

# Expandiendo el escritorio

T. Baermann<sup>1</sup>, M. Agustí<sup>2</sup>, J. V. Benlloch<sup>2</sup>

<sup>1</sup>Technische Universität Berlin  
Straße des 17. Juni 135  
10623 Berlin, Germany  
t.baermann@epost.de

<sup>2</sup>Departamento de Informática de Sistemas y Computadores  
Universidad Politécnica de Valencia  
Camino de Vera s/n  
46022 Valencia  
{magusti,jbenlloc}@disca.upv.es

**Resumen.** Inspirado en la conocida metáfora del escritorio, este trabajo tiene como objetivo el diseño e implementación de una interfaz de usuario óptica amigable, en la que colocando objetos reales sobre una pantalla de computador situada horizontalmente, el usuario pueda interactuar con aquél. Se introducen tanto las ideas claves de diseño como la solución final adoptada, basada en técnicas de visión por computador. Las aplicaciones de este tipo de interfaz pueden ser muy diversas, desde las puramente artísticas a otras mucho más académicas, como las que simulan una mesa en una sala de conferencias.

**Palabras clave:** Interfaz Multimedia, Interfaz Adaptable, Realidad Aumentada, Visión por Computador.

## 1 Introducción

La metáfora del escritorio ("*desktop metaphor*") existe desde hace unos 20 años cuando se inventó en el Rank Xerox Palo Alto Research Center y se introdujo en 1981 con el computador "Star" [1][2]. La idea se hizo bastante popular con los computadores Lisa de Apple [3] y finalmente, su uso en los sistemas Windows de Microsoft, la extendió a una gran mayoría de computadores en todo el mundo. Si bien la implementación de esta metáfora ha sufrido una gran metamorfosis desde su aplicación en los computadores Star hasta su uso en las modernas interfaces, las ideas básicas de manipulación directa de objetos, principalmente mediante el teclado o el ratón, apenas han cambiado. Sin embargo, la omnipresencia del computador en la sociedad digital nos hace pensar en planteamientos algo distintos.

Para dotar a la metáfora de una nueva dimensión que, al mismo tiempo, recordara su concepción original, la pantalla se dispondría horizontalmente, de forma que se pudieran depositar sobre ella objetos reales. De ese modo, los objetos tendrían influencia sobre el escritorio virtual del computador. Las aplicaciones de este tipo de interfaz pueden ser muy diversas, desde las puramente artísticas a otras mucho más académicas, como las que simulan una mesa interactiva en una sala de conferencias, donde diferentes recursos (documentos multimedia, secuencias de vídeo, aplicaciones

en ejecución, etc.) podrían intercambiarse entre los participantes. En este contexto, un gráfico que queda tapado por una hoja de papel, podría ser desplazado automáticamente por la interfaz, hasta visualizarse de nuevo por completo.

Imagine una mesa, por ejemplo, en un club, en una sala de espera o simplemente en una cocina. Entonces, cuando usted deja su copa sobre un mantel a cuadros, rojo y blanco, el patrón empieza a moverse. Al principio, los campos próximos a la copa comienzan a retorcerse para luego, uno tras otro, ir perdiendo su forma original hasta formar un efecto lava alrededor de la copa. Se escucha el ritmo de un tambor que aumenta su frecuencia conforme se mueve la copa hacia el centro de la mesa. Cuando se coloca un plato al lado, se producen dos flujos de lava que se mezclan junto a un segundo sonido más grave. Cada nuevo objeto colocado en el escritorio, genera un nuevo sonido mientras el mantel muestra unos colores psicodélicos adecuados.

En este trabajo se presenta un desarrollo experimental de esta concepción de escritorio.

## 2 Interacción

La idea central es que se puedan manipular directamente objetos situados sobre el escritorio, apuntando con el dedo, pasando la mano por encima o marcando determinadas áreas de la pantalla.

En esta línea, diferentes trabajos ([4] [5] [6]) muestran ejemplos de lo que sería un sistema completo de inmersión e interacción hombre-máquina que constituyen la base teórica de otras aproximaciones similares. Sin embargo, nuestro planteamiento es más simple, ya que no se plantea utilizar tabletas virtuales, proyecciones transparentes de barras de iconos y menús, ni otros periféricos usuales en estos sistemas. En su lugar, las manos y unos pocos objetos reales serán los instrumentos a través de los cuales se conducirá la interacción. Pero las ideas fundamentales del paradigma, la exploración y el reconocimiento de gestos y posiciones sí serán empleados.

### 2.1 Elementos de la Interfaz de Usuario

La interfaz de usuario con el escritorio se realiza a través del uso de objetos y zonas. Respecto a los primeros, su posición, el movimiento de los mismos, el área que determinan entre ellos o su número, pueden ser considerados parámetros que determinan las acciones del sistema. En lo que respecta a las zonas, determinadas áreas prefijadas van a actuar como si de botones se trataran; es decir, cuando un objeto permanezca un cierto tiempo sobre una de estas zonas "activas" el sistema lo interpretará como la pulsación del botón izquierdo del ratón sobre un objeto, ejecutando la acción correspondiente asociada.

Otros elementos podrían formar parte de este sistema como desencadenantes que dan lugar a respuestas del sistema. Así, las manos del usuario podrían ser útiles medios de entrada, no sólo por su posición, sino también por los gestos que realicen, en la línea de los trabajos del grupo *Autonomous Agents* del Media Lab [7].

## 2.2 Modos de Respuesta

En el momento actual el sistema incorpora tres modos de respuesta diferentes. El primero permite mostrar diferentes contenidos a pantalla completa, de tal forma que la transición viene determinada por el movimiento de un objeto en la escena (se podría utilizar la propia mano para pasar "virtualmente" las diapositivas).

El segundo permite configurar el color de fondo del escritorio en función de los objetos presentes en la escena. A cada uno se le asignará un color que, en función del tamaño del objeto, se extenderá en un determinado radio alrededor del centro de aquél. Si dos o más objetos están próximos el color resultante será la suma de las "contribuciones" de los objetos cercanos.

La tercera de las respuestas del sistema se asemeja a una pantalla nevada o una caja de arena. De la parte superior de la misma, van deslizándose verticalmente puntos de color que en su camino pueden encontrarse con un obstáculo (un objeto que aparece en el escritorio), en ese caso los puntos se desviarán de su trayectoria y bordearán el objeto en cuestión hasta encontrar de nuevo un camino vertical libre para continuar su descenso. La aparición de un nuevo objeto, o la desaparición de uno que estuviese ya presente, modificará consecuentemente el comportamiento del sistema.

## 3 Metodología de Diseño y Desarrollo

El reconocimiento de los objetos situados en el escritorio podría llevarse a cabo utilizando diferentes aproximaciones, como por ejemplo: ultrasonidos, células de carga, ópticas específicas, matrices eléctricas o visión por computador. En un contexto de bajo coste, la última solución parece la más realista, ya que no implica un hardware complejo. En concreto, se ha utilizado un computador portátil con un procesador Celeron a 333 MHz y 64 MB de RAM, cuya pantalla constituye el escritorio y sobre la misma, se ha situado una cámara de bajo coste (Video Blaster WebCam 3, de Creative Labs), tal y como se muestra en la Fig.1.

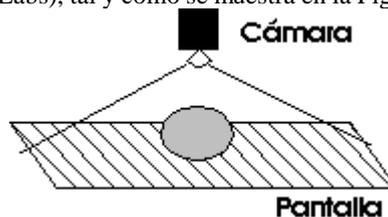


Fig. 1. Esquema general del sistema.

Sin embargo, mientras muchas aplicaciones de visión por computador trabajan con un fondo uniforme, en este caso, supone un reto detectar una serie de objetos sobre un fondo cambiante que, en realidad, transporta la información gráfica de nuestra pantalla. Por tanto, restando la imagen capturada por una cámara enfocada sobre el escritorio, respecto de la imagen enviada a la pantalla se podrían determinar, al menos teóricamente, los objetos situados sobre el escritorio.

En la práctica, esta aproximación tan simple tiene una serie de inconvenientes, ya que el problema de la correspondencia entre la imagen capturada por la cámara y la que realmente se está pintando en la pantalla (que se puede obtener interrogando al propio sistema), no es en absoluto trivial. La distorsión impuesta por la óptica de la cámara y las más que probables diferencias en resolución entre los dos mapas de bits, además de la influencia de la luz ambiente en la imagen de la cámara, obligan a un planteamiento mucho más elaborado. Lo que sucede es que varios puntos de la imagen real se "reagrupan" en la imagen capturada. Esta proyección no es conocida y puede variar durante la ejecución del sistema.

### **3.1 Segmentación**

El principal problema en nuestra aplicación es el de reconocimiento de los objetos que entran en la escena captada por la cámara, esto es, decidir qué puntos de la imagen pertenecen al objeto y cuáles al fondo (lo que se está pintando en la pantalla del ordenador).

La primera idea fue implementar una transformación geométrica que resolviera el problema de la correspondencia de imágenes, pero suponía un coste inaceptable para una aproximación en tiempo real. Lo siguiente fue utilizar redes neuronales para aprender la relación entre las dos imágenes. Esta aproximación supondría un alto coste en la etapa de aprendizaje, pero sería mucho más rápido en la ejecución final. Con las imágenes capturadas por ese tipo de cámara de bajas prestaciones, esta solución no resulta utilizable ya que la red, independientemente de su estructura, converge con un error del 10%, el cual resulta ser inviable para la discriminación de los colores. Es posible que utilizando cámaras de más calidad, este enfoque fuera adecuado.

Por último, dado que se trabaja con un monitor LCD que polariza la luz emitida, es posible utilizar otra aproximación más simple para resolver este problema. Consiste en montar un filtro polarizado ortogonalmente a la polarización de la pantalla LCD, de tal modo que la imagen del monitor queda totalmente bloqueada.

### **3.2 Post-proceso**

Una vez completada la etapa de segmentación, es necesario un proceso adicional que acabe de definir adecuadamente los objetos, para su posterior caracterización. En concreto, un operador morfológico de apertura permite eliminar puntos de ruido aleatorio que aparecen como resultado de la etapa anterior. Sobre esta imagen final ya es posible realizar un procedimiento de detección de objetos buscando áreas homogéneas, cuya parametrización (centro de masas, área, ...) servirá de entrada al sistema y determinará la respuesta del mismo en función de las características de los objetos.

El proyecto se ha desarrollado en C++ y, como interfaz de programación, se ha utilizado la librería DirectX-8 de Microsoft.

## 4 Conclusiones

La solución presentada, utilizando la pantalla del portátil como escritorio, permite evaluar la viabilidad de la nueva metáfora, así como experimentar con diferentes modos de interacción. No obstante, tanto el pequeño tamaño del escritorio, como las prestaciones limitadas en cuanto a ángulo de visión, hacen pensar en configuraciones más profesionales, utilizando proyectores de vídeo LCD que generen una imagen de mayor tamaño en el escritorio y sobre el cual, se dispondría la cámara.

Por otro lado, dadas las limitadas prestaciones de la cámara empleada, en aplicaciones donde la precisión (en cuanto a situación de los objetos y determinación de tamaño de los mismos) sea importante, el proceso de visión por computador se simplificará usando otros modelos de cámara y de sistema de adquisición.

En el futuro, se piensan añadir nuevas prestaciones al sistema incluyendo entre otras: seguimiento activo del movimiento de los objetos, otros modos de respuesta con apoyo de audio o de sonidos instrumentales, así como el reconocimiento del habla o de sonidos. También se va a considerar la utilización de algún código sencillo de gestos para utilizar las manos como un objeto más, cuyo significado venga dado por la posición que, de acuerdo a ese código, adopte. Por último, la clasificación de objetos permitiría incorporar objetos de la vida cotidiana para su utilización en el escritorio (tazas, platos, ...).

## 5 Referencias

1. Seybold, J.: "Xerox's `Star`". The Seybold Report 10,16.
2. Bruce Damer's Personal Histories of the Desktop User Interface. A Retrospective of the Xerox Alto, Star 8010 System and Elixir Desktop at the Birth of the Desktop Metaphor. (<http://www.damer.com/pictures/elixir/products/star.html> ).
3. The Story Behind the Lisa (and Macintosh) Interface. (<http://home.san.rr.com/deans/lisagui.html> ).
4. Kurtenbach, G., Fitzmaurice, G., Baudel, T., Buxton, B.: The Design of a GUI Paradigm Based on Tablets, Two-Hands, and Transparency. In: Pemberton, S. (ed.): Proceedings of ACM CHI 97 Conference on Human Factors in Computing Systems, Atlanta, Georgia, Vol 1. Addison-Wesley Pub Co (1997) 35-42.
5. Encarnação, L.M., Bimber, O., Schmalstieg, D., Chandler S.D.: A Translucent Sketchpad for the Virtual Table Exploring Motion-based Gesture Recognition. In: Eurographics'99, Milano, Italy (1999).
6. Arai, T., Machii, K., Kusunuki, S., Shojima, H.: InteractiveDESK: a Computer-Augmented Desk which responds to operations on Real Objects. In: CHI'95 Conference companion on Human factors in computing systems, Denver, Colorado (1995) 141-142.
7. Maes, P., Darrell, T., Blumberg, B., Pentland, A.: The ALIVE System: Full-Body Interaction with Autonomous Agents. In: Proceedings of the Computer Animation '95 Conference, Geneva, Switzerland. IEEE Press (1995) 11-18.