Imágenes de la instalación reacTables desarrollada con reacTIVision (http://www.iua.upf.es/mtg/reactable)

Estas mesas tangibles están constituidas por una superficie translucida, sobre la cual se colocan diversos objetos que tienen la impronta de las diferentes marcases a reconocer. Una cámara registra la posición de las marcas y envía los datos a reacTIVision.

La aplicación envía mensajes OSC (OpenSound Control) a través del puerto UDP. Se implemento el protocolo TUIO, especialmente diseñado para transmitir el estado de cada uno de los objetos.

Finalmente la imagen procesada según los parámetros de las marcas obtenidos por la aplicación TUIO, es retro-proyectada sobre la superficie translucida.

De esta manera todo el sistema queda debajo de la mesa, oculto a la vista del usuario.

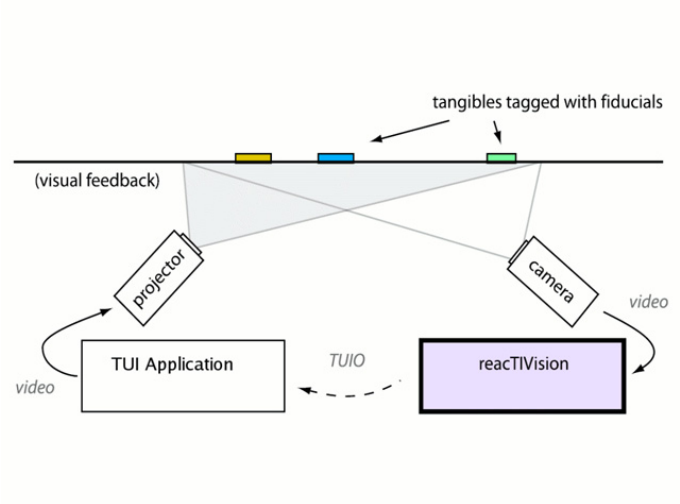

Figura 10: Gráfico que describe el dispositivo empleado en reacTables (http://www.iua.upf.es/mtg/reactable)

3.2.2.6. Librerias de detección de manos - HandVu

http://www.movesinstitute.org/~kolsch/HandVu/HandVu.html

Este software implementa una interface basada en el reconocimiento de las manos. HandVu detecta la mano en una postura estándar y luego realiza un seguimiento y reconoce posturas determinadas, todo en tiempo real.

La herramienta esta compuesta por una librería principal y una serie de aplicaciones de captura relacionadas con OpenCV's highgui, DirectShow, y ARtoolkit.

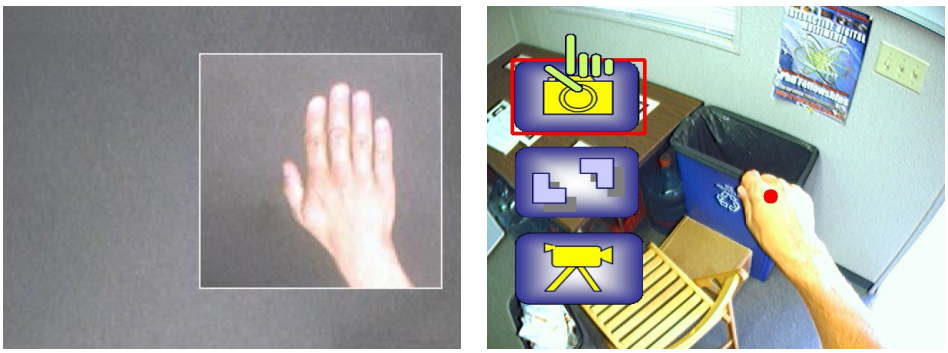

Figura 12: Capturas de diferentes aplicaciones realizadas con HandVu. Imágenes obtenidas en http://www.movesinstitute.org/~kolsch/HandVu/HandVu.html

Por último para finalizar este apartado, es interesante mencionar que la mayoría de estas librerías son de “Open Source” (código abierto), y están registradas bajo la licencia GPL (General Public License).

4. Artistas

Al margen de las obras puntualmente mencionadas a lo largo de los informes anteriores, es oportuno profundizar en este apartado haciendo referencia a algunos artistas que han empleado alguna de las herramientas mencionadas en sus obras. Obras que situamos dentro de un proceso de evolución artística hacia una nueva interacción hombre - computadora.

4.1. Messa di Voce

http://www.tmema.org/messa/messa.html

Golan Levin es una artista, diseñador y compositor interesado en la creación de dispositivos y experiencias que exploren nuevos modos de expresión audiovisual. Su trabajo se enfoca en el diseño de sistemas para la creación de performances tanto de sonido como imagen. Como resultado de esto surge entre otros proyectos, Messa di Voce, una serie performances audiovisuales en las cuales la voz, los gritos y canciones producidas por vocalistas son "materializadas" en tiempo real por aplicaciones de visualización interactiva especialmente desarrolladas.

Independientemente de las variaciones específicas que se pueden observar en cada una de las obras que conforman la serie realizada por Messa di Voce, las herramientas de software desarrolladas realizan básicamente tres tareas: la captación del movimiento y ubicación de los vocalistas, la captación del gesto vocal o sonoro y por último la síntesis de la imagen.

Dentro de la serie, tomemos como ejemplo "Fluid", en donde una especie de flujo de plasma parece emerger de las bocas de los performers cuando estos comienzan a cantar o hablar. En este caso el sistema se divide en tres módulos según las tareas ya mencionadas. El primer módulo determina la ubicación de los vocalistas y la posición aproximada de sus bocas combinando un dispositivo óptico provisto de cámaras infrarrojas con una aplicación de visión artificial (CV). El segundo módulo capta la gestualidad de la voz, los performers están provistos de micrófonos inalámbricos estereofónicos que captan los sonidos que generan.

A través de estos receptores la información ingresa al sistema y es analizada para determinar el carácter del sonido. Aquí no sólo hablamos de poder determinar ciertos parámetros como el volumen y la altura, sino que el sistema analiza las características espectrales de la voz para reconocer las vocales. Esto determinará ciertas cualidades de la imagen, como por ejemplo el color, las vocales brillantes como la "e" se traducen en tonalidades verdosas o amarillentas. El tercer módulo toma los datos referentes al carácter de la voz, que junto con los datos de ubicación obtenidos en primer lugar, se utilizan para la generación de la imagen, es decir la síntesis de esta especie de plasma a partir de ciertos algoritmos genéticos para la simulación de fluidos. Por último la imagen resultante, esta "materialización" del sonido, es incluida dentro de la escena, siendo proyectada sobre las pantallas.

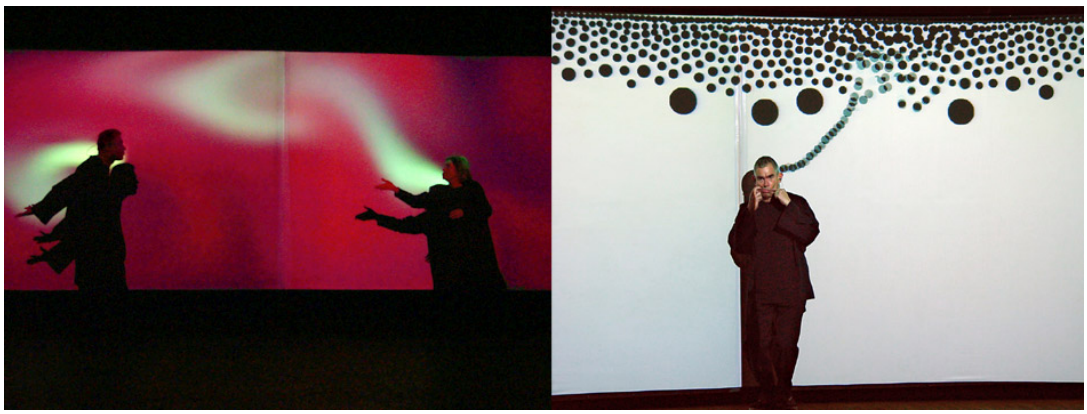

Figura 13: Imágenes de Messa di Voce. A la izquierda escena de la performance Fluid. A la derecha imagen de Bounce (Jaap's Solo). Imágenes obtenidas en http://www.tmema.org/messa/messa.html

4.2. Mine-Control

http://www.mine-control.com/

Mine-Control es un emprendimiento artístico colectivo encabezado por Zachary Booth Simpson que se especializa en la realización de instalaciones de obras de arte interactivas, con temáticas naturales y científicas desarrolladas para diversos museos alrededor del mundo.

Empleando conocimientos adquiridos en la industria del video juego, Mine-Control desarrolla sus propias herramientas para introducir a los participantes en novedosas técnicas de interacción que comprometen íntegramente cuerpo del usuario.

La mayoría de las obras utilizan como recurso principal la captación de la silueta para el control del sonido y la imagen, lo que se denomina como Sistemas de Detección de Sombras y Luces (Shadow and Flashlight Detection Systems).

Dentro de estos sistemas, las instalaciones pueden ser single-sided o double-sided. El primer término se refiere a aquellos sistemas en donde la captación y proyección se realiza desde un mismo punto de vista, en general de manera frontal. Es el más tradicional de los métodos, que tiene como desventaja los típicos problemas de oclusión que se generan cuando el usuario se acerca demasiado a la superficie de proyección.

Como ejemplo de este tipo de dispositivos, es oportuno hacer referencia a "Mariposa" o “Sand”, dos instalaciones que ofrecen al usuario la posibilidad de interactuar con diversos elementos virtuales a través de la propia sombra que proyecta el cuerpo sobre la pantalla.

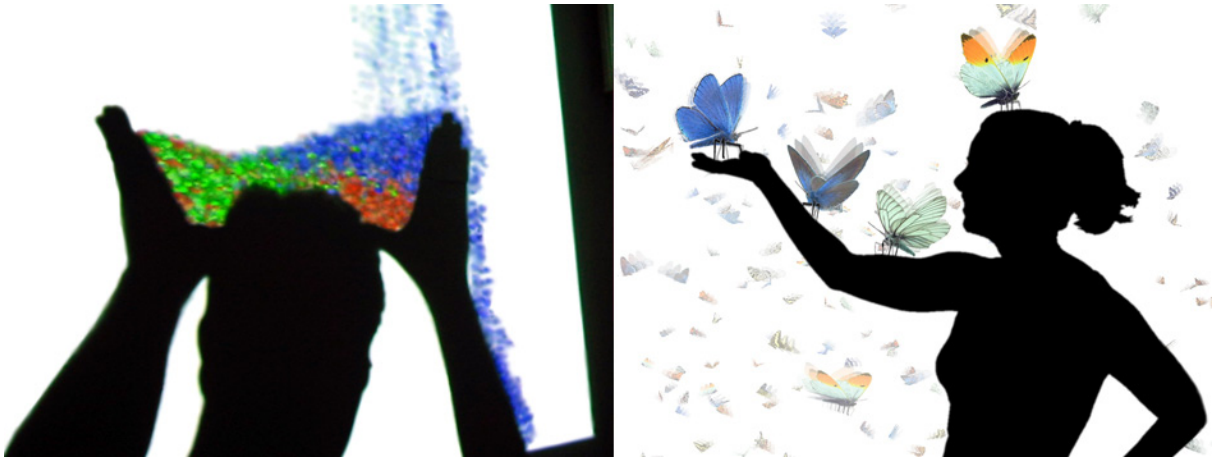

Figura 14: Imágenes de “Sand” y “Mariposa”, obtenidas en http://www.mine-control.com/downloads.html

El segundo término hace referencia a los métodos que utilizan pantallas translucidas, del lado frontal se proyecta la sombra del usuario y del otro lado de la pantalla se coloca la cámara para realizar la captación.

La utilización de dispositivos con pantallas translúcidas es uno de los principales aportes que se puede mencionar dentro de los sistema desarrollados por Mine-Control, ya que permite a los participantes permanecer frente a la pantalla e incluso tocarla, sin generar los problemas de oclusión ya mencionados.

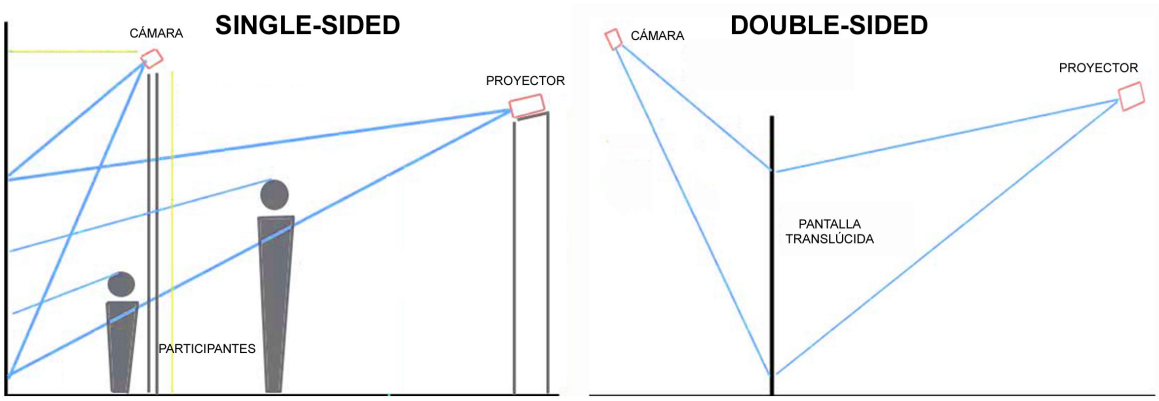

Figura 15: A la izquierda esquema del sistema Single-Sided. A la derecha esquema del sistema Double-Side. Imágenes obtenidas en http://www.mine-control.com/installation_details.html

El perfeccionamiento en el desarrollo de estas pantallas translucidas, y su combinación con luces infrarrojas, determina la aparición de variantes, como los sistemas de oclusión infrarrojo, similares a los mencionados en el apartado 2.5.1 de este trabajo.

En este caso se utiliza un sistema de retro-proyección y captación, el proyector se coloca por detrás de la superficie translúcida, al igual que la cámara, en este caso una cámara infrarroja, que capta la sombra que produce el usuario cuando intercepta la luz que emiten varias lámparas infrarrojas orientadas oblicuamente por delante de la pantalla.

Como ejemplo de este sistema se puede mencionar a Mondriaan, una instalación que utiliza este dispositivo, una obra en la cual los participantes pueden bosquejar y editar composiciones de género abstracto al estilo de Piet Mondrian.

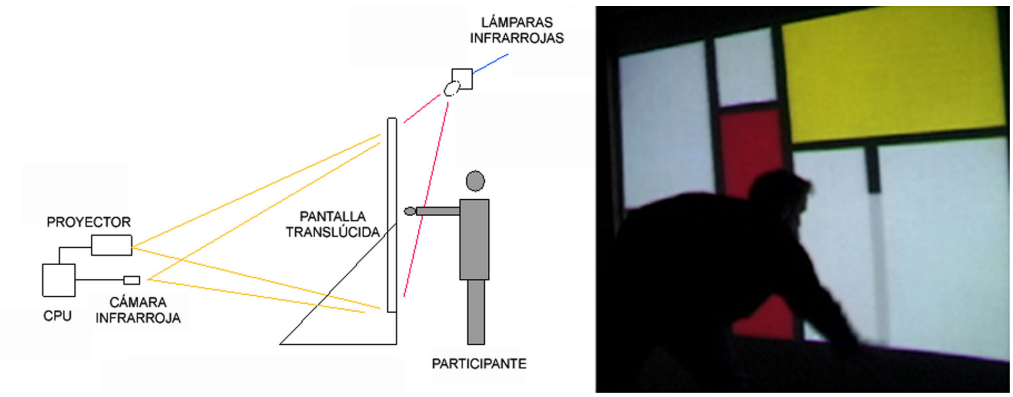

Figura 16 A la izquierda esquema del sistema de oclusión infrarroja. A la derecha imagen la instalación Mondriaan). Imágenes obtenidas en http://www.mine-control.com/installation_details.html y http://www.mine-control.com/mondrian.html respectivamente

5. Conclusión

Luego del recorrido realizado por las diferentes herramientas de software que realizan captura y análisis del gesto humano es posible concluir en primer lugar que el avance tecnológico de los últimos años ha favorecido el desarrollo de estas herramientas y contribuido a una redefinición de las interfaces de control de imagen y sonido en tiempo real, situando al cuerpo y a la gestualidad del usuario, participante o performer, de acuerdo al caso, en el centro de la interacción.

Es el propio cuerpo el que se relaciona directamente con la obra, creando así una comunicación directa, libre e intuitiva entre el hombre y los procesos de control de imagen y sonido.

En segundo lugar la implementación de estas herramientas en el proceso de creación artística es un aporte positivo, ya que pone a disposición del artista un recurso más de expresión y comunicación, donde la tecnología contribuye a generar sentido y en cierta medida forma parte del discurso.