(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 301 256**

(51) Int. Cl.:
H04N 5/232 (2006.01)
H04N 7/15 (2006.01)
G01S 3/786 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Número de solicitud europea: **99964595 .5**
(86) Fecha de presentación : **14.12.1999**
(87) Número de publicación de la solicitud: **1057326**
(87) Fecha de publicación de la solicitud: **06.12.2000**

(54) Título: **Determinación automática de posiciones preajustadas correspondientes a participantes de videoconferencias.**

(30) Prioridad: **22.12.1998 US 218554**

(45) Fecha de publicación de la mención BOPI:
16.06.2008

(45) Fecha de la publicación del folleto de la patente:
16.06.2008

(73) Titular/es: **Koninklijke Philips Electronics N.V.**
Groenewoudseweg 1
5621 BA Eindhoven, NL

(72) Inventor/es: **Cohen-Solal, Eric;**
Martel, Adrian, P.;
Sengupta, Soumitra;
Strubbe, Hugo;
Caviedes, Jorge;
Abdel-Mottaleb, Mohamed y
Elgammal, Ahmed

(74) Agente: **Zuazo Araluze, Alexander**

DESCRIPCIÓN

Determinación automática de posiciones de preajustadas correspondientes a participantes de videoconferencias.

Esta invención se refiere al campo de tecnología de videoconferencia y específicamente a un procedimiento para determinar automáticamente los parámetros de giro, inclinación y zoom apropiados de una cámara que corresponden a vistas deseadas de participantes en un escenario de videoconferencia.

Durante una videoconferencia es necesario saber los parámetros de la cámara apropiados para cada participante de modo que la vista de la cámara pueda cambiar rápidamente de un participante a otro. Estos parámetros incluyen el zoom, el giro y la inclinación apropiados de la cámara, y se denominarán colectivamente como "parámetros" de la cámara siendo los valores de estos parámetros asociados con cada participante los "preajustes". Mientras tiene lugar la conferencia, los usuarios requieren la capacidad de poder ver a diferentes participantes rápidamente; cambiando frecuentemente de un participante a otro en un breve periodo de tiempo.

Dispositivos de la técnica anterior requieren que un usuario ajuste manualmente los parámetros de la cámara para cada participante implicado en la videoconferencia. Cada cámara que se utiliza enfoca a un participante y se acciona un conmutador de preajuste. Por ejemplo, si hay tres personas en la conferencia, el conmutador 1 se utiliza para representar los parámetros de la cámara apropiados para el participante 1; el conmutador 2 para el participante 2; y el conmutador 3 para el participante 3. Cuando un usuario desea conmutar la vista entre el participante 1 y el 2, sólo necesita activar el conmutador 2 y la cámara se mueve y enfoca en consecuencia. Sin embargo, ajustar una cámara para cada participante es frecuentemente un proceso tedioso que requiere dedicación de tiempo por parte del usuario u operador de cámara. Adicionalmente, cada vez que un participante abandona o entra en la sala, los preajustes tienen que reajustarse en consecuencia. Si un participante simplemente se mueve de su ubicación original, los ajustes de la cámara originales ya no se aplicarán. Claramente este es un problema si un participante se mueve de una ubicación a otra dentro de la sala. Sin embargo, incluso si el participante se mueve dentro de su propia silla (es decir, hacia delante, hacia atrás, se inclina hacia un lado, etc.) los parámetros pueden cambiar y ese participante puede ya no estar enfocado, en el centro de la vista de la cámara, o del tamaño deseado con respecto a la vista de la cámara.

En la patente estadounidense 5.598.209, un usuario puede apuntar a un objeto o persona que desea ver y el sistema almacena automáticamente los parámetros de giro e inclinación de la cámara que se refieren al centro de ese objeto. Sin embargo, todos los objetos o personales en la sala tienen que seleccionarse y almacenarse con afirmación bajo el control de un usuario, lo que de nuevo lleva tiempo. Tampoco se proporciona actualizar los parámetros cuando un participante abandona o entra en la sala.

La capacidad de determinar automáticamente posiciones de preajuste es útil también en una distribución de congreso. En general, en estos tipos de salas, los preajustes de la cámara se basan en el micrófono que se utiliza para cada individuo. Cuando un participante enciende su micrófono, se utilizan los preajustes de la cámara que se refieren a la posición de ese micrófono. Esto es problemático porque si el micrófono no funciona o si otro hablante utiliza un micrófono particular, podría no tener lugar la correlación apropiada entre el hablante y la vista de la cámara.

Por lo tanto, existe una necesidad de un sistema de videoconferencia que determine automáticamente los parámetros de la cámara apropiados para todos los participantes y que también pueda autoajustarse cuando los participantes entran en y abandonan la sala. El objetivo de una videoconferencia es la conversación y comunicación eficaces. Si un usuario tiene que reajustar continuamente el sistema para inicializar o actualizar los parámetros de preajuste, se frustra este objetivo. La dinámica de conversación entre los usuarios finales es diferente de la de una producción (como en un programa de televisión). Para facilitar esta dinámica, es deseable automatizar tanto del sistema como sea posible sin recurrir a una vista de zoom de alejamiento estática que proporcionaría comunicación menos significativa.

Un aspecto de la invención es un procedimiento de cálculo de preajustes de los parámetros de la cámara correspondientes a participantes en un sistema de videoconferencia. El procedimiento incluye proporcionar una cámara que tenga parámetros de giro, inclinación y zoom, y definir un espacio basándose en una distribución del sistema de videoconferencia. El procedimiento incluye además realizar uno de mover la cámara a través de todos los valores de giro pertinentes, definiéndose los valores de giro pertinentes por el espacio en el que está ubicado el sistema de videoconferencia, y hacer un zoom de alejamiento en la cámara de tal modo que pueda verse a todos los posibles participantes mediante la cámara y de tal modo que pueda determinarse una ubicación de cada participante en el espacio. El procedimiento prevé además detectar los participantes dentro del espacio y detectar los preajustes correspondientes a los participantes, definiendo los preajustes una vista de la cámara, basándose los preajustes en al menos una de una posición óptima de los participantes en la vista de la cámara, una alineación del centro de una cabeza de los participantes con un centro de la vista de la cámara, y una alineación de un centro de un participante con el centro de la vista de la cámara.

Este aspecto, como los siguientes, permite la detección y actualización automáticas de parámetros de la cámara correspondientes a participantes en una videoconferencia.

Según otro aspecto de la invención, un sistema de videoconferencia incluye al menos una cámara que tiene parámetros de giro, inclinación y zoom. Los parámetros tienen valores de preajuste asignados a participantes correspondientes del sistema de videoconferencia. Cada uno de los preajustes define una vista de la cámara y se determinan mediante:

uno de realizar un movimiento panorámico y hacer zoom en la cámara por todo un espacio definido por el sistema de videoconferencia, detectar un participante, y definir un preajuste basándose en una posición de la cámara que colocaría al participante en una de una posición óptima, una posición en la que una cabeza del participante está en alineación con un centro de la vista de la cámara, y una posición en la que un centro del participante está alineado con el centro de la vista de la cámara.

Según aún otro aspecto de la invención, un sistema de videoconferencia comprende al menos una cámara que tiene parámetros de giro, inclinación y zoom. Los parámetros tienen valores de preajuste asignados a participantes correspondientes del sistema de videoconferencia; definiendo los preajustes una vista de la cámara. El sistema incluye además al menos uno de medios de giro para que la cámara realice un movimiento panorámico por todo un espacio definido por el sistema de videoconferencia, y medios de zoom para hacer zoom de alejamiento en la cámara para permitir de ese modo que la cámara vea el espacio definido por el sistema de videoconferencia. Se utilizan medios de detección para detectar a los participantes en el espacio. Se utilizan medios de determinación para determinar los preajustes de la cámara basándose en una posición de la cámara que colocaría a uno de los participantes en una de una posición óptima, una posición en la que una cabeza del participante está en alineación con un centro de dicha vista de la cámara, y una posición en la que un centro del participante está alineado con el centro de la vista de la cámara.

Es un objeto de la invención proporcionar un sistema y procedimiento de videoconferencia que pueda determinar automáticamente los preajustes de los parámetros de la cámara que se refieren a las vistas apropiadas de los participantes.

Es otro objeto de la invención proporcionar un sistema y procedimiento de videoconferencia que pueda actualizar continuamente preajustes de la cámara según los cambios en el número y ubicación de los participantes.

Estos objetos, así como otros, serán más evidentes a partir de la siguiente descripción leída en conjunción con los dibujos adjuntos, en los que se prevé que números de referencia iguales designen los mismos elementos.

Las figuras 1A, 1B y 1C son diagramas de distribuciones de sala, congreso y mesa respectivamente de un sistema de videoconferencia según la invención;

las figuras 2A, 2B y 2C son diagramas que muestran a un participante entrando en una vista de la cámara cuando la cámara realiza un movimiento panorámico en una sala en un sistema de videoconferencia según la invención;

la figura 3 es un modelo en perspectiva de una cámara utilizada en la invención;

la figura 4 es un diagrama que muestra a los participantes en una videoconferencia con preajustes temporales respectivos indicados;

la figura 5 es un diagrama que muestra el centro de un participante desplazado del centro de la vista de la cámara de ese participante;

la figura 6 es un diagrama que muestra a los participantes en una videoconferencia con preajustes actualizados respectivos indicados;

la figura 7 es un diagrama que muestra una realización alternativa de la invención que utiliza dos cámaras;

la figura 8 es un diagrama de un sistema de coordenadas cilíndricas utilizado para colores de representación de píxeles en las imágenes;

la figura 9 son tres gráficos que representan proyecciones del espacio de color YUV que indican las zonas en las que se encuentras los píxeles del color de la piel;

las figuras 10A a 10F son imágenes originales e imágenes binarias respectivas que se forman separando los píxeles basándose en el color;

la figura 11 es un diagrama que ilustra cómo se utiliza una máscara 3x3 como parte de la detección de la variación de luminancia según la invención;

las figuras 12A y 12B son diagramas que ilustran conectividad de tipo 4 y 8 respectivamente;

las figuras 13A y 13B son imágenes que muestran cómo aparecería la imagen de las figuras 3C y 3E después de que se eliminen los bordes según la invención;

la figura 14 es una imagen que muestra ejemplos de cuadros delimitadores aplicados a la imagen de la figura 3F;

la figura 15 es una secuencia de diagramas que muestra cómo se representan los componentes de una imagen por vértices y se conectan para formar un gráfico según la invención;

las figuras 16A a 16D son una secuencia de imágenes que ilustran la aplicación de una heurística según la invención; y

la figura 17 es un diagrama de flujo que detalla las etapas generales implicadas en la detección de caras.

En la figura 1A, se muestra un sistema de videoconferencia en el que los participantes están sentados alrededor de una mesa. La figura 1B muestra los participantes en una distribución de estilo congreso. Un cámara 50 se controla mediante un controlador 52 para realizar un movimiento panorámico desde un lado de la sala al otro. Claramente, el movimiento panorámico puede comenzar y terminar en el mismo lugar. Por ejemplo, tal como se muestra en la figura 1C, la cámara 50 podría disponerse en el medio de una sala con los participantes ubicados todos alrededor de la misma. En este tipo de situación, la cámara 50 giraría complemente en un círculo con el fin de realizar un movimiento panorámico completo en toda la sala. En la distribución de congreso mostrada en la figura 1B, la cámara 50 podría realizar múltiples trayectorias panorámicas para cubrir las diferentes filas. Cada una de estas trayectorias tendría una inclinación diferente y probablemente un zoom diferente (aunque el zoom puede ser el mismo si los participantes están colocados directamente unos por encima de otros a sustancialmente la misma distancia radial desde la cámara). De nuevo, en la distribución de congreso, la cámara 50 podría disponerse en el centro de la sala y entonces el movimiento panorámico puede requerir un giro completo tal como se muestra en la figura 1C.

Para mayor simplicidad, ahora se describirá adicionalmente la distribución mostrada en la figura 1A aunque debería ser evidente que se aplicarían las mismas ideas a todas las distribuciones mencionadas y también a otras distribuciones evidentes para los expertos en la técnica. La invención funcionará para cualquier espacio definido por la ajustabilidad del sistema de videoconferencia. Se muestran tres participantes ($Part_A$, $Part_B$, $Part_C$) pero, de nuevo, podrían implicarse más participantes.

Cuando la cámara 50 realiza un movimiento panorámico desde un lado de la sala al otro, los participantes parecerán moverse por y a través de la vista de la cámara. Tal como se muestra en las figuras 2A a 2C, un participante aparece en diferentes partes de la vista de la cámara dependiendo de la posición de giro de la cámara. Tal como puede distinguirse también a partir de la figura, para tres posiciones de giro diferentes (P1, P2, P3) la inclinación (T) y el zoom (Z) permanecen iguales. También es posible que durante la exploración de la cámara inicial, pudiera moverse uno de los otros parámetros (es decir, la inclinación o el zoom) a través de un intervalo apropiado mientras que los dos parámetros restantes se mantienen constantes. Otra posibilidad es si la cámara 50 tiene su ajuste de parámetro de zoom de tal modo que pudiera verse de una vez toda la sala (suponiendo que puede recogerse suficiente información para determinar la posición de participantes estacionarios tal como se da a conocer posteriormente de manera más clara). De nuevo, para mayor simplicidad, se describirá la idea de realizar un movimiento panorámico con la cámara pero debería ser evidente que las otras sugerencias podrían implementarse con cambios apropiados que estarían claros para los expertos en la técnica.

Durante el inicio de la realización de un movimiento panorámico, cada fotograma que procesa la cámara se analiza para determinar si un participante está dispuesto dentro del fotograma. Un procedimiento para realizar esta determinación se detalla posteriormente en la sección de detección de participantes. Claramente, podrían implementarse otros procedimientos. Para cada participante que se detecta, una cámara que realiza un movimiento panorámico detectará una multiplicidad de fotogramas que incluirían a ese participante. Por ejemplo, si una cámara procesa mil fotogramas para una sala, éste podría interpretarse como que son mil participantes, si se muestra un participante en cada fotograma.

Para evitar este problema de multiplicar el número real de participantes, se etiqueta cada participante detectado. Se calcula el centro de masas para cada participante detectado para cada fotograma procesado. Entonces, se compara un segundo fotograma, sucesivo que contiene participantes potenciales con el primer fotograma, anterior para ver si la cámara está viendo a un nuevo participante o sólo otro fotograma que incluye al mismo participante. Un procedimiento para llevar a cabo esta comparación es realizar una extrapolación geométrica basada en el primer centro y la cantidad que la cámara se ha movido desde la primera posición. Esto mostraría aproximadamente dónde debería estar el centro si el segundo fotograma contiene al mismo participante que el primer fotograma. De manera similar, se calcularía el centro de masas del segundo fotograma y luego se compararía con el primer centro junto con el movimiento conocido de la cámara entre la posición en la que se ve el primer fotograma y la posición en la que se ve el segundo fotograma. Como alternativa, podría crearse una signatura para cada participante detectado y entonces podrían compararse las signaturas de los participantes en fotogramas sucesivos con esa signatura inicial. Las signaturas se conocen en la técnica. Algunos ejemplos de técnicas de signatura se tratan posteriormente en la sección de identificación de participantes y de actualización de posición. Una vez que se determina que la imagen de un participante está dispuesta dentro de un fotograma, pueden calcularse preajustes temporales.

En referencia a la figura 3, se muestra un modelo en perspectiva de una cámara. Un sensor 56 de la cámara tiene un punto principal PP que tiene una coordenada x e y PPx y PPy respectivamente. Una lente 58 tiene un centro que está dispuesto a una longitud focal f del punto principal PP. Un cambio en el zoom de la cámara se lleva a cabo mediante un cambio en la distancia focal f. Una f más corta significa una vista amplia ("alejamiento de zoom"). Un cambio en el parámetro de giro es efectivamente un giro del sensor alrededor del eje de giro. Un cambio en el parámetro de inclinación es un giro del sensor alrededor del eje de inclinación.

Cuando un objeto o participante 62 entra en el campo de visión de la cámara, puede determinarse la ubicación de ese participante en el espacio utilizando procedimientos convencionales si están disponibles dos fotogramas que

contienen a ese participante. Esto es porque se conoce la ubicación del punto principal PP (no mostrado en 60) y el enfoque f. Cuando la cámara 50 realiza un movimiento panorámico en una sala, adquiere múltiples fotogramas que contienen participantes y así puede determinarse la ubicación de cada participante en el espacio. Si la cámara está haciendo un zoom de alejamiento en lugar de un movimiento panorámico, pueden necesitarse dos mediciones distintas para determinar la ubicación. Una vez que se conoce la ubicación de un participante, puede calcularse el preajuste temporal mediante un procesador 54 (figuras 1A a 1C).

Para calcular el preajuste temporal, se determina el centro de un participante, tal como anteriormente para el etiquetado de participantes, utilizando técnicas conocidas. Por ejemplo, puede calcularse la media del contorno del participante y su centro de masas. El punto central se coloca entonces en el centro de la vista de la cámara para producir, por ejemplo, los preajustes Psa, Tsa y Zsa para el $Part_A$ de la figura 1. Estos procesos de realización de un movimiento panorámico y cálculo de preajustes se repiten para todos los participantes en la sala y, en consecuencia, también determina cuántos participantes hay inicialmente en la sala. Esto se realiza durante una parte de iniciación de la conferencia y puede repetirse posteriormente durante una rutina de actualización tal como se describe posteriormente de manera más completa.

Una vez que todos los participantes en la sala están etiquetados y se calculan todos los parámetros temporales tal como se muestra en la figura 4, la cámara 50 realiza un segundo movimiento panorámico (o zoom de alejamiento) en la sala. Cada vista de preajuste se perfecciona adicionalmente porque la calibración realizada en la fase de realización de movimiento panorámico inicial no será en general lo suficientemente precisa.

Tal como se muestra en la figura 5, el centro de la vista de la cámara se compara con el centro de la cabeza de cada participante respectivo. Los parámetros se ajustan de tal modo que en la vista de la cámara, se alinean los centros. Una vez que se perfecciona el preajuste, se calcula el preajuste correspondiente a una vista “óptima” de cada participante. Esto puede ser diferente dependiendo de las culturas sociales. Por ejemplo, la cabeza y torso de un participante puede ocupar cualquier lugar del 30 al 60% de todo el fotograma, tal como en un programa de noticias en Estados Unidos. La vista óptima produce preajustes actualizados Psn’, Tsn’ y Zsn’ tal como se muestra en la figura 6. Estos valores se actualizan continuamente dependiendo de cómo se estructure el sistema y cómo deben realizarse las actualizaciones tal como se explica posteriormente. Si una cámara está mirando a un participante y ese participante se mueve, se calcularía la nueva posición óptima y el preajuste de la cámara se ajustará continuamente en consecuencia.

La cámara puede enfocar a participantes basándose en seguimiento de audio, seguimiento de vídeo, una selección realizada por un usuario, o mediante cualquier otra técnica conocida en la técnica. El seguimiento de audio por sí solo está limitado porque disminuye en precisión a medida que las personas se alejan y no puede utilizarse por sí mismo porque generalmente tiene un error de 4 a 5 grados y no puede haber seguimiento cuando un participante para de hablar.

Puede asociarse un nombre con cada participante una vez que se detecta. Por ejemplo, los tres participantes de la figura 1 podrían identificarse como A, B y C de tal modo que un usuario podría simplemente indicar que desea ver al participante A y la cámara se moverá al preajuste optimizado para A. Adicionalmente, el sistema podría programarse para aprender algo específico sobre cada participante y por tanto etiquetar a ese participante. Por ejemplo, podría crearse una signatura para cada participante, el color de la camiseta de la persona, podría tomarse un patrón de voz, o podría utilizarse una combinación de la cara y la voz para formar la etiqueta asociada con un participante. Con esta información adicional, si el participante A se mueve por la sala, el sistema sabrá qué participante está moviéndose y no estará confundido porque el participante A ande a través de la vista correspondiente a parámetros para el participante B. Además, si dos participantes están ubicados lo suficientemente próximos el uno al otro de tal modo que comparten una vista de la cámara, los dos participantes pueden considerarse como un participante con la cámara enfocando al centro de la combinación de sus imágenes.

Tal como se expuso anteriormente, un beneficio de este sistema es que permite que se ajusten automáticamente los preajustes cuando cambia la dinámica de los participantes de la sala. Claramente, si se selecciona un preajuste y el participante correspondiente ha abandonado la sala, el sistema lo detectará y actualizará los preajustes. Otro procedimiento de actualización es que cada vez que se selecciona un nuevo preajuste, la cámara 50 hará un zoom de alejamiento (o un movimiento panorámico en la sala) para ver si alguien ha entrado en o ha abandonado la sala y actualizará los preajustes antes de que la cámara 50 se mueva al preajuste seleccionado. La cámara 50 podría controlarse periódicamente, incluso mientras se le ordena ver a un participante seleccionado, detener temporalmente la visión de ese participante, y realizar un movimiento panorámico en la sala o zoom de alejamiento para ver si el número de participantes ha cambiado. Otra técnica es reconocer que un participante no está donde debería estar. Por ejemplo, si se le dice a la cámara 50 que se mueva desde el preajuste para el participante C al participante A por ejemplo (figura 1), si el participante B ha abandonado la sala, el sistema podría aprenderlo y realizar los ajustes apropiados. Aún otra técnica de actualización implica que la cámara 50 realice un movimiento panorámico a través de la sala (o zoom de alejamiento) o bien periódicamente o bien cada vez que se selecciona un nuevo preajuste.

En referencia a la figura 7, se muestra una segunda realización. Esta realización muestra las mismas características que las de la figura 1A excepto que se añade una segunda cámara 64. La calibración inicial se realiza de la misma manera que se describió anteriormente. Sin embargo, durante la conferencia, se utiliza una cámara para enfocar al participante pertinente mientras que la otra se utiliza para actualizar continuamente los preajustes. La cámara de actualización puede estar continuamente en un zoom de alejamiento de tal modo que pueda determinar cuándo un

participante abandona o entra en la sala. Como alternativa, la cámara de actualización podría realizar continuamente un movimiento panorámico en la sala y realizar las actualizaciones apropiadas para los preajustes. Las dos cámaras comparten la información de preajustes a través, por ejemplo, de un procesador 54. Claramente, podrían utilizarse más cámaras. Por ejemplo, podría asignarse una cámara a cada individuo que se planee que esté en la reunión y entonces podría utilizarse una cámara adicional como la cámara de actualización.

Una manera de determinar si un participante está ubicado dentro de una vista de la cámara es determinar si hay una cara dispuesta dentro de la imagen que se ve mediante la cámara. Cada píxel en una imagen se representa generalmente en el espacio de color HSV (tonalidad, saturación, valor). Estos valores se mapean sobre un sistema de coordenadas cilíndricas tal como se muestra en la figura 8, donde P es un valor (o luminancia), θ es la tonalidad, y r es la saturación. Debido a la no linealidad de sistemas de coordenadas cilíndricas, se utilizan otros espacios de color para aproximar el espacio HSV. En la presente solicitud, se utiliza el espacio de color YUV porque la mayoría del material de vídeo almacenado en un medio magnético y el estándar MPEG2 utilizan ambos este espacio de color.

Transformar una imagen RGB al espacio YUV, y proyectar además en los planos VU, VY y VU, produce gráficos como los mostrados en la figura 9. Los segmentos de círculo representan la aproximación del espacio HSV. Cuando se representan los píxeles correspondientes al color de la piel en el espacio YUV, caen generalmente en esos segmentos de círculo mostrados. Por ejemplo, cuando la luminancia de un píxel tiene un valor entre 0 y 200, la crominancia U tiene generalmente un valor entre -100 y 0 para un píxel del color de la piel. Estos son valores generales basados en la experimentación. Claramente, podría realizarse una operación de entrenamiento de color para cada cámara que se utiliza. Los resultados de ese entrenamiento se utilizarían entonces para producir segmentos del color de la piel más precisos.

Para detectar una cara, se examina cada píxel en una imagen para distinguir si es del color de la piel. Aquellos píxeles que son del color de la piel se agrupan respecto al resto de la imagen y por tanto se quedan como candidatos a cara potenciales. Si al menos una proyección de un píxel no cae dentro de los límites del segmento de agrupamiento de la piel, se considera que el píxel no es del color de la piel y se excluye de la consideración como un candidato a cara potencial.

La imagen resultante formada por la detección del color de la piel es binaria porque muestra o bien partes de la imagen que son del color de la piel o bien partes que no son del color de la piel tal como se muestra en las figuras 10B, 10 D y 10F que corresponden a las imágenes originales en las figuras 10A, 10C y 10E. En las figuras, se muestra blanco para el color de la piel y negro para el color que no es de la piel. Tal como se muestra en las figuras 10A y 10 B, esta etapa de detección por sí sola puede descartar que grandes partes de la imagen tengan una cara dispuesta dentro de la misma. Técnicas de la técnica anterior que utilizan el color y la forma pueden funcionar por tanto para fondos sencillos tales como el mostrado en la figura 10A. Sin embargo, mirando a las figuras 10C y 10D y a las figuras 10E y 10F, está claro que la detección por sólo color y forma puede no ser suficiente para detectar las caras. En las figuras 10C a 10F, objetos en el fondo como cuero, madera, ropas, y pelo, tienen colores similares a la piel. Tal como puede verse en las figuras 10D y 10F, estos objetos del color de la piel están dispuestos inmediatamente adyacentes a la piel de las caras y por tanto las propias caras son difíciles de detectar.

Después de que los píxeles se separan por color, los píxeles ubicados en los bordes se excluyen de la consideración. Un borde es un cambio en el nivel de brillo de un píxel al siguiente. La eliminación se lleva a cabo tomando cada píxel del color de la piel y calculando la varianza en los píxeles alrededor del mismo en la componente de luminancia; siendo indicativa una alta varianza de un borde. Tal como se muestra en la figura 11, se coloca un cuadro ("ventana") del tamaño o bien de 3x3 o bien de 5x5 píxeles, en la parte superior de un píxel del color de la piel. Claramente, podrían utilizarse otras máscaras además de un cuadro cuadrado. La varianza se define como

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_x)^2$$

donde $\bar{a}_x$ es la media de todos los píxeles en la ventana examinada. Un nivel de varianza "alto" será diferente dependiendo de la cara y la cámara utilizada. Por lo tanto, se utiliza una rutina iterativa empezando con un nivel de varianza muy alto y bajando hasta un nivel de varianza bajo.

En cada etapa de la iteración de varianza, se excluyen de la consideración facial los píxeles si la varianza en una ventana alrededor del píxel del color de la piel es superior al umbral de varianza que se prueba para esa iteración. Después de que se examinan todos los píxeles en una iteración, se examinan las componentes conectadas resultantes en busca de características faciales tal como se describe posteriormente de manera más completa. Las componentes conectadas son píxeles que son del mismo valor binario (blancos para color facial) y están conectadas. La conectividad puede ser conectividad o de tipo 4 u 8. Tal como se muestra en la figura 12A, para conectividad de tipo 4, el píxel central se considera "conectado" a sólo los píxeles directamente adyacentes al mismo tal como se indica mediante el "1" en los cuadros adyacentes. En conectividad de tipo 8, tal como se muestra en la figura 12B, los píxeles que tocan diagonalmente el píxel central también se consideran que están "conectados" a ese píxel.

Tal como se expuso anteriormente, después de cada iteración, se examinan las componentes conectadas en una etapa de clasificación de componentes para ver si podrían ser una cara. Este examen implica estudiar 5 criterios

distintos basándose en un cuadro delimitador dibujado alrededor de cada componente conectada resultante; ejemplos de lo cual se muestran en la figura 14 basada en la imagen de la figura 10E. Los criterios son:

1. El área del cuadro delimitador comparado con un umbral. Esto reconoce el hecho de que una cara no será en general muy grande o muy pequeña.

2. La relación de aspecto (altura comparada con el ancho) del cuadro delimitador comparada con un umbral. Esto reconoce que las caras humanas caen generalmente en un intervalo de relaciones de aspecto.

3. La relación del área de píxeles del color de la piel detectados con el área del cuadro delimitador, comparada con un umbral. Este criterio reconoce el hecho de que el área cubierta por una cara humana caerá en un intervalo de porcentajes del área del cuadro delimitador.

4. La orientación de objetos alargados dentro del cuadro delimitador. Hay muchas maneras conocidas de determinar la orientación de una serie de píxeles. Por ejemplo, puede determinarse el eje medio y puede encontrarse la orientación a partir de ese eje. En general, las caras no están giradas significativamente alrededor del eje ("eje-z") que es perpendicular al plano que tiene la imagen y por tanto las componentes con objetos alargados que están giradas con respecto al eje z se excluyen de la consideración.

5. La distancia entre el centro del cuadro delimitador y el centro de masas de la componente que se examina. En general, las caras están ubicadas dentro del centro del cuadro delimitador y no estarán, por ejemplo, ubicadas totalmente a un lado.

Se continúan las iteraciones para la varianza descomponiendo de ese modo la imagen en componentes más pequeñas hasta que el tamaño de las componentes es inferior a un umbral. Las imágenes de las figuras 10C y 10E se muestran transformadas en las figuras 13A y 13B respectivamente después del proceso de iteración de varianza. Tal como puede distinguirse, las caras en la imagen se separaron de las zonas del color de la piel no faciales en el fondo como resultado de la iteración de variación. Frecuentemente, esto provoca que la zona con color de la piel detectado se fragmente como se muestra a modo de ejemplo en la figura 13B. Esto tiene lugar porque o bien hay objetos que ocultan partes de la cara (como gafas o vello facial) o porque se eliminaron partes debido a una alta varianza. Por tanto, sería difícil buscar una cara utilizando las componentes resultantes por sí mismas. Las componentes que todavía pueden ser parte de la cara después de las etapas de iteración de varianza y clasificación de componentes, se conectan para formar un gráfico tal como se muestra en la figura 15. De esta manera, las componentes del color de la piel que tienen características similares, y están próximas en el espacio, se agrupan juntas y se examinan adicionalmente.

En referencia a la figura 15, cada componente resultante (que sobrevive a las etapas de detección de color, eliminación de bordes, y clasificación de componentes) se representa por un vértice de un gráfico. Los vértices se conectan si están próximos en el espacio en la imagen original y si tienen un color similar en la imagen original. Dos componentes, i y j, tienen un color similar si:

$$|Y_i - Y_j| < t_y \wedge |U_i - U_j| < t_u \quad \text{Y LÍNEA} \quad V_i - V_j| < t_v$$

donde Y_n, U_n y V_n son los valores medios de la luminancia y crominancia de la enésima componente y t_n son valores umbrales. Los umbrales se basan en variaciones en los valores Y, U y V en las caras y se mantienen lo suficientemente altos de tal modo las componentes de la misma cara se considerarán similares. Las componentes se consideran próximas en el espacio si la distancia entre las mismas es inferior a un umbral. El requisito espacial garantiza que las componentes distantes espacialmente no se agrupan juntas porque las partes de una cara no estarían ubicadas normalmente en partes distantes espacialmente de una imagen.

La conexión entre vértices se denomina un borde. Se le da a cada borde un peso que es proporcional a la distancia euclídea entre los dos vértices. Conectar los vértices juntos dará como resultado un gráfico o un conjunto de gráficos inconexos. Para cada uno de los gráficos resultantes, se extrae el árbol de expansión mínima. El árbol de expansión mínima se define en general como el subconjunto de un gráfico en el que todos los vértices todavía están conectados y la suma de las longitudes de los bordes del gráfico es tan pequeña como sea posible (mínimo peso). Las componentes correspondientes a cada gráfico resultante se clasifican entonces como o bien cara o bien no cara utilizando los parámetros de forma definidos en la etapa de clasificación de componentes mencionada anteriormente. Entonces se divide cada gráfico en dos gráficos eliminando el borde más débil (el borde con el mayor peso) y las componentes correspondientes de los gráficos resultantes se examinan de nuevo. La división continúa hasta que un área de un cuadro delimitador formado alrededor de los gráficos resultantes es inferior a un umbral.

Descomponiendo y examinando cada gráfico en busca de una cara, se determina un conjunto de todas las posibles ubicaciones y tamaños de caras en una imagen. Este conjunto puede contener un gran número de falsos positivos y por ello se aplica una heurística para eliminar algunos de los falsos positivos. Buscar todas las características faciales (es decir, nariz, boca, etc.) requeriría una plantilla, lo que proporcionaría un espacio de búsqueda demasiado grande. Sin embargo, la experimentación ha mostrado que esas características faciales tienen bordes con una alta varianza. Muchos falsos positivos pueden eliminarse examinando la relación de píxeles de alta varianza dentro de una cara potencial con el número total de píxeles en la cara potencia.

La heurística mencionada anteriormente se lleva a cabo aplicando en primer lugar una operación de cierre morfológico a los candidatos faciales dentro de la imagen. Tal como se conoce en la técnica, se elige y se aplica una máscara a cada píxel dentro de una zona facial potencial. Por ejemplo, podría utilizarse una máscara 3x3. Se aplica un algoritmo de dilatación para expandir los bordes de componentes candidatas a cara. Entonces se utiliza un algoritmo de erosión para eliminar píxeles de los bordes. Un experto en la técnica apreciará que estos dos algoritmos, realizados en este orden, rellenarán los huecos entre las componentes y también mantendrá las componentes a sustancialmente la misma escala. Claramente, se podrían realizar etapas de dilataciones múltiples y luego de erosiones múltiples siempre que ambas se apliquen un número igual de veces.

Ahora, la relación de píxeles con una vecindad de alta varianza dentro de la zona candidata a cara se compara con el número total de píxeles en la zona candidata a cara. En referencia a las figuras 16A a 16D, se examina una imagen original en la figura 16A en busca de candidatos a cara potenciales utilizando los procedimientos descritos anteriormente para conseguir la imagen binaria mostrada en la figura 16B. La operación de cierre morfológico se realiza sobre la imagen binaria dando como resultado la imagen mostrada en la figura 16C. Finalmente, se detectan los píxeles con alta varianza ubicados en la imagen de la figura 16C tal como se muestra en la figura 16D. Entonces puede determinarse la relación de los píxeles de alta varianza con el número total de píxeles. Todo el procedimiento de detección de participantes se resume mediante las etapas S2 a S16 mostradas en la figura 17.

Tal como puede distinguirse, controlando una cámara para ver un espacio definido por un sistema de videoconferencia, pueden calcularse automáticamente y actualizarse continuamente preajustes de parámetros de la cámara correspondientes a participantes.

Habiendo descrito las realizaciones preferidas debería ser evidente que podrían realizarse diversos cambios sin apartarse del alcance de la invención que se define mediante las reivindicaciones adjuntas.

REIVINDICACIONES

1. Procedimiento de cálculo de preajustes de parámetros de la cámara correspondientes a participantes (Part A, Part B, Part C) en un sistema de videoconferencia, comprendiendo dicho procedimiento:

- proporcionar una cámara que tenga parámetros (50) de giro, inclinación y zoom;

- definir un espacio basándose en una distribución de dicho sistema de videoconferencia;

realizando uno de

- mover dicha cámara a través de todos los valores de giro pertinentes, definiéndose dichos valores de giro pertinentes por dicho espacio en el que está ubicado dicho sistema de videoconferencia, y

- hacer un zoom de alejamiento en dicha cámara de tal modo que pueda verse a todos los posibles participantes mediante dicha cámara y de tal modo que pueda determinarse una ubicación de cada participante en dicho espacio;

- detectar y etiquetar dichos participantes para obtener participantes etiquetados dentro de dicho espacio;

- calcular dichos preajustes correspondientes a dichos participantes etiquetados, definiendo dichos preajustes una vista de la cámara, basándose dichos preajustes en al menos una de: (i) una posición óptima de dichos participantes etiquetados en dicha vista de la cámara, (ii) una alineación del centro de una cabeza de dichos participantes etiquetados con un centro de dicha vista de la cámara, y (iii) una alineación de un centro de un participante etiquetado con dicho centro de dicha vista de la cámara; y

- actualizar los preajustes asociados con un participante etiquetado particular si ha cambiando una ubicación del participante etiquetado particular, siendo la actualización continua, periódica, o cuando se selecciona un nuevo preajuste.

2. Procedimiento según la reivindicación 1, que comprende además: proporcionar al menos una segunda cámara para actualizar dichos preajustes ejecutando dicha actuación.

3. Procedimiento según la reivindicación 1, que comprende además el seguimiento de dichos participantes etiquetados.

4. Procedimiento según la reivindicación 1, comprendiendo además la etapa de actualizar dicho preajuste actualizar dichos preajustes teniendo dicho sistema de videoconferencia que realizar al menos uno de ajustar un preajuste cuando un usuario elige ese preajuste, borrar un preajuste cuando el participante correspondiente al preajuste abandona dicho espacio, y repetir dicha actuación.

5. Procedimiento según la reivindicación 1, en el que en dicha etapa de cálculo, cuando más de un participante está dentro de dicha vista de la cámara, los participantes se combinan en una imagen combinada y el centro de la imagen combinada se utiliza para determinar dichos preajustes.

6. Procedimiento según la reivindicación 1, en el que dicha etapa de detección comprende:

- proporcionar una imagen digital compuesta por una pluralidad de píxeles (52);

- producir una imagen binaria a partir de la imagen digital detectando píxeles (54) del color de la piel;

- eliminar píxeles correspondientes a bordes en la componente de luminancia de dicha imagen binaria produciendo de ese modo componentes (56) de imagen binaria;

- mapear dichas componentes de imagen binaria en al menos un gráfico (512); y

- clasificar dichas componentes de imagen binaria mapeadas como tipos faciales y no faciales en el que los tipos faciales sirven como candidatos (514) faciales.

7. Procedimiento según la reivindicación 6, que comprende además la etapa de aplicar una heurística, incluyendo dicha heurística las siguientes etapas:

- aplicar una operación de cierre morfológico sobre cada uno de dichos candidatos faciales para producir al menos un candidato facial cerrado;

- determinar píxeles de alta varianza en dicho candidato facial cerrado;

- determinar la relación entre dichos píxeles de alta varianza y el número total de píxeles en dicho candidato a cara cerrado; y

- comparar dicha relación con un umbral.

8. Procedimiento según la reivindicación 6, en el que dicha etapa de eliminación incluye:

- aplicar una máscara a una pluralidad de píxeles que incluye un píxel examinado;

- determinar la varianza entre dicho píxel examinado y píxeles dispuestos dentro de dicha máscara; y

- comparar dicha varianza con un umbral de varianza.

9. Procedimiento según la reivindicación 8, en el que:

- dicha etapa de eliminación se repite para disminuir umbrales de varianza hasta que un tamaño de dichas componentes de imagen binaria sea inferior a un umbral de tamaño de componente; y

- después de cada etapa de eliminación se realiza dicha etapa de clasificación de dichas componentes.

10. Procedimiento según la reivindicación 6, en el que dichas componentes de imagen binaria están conectadas.

11. Procedimiento según la reivindicación 6, en el que dicha etapa de clasificación comprende formar un cuadro delimitador alrededor de una componente clasificada de dichas componentes y realizar al menos uno de:

- formar un cuadro delimitador alrededor de una componente clasificada de dichas componentes;

- comparar un área del cuadro delimitador con un umbral de cuadro delimitador;

- comparar una relación de aspecto del cuadro delimitador con un umbral de relación de aspecto;

- determinar una relación de área, siendo dicha relación de área la comparación entre el área de dicha componente clasificada y el área de dicho cuadro delimitador, y comparar dicha relación de área con un umbral de relación de área;

- determinar una orientación de objetos alargados dentro de dicho cuadro delimitador; y

- determinar una distancia entre un centro de dicho cuadro delimitador y un centro de dicha componente clasificada.

12. Procedimiento según la reivindicación 6, en el que dicha etapa de mapeo comprende las siguientes etapas:

- representar cada componente como un vértice;

- conectar vértices con un borde cuando están próximos en el espacio y son similares en color, formando de ese modo dicho al menos un gráfico.

13. Procedimiento según la reivindicación 12, en el que dicho borde tiene un peso asociado y que comprende además las etapas de:

- extraer el árbol de expansión mínima de cada gráfico;

- clasificar las componentes de imagen binaria correspondientes de cada gráfico como o bien cara o bien no cara;

- eliminar el borde en cada gráfico con el mayor peso formando de ese modo dos gráficos más pequeños; y

- repetir dicha etapa de clasificación de las componentes de imagen binaria correspondientes para cada uno de dichos gráficos más pequeños hasta que un cuadro delimitador alrededor de dichos gráficos más pequeños sea inferior a un umbral de gráfico.

14. Sistema de videoconferencia que comprende:

- al menos una cámara que tiene parámetros (50) de giro, inclinación y zoom;

- teniendo dichos parámetros valores asignados a participantes correspondientes de dicho sistema de videoconferencia, siendo los valores preajustes, definiendo dichos preajustes una vista de la cámara;

- al menos uno de medios de giro para que dicha cámara realice un movimiento panorámico por todo un espacio definido por dicho sistema de videoconferencia, y medios de zoom para hacer zoom de alejamiento en dicha cámara para permitir de ese modo que dicha cámara vea el espacio definido por dicho sistema de videoconferencia;

- medios de detección y etiquetado para detectar y etiquetar los participantes para obtener participantes etiquetados en dicho espacio; y

- medios de determinación para determinar los preajustes de dicha cámara basándose en una posición de la cámara que colocaría a uno de dichos participantes etiquetados en una de: (i) una posición óptima, (ii) una posición en la que una cabeza de dicho participante etiquetado está en alineación con un centro de dicha vista de la cámara, y (iii) una posición en la que un centro de dicho participante etiquetado está alineado con dicho centro de dicha vista de la cámara

- medios para actualizar los preajustes asociados con un participante etiquetado particular si la ubicación de un participante etiquetado particular ha cambiando, siendo la actualización continua, periódica, o cuando se selecciona un nuevo preajuste.

15. Sistema de videoconferencia según la reivindicación 14, en el que los medios para la actualización comprenden al menos una segunda cámara para actualizar dichos preajustes.

16. Sistema de videoconferencia según la reivindicación 14, que comprende además medios para el seguimiento de dichos participantes asociando una etiqueta con cada uno de dichos participantes.

17. Sistema de videoconferencia según la reivindicación 14, en el que los medios para la actualización se disponen para actualizar dichos preajustes teniendo dicho sistema de videoconferencia que realizar al menos uno de ajustar un preajuste cuando un usuario elige ese preajuste, borrar un preajuste cuando el participante correspondiente al preajuste abandona dicho espacio, realizar un movimiento panorámico de dicha cámara por dicho espacio, y hacer un zoom en dicha cámara por dicho espacio.

18. Sistema de videoconferencia según la reivindicación 14, en el que cuando hay más de un participante dentro de dicha vista de la cámara, los participantes se combinan en una imagen combinada y el centro de la imagen combinada se utiliza para determinar dichos preajustes.

19. Sistema de videoconferencia según la reivindicación 14, en el que dicha detección comprende:

- proporcionar una imagen digital compuesta por una pluralidad de píxeles (52);

- producir una imagen binaria a partir de la imagen digital detectando píxeles (54) del color de la piel;

- eliminar píxeles correspondientes a bordes en la componente de luminancia de dicha imagen binaria produciendo de ese modo componentes (56) de imagen binaria;

- mapear dichas componentes de imagen binaria en al menos un gráfico (512); y

- clasificar dichas componentes de imagen binaria mapeadas como tipos faciales y no faciales en el que los tipos faciales sirven como candidatos (514) faciales.

20. Sistema de videoconferencia según la reivindicación 15, disponiéndose la al menos una segunda cámara para actualizar dichos preajustes para realizar al menos uno de realizar un movimiento panorámico de dicha cámara por dicho espacio, y hacer zoom en dicha cámara por dicho espacio.

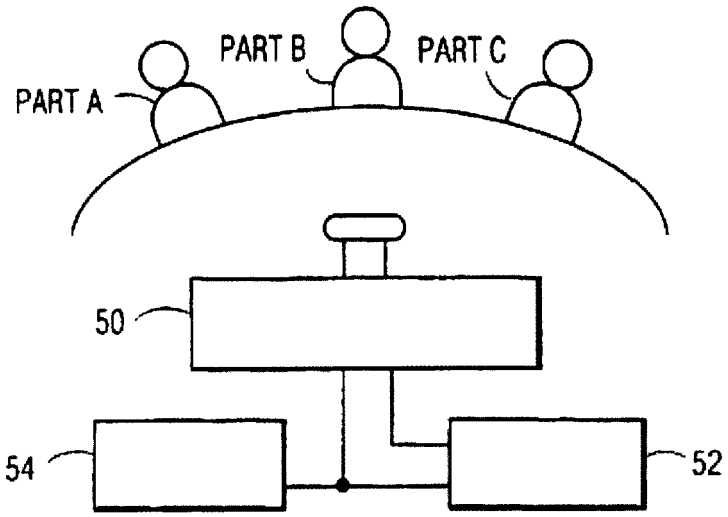
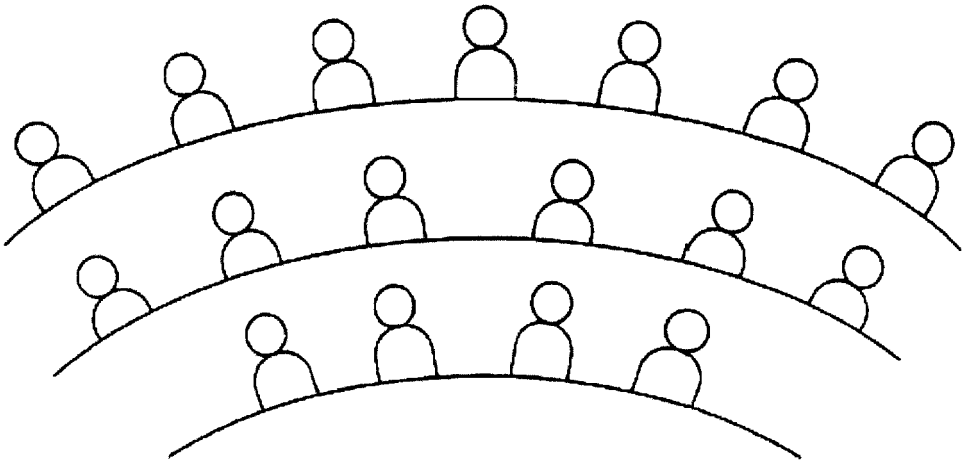


FIG. 1A



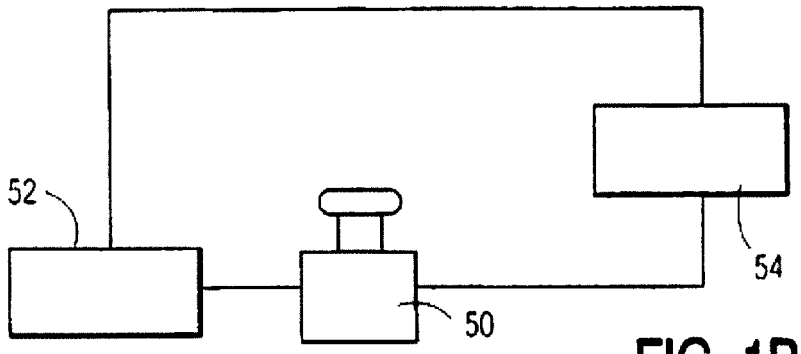


FIG. 1B

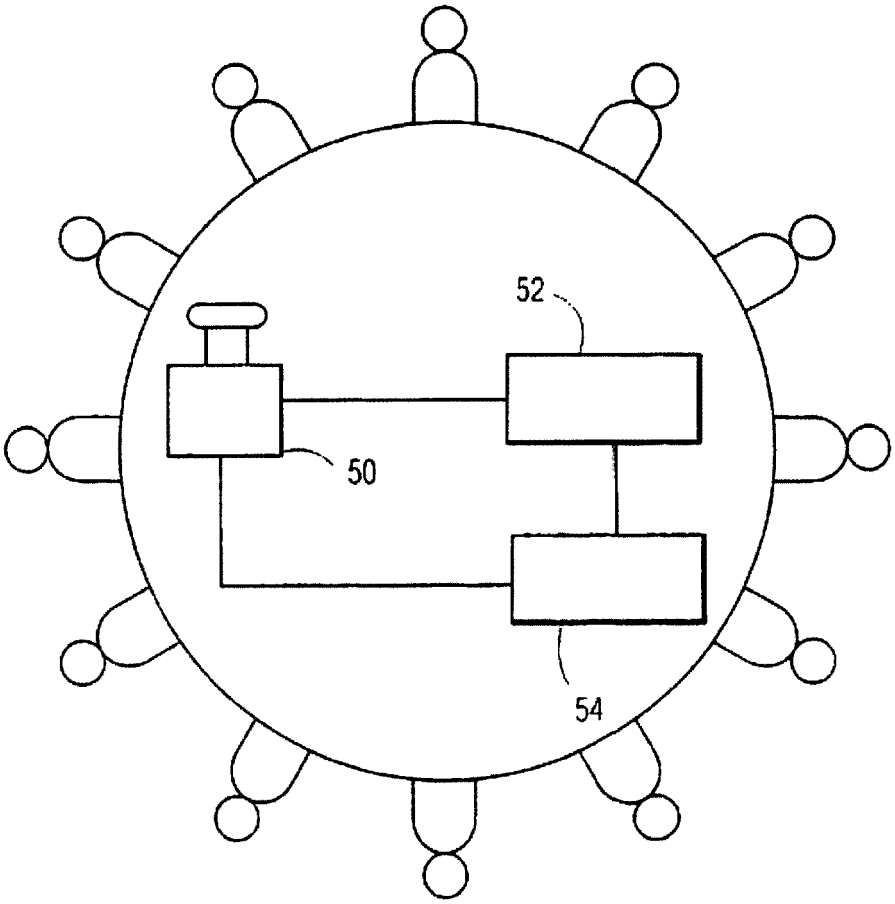


FIG. 1C

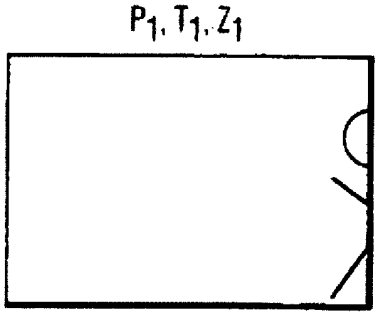


FIG. 2A

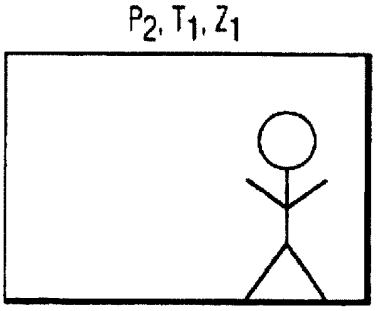


FIG. 2B

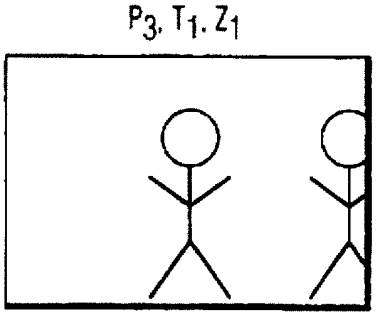


FIG. 2C

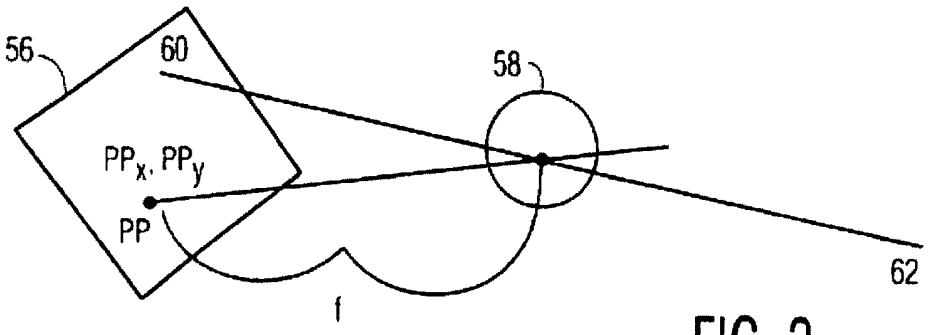


FIG. 3

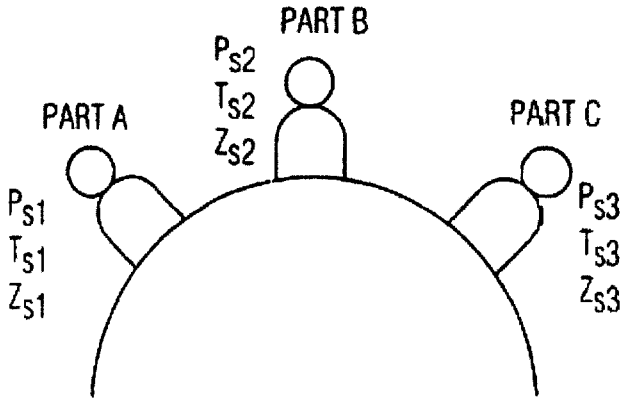


FIG. 4

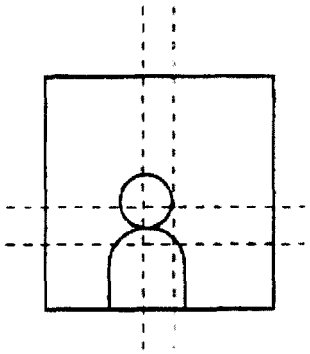


FIG. 5

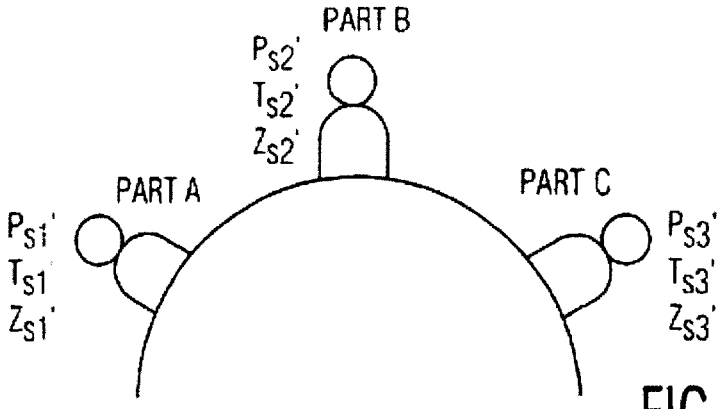


FIG. 6

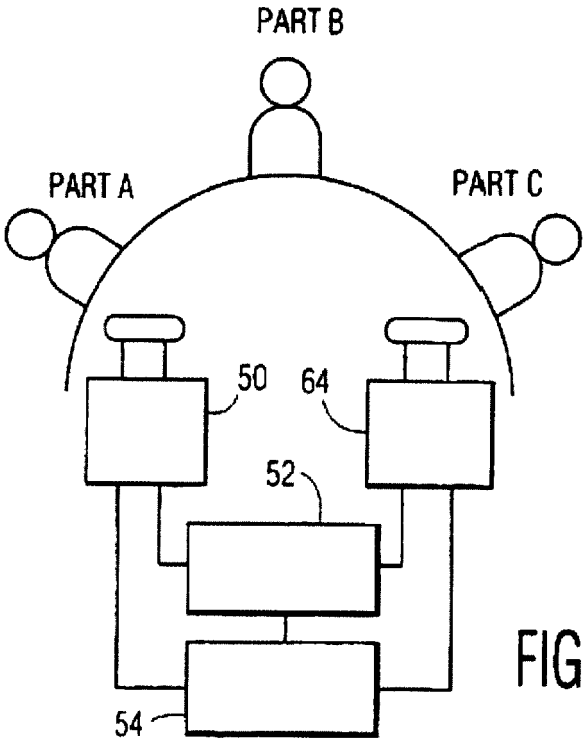


FIG. 7

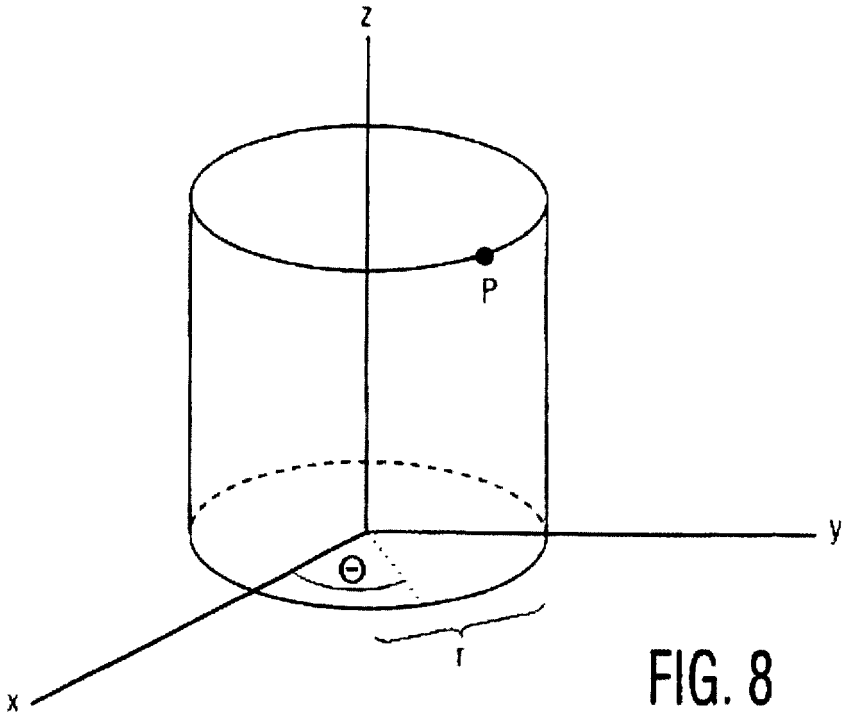


FIG. 8

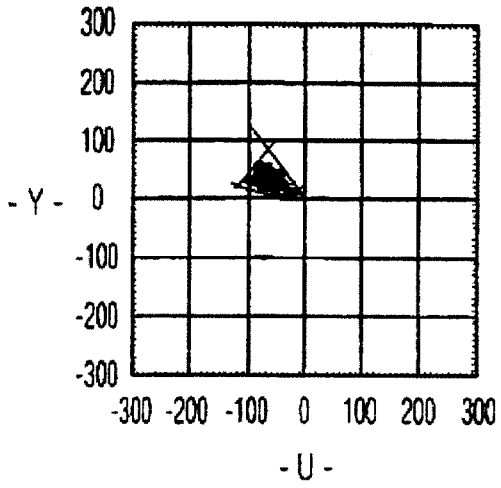


FIG. 9A

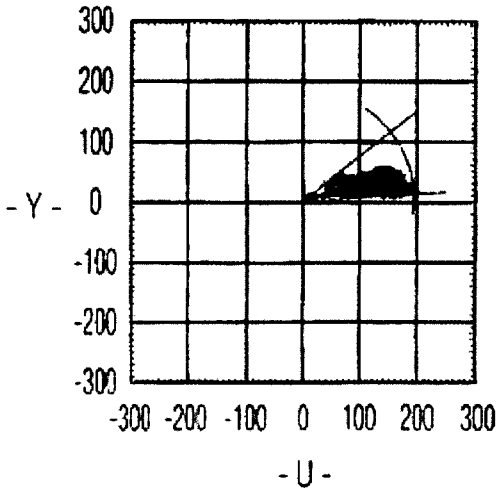


FIG. 9B

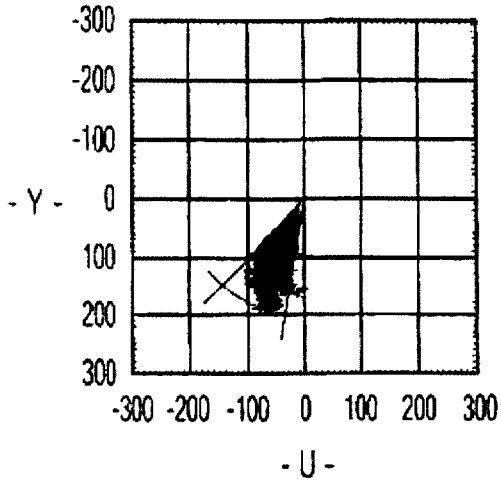


FIG. 9C

FIG. 10A

FIG. 10B

FIG. 10C

FIG. 10D

FIG. 10E

FIG. 10F

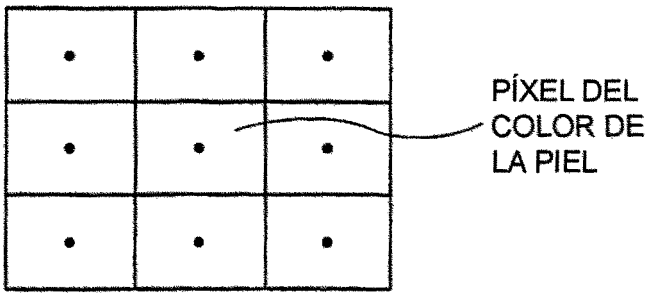


FIG. 11

	1	
1	1	1
	1	

FIG. 12A

1	1	1
1	1	1
1	1	1

FIG. 12B

FIG. 13A

FIG. 13B

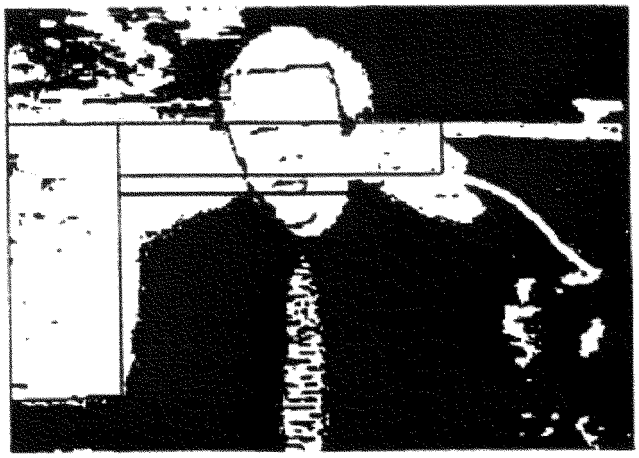


FIG. 14

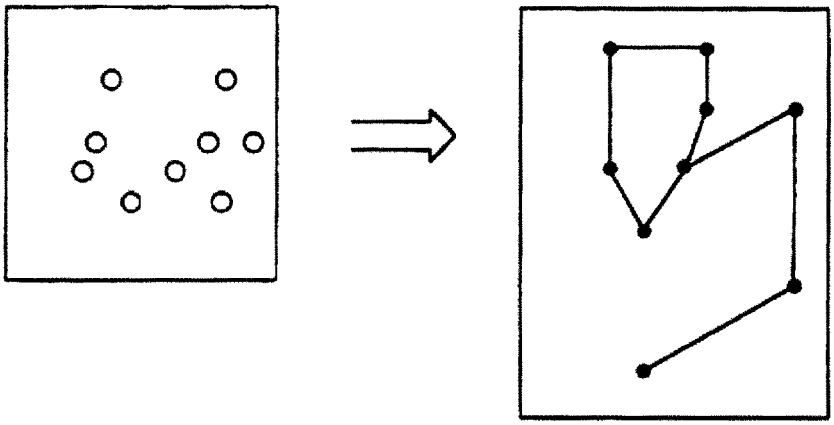


FIG. 15

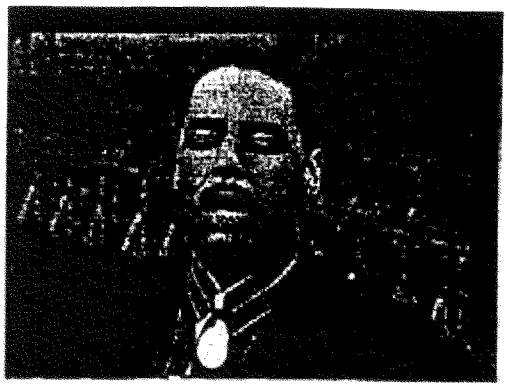


FIG. 16A

FIG. 16B

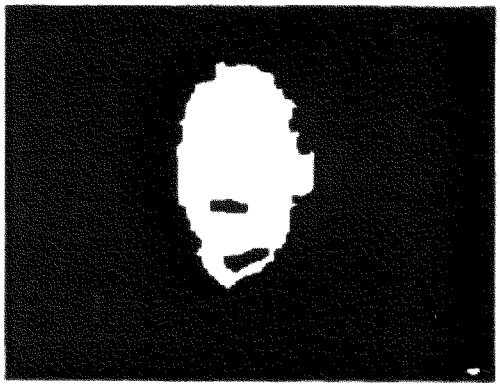


FIG. 16C

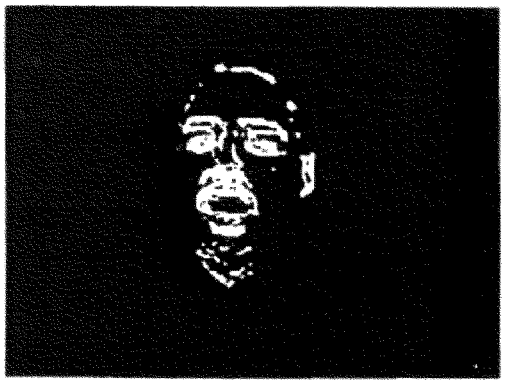


FIG. 16D

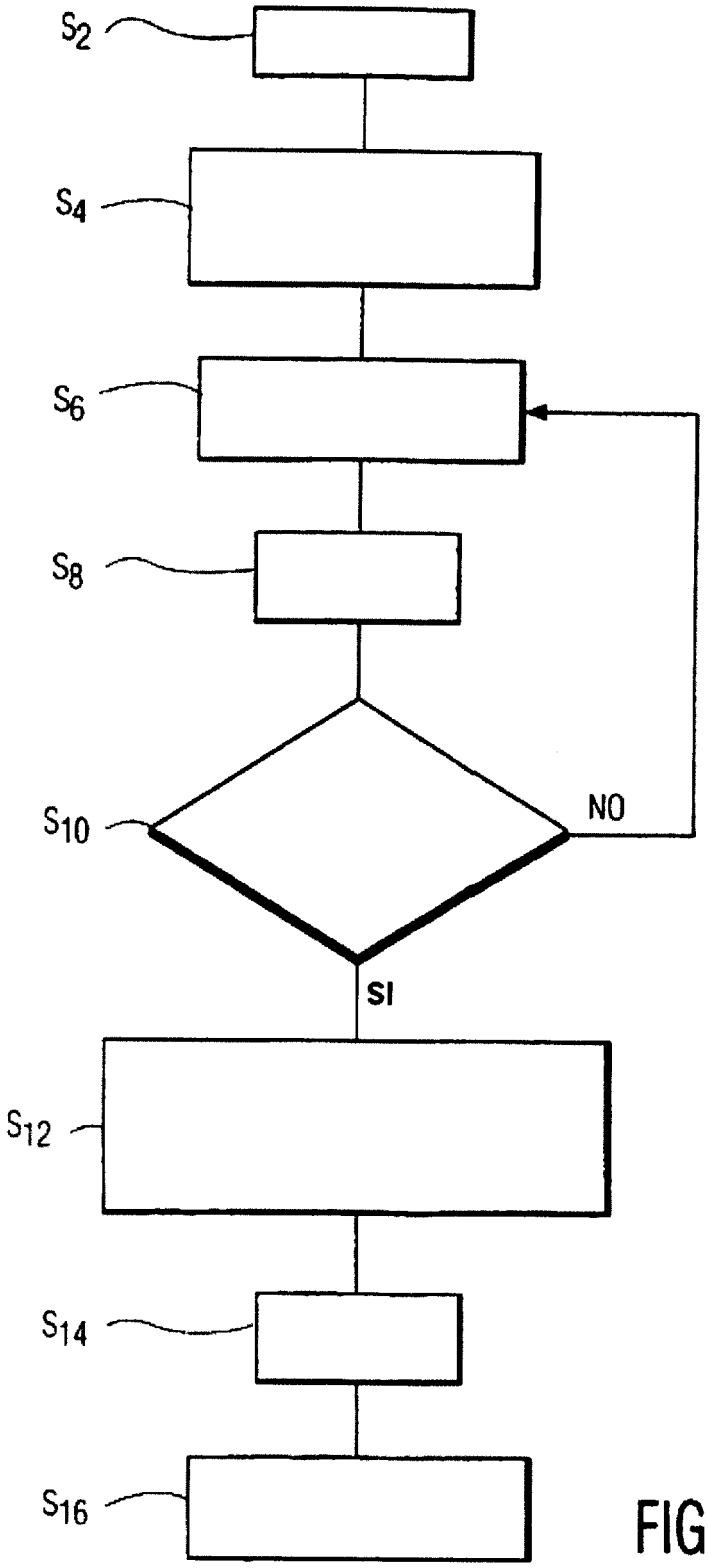


FIG. 17