

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 980 940**

51 Int. Cl.:

G06V 40/16

(2012.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **21.02.2012** **PCT/EP2012/052958**

87 Fecha y número de publicación internacional: **30.08.2012** **WO12113805**

96 Fecha de presentación y número de la solicitud europea: **21.02.2012** **E 12708270 (9)**

97 Fecha y número de publicación de la concesión europea: **10.04.2024** **EP 2678804**

54 Título: **Método para detectar un conjunto predefinido de puntos característicos de un rostro**

30 Prioridad:

22.02.2011 FR 1151424

45 Fecha de publicación y mención en BOPI de la
traducción de la patente:

03.10.2024

73 Titular/es:

FITTINGBOX (100.0%)

**644 Voi l'Occitane, Immeuble Arizona-Bâtiment A
31670 Labège, FR**

72 Inventor/es:

**CHOUKROUN, ARIEL y
LE GALLOU, SYLVAIN**

74 Agente/Representante:

SÁEZ MAESO, Ana

ES 2 980 940 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Método para detectar un conjunto predefinido de puntos característicos de un rostro

- 5 La presente invención se refiere al campo del procesamiento de imágenes. Esta se refiere más particularmente a la detección y al seguimiento de puntos característicos de un rostro.

Contexto de la invención y problemas planteados

- 10 La identificación de puntos característicos de un rostro en una imagen tiene numerosas aplicaciones, entre ellas la seguridad o la realidad aumentada.

- En este último ámbito, podemos citar, por ejemplo, la prueba virtual de gafas totalmente automática, es decir, la superposición en una pantalla de una imagen real del rostro de un usuario y una imagen virtual de unas gafas cuyos parámetros característicos, como la forma y la textura, han sido previamente memorizados.
- 15

- Entonces es importante poder detectar los puntos característicos de un rostro a partir de una simple imagen de un usuario. Este proceso se llama

- 20 alineación facial.

Entre los métodos de alineación, podemos citar los métodos de modelado estadístico. Estos suelen tener dos fases:

- 25 - una fase de aprendizaje que consiste en la creación del modelo estadístico y
- una fase de detección que consiste en hacer converger la forma y/o textura del modelo estadístico sobre datos reales.

- Los tres algoritmos principales que utilizan modelos estadísticos son los modelos de forma activa ASM (Active Shape Models), los modelos de apariencia activa AAM (Active Appearance Models) y los modelos locales restringidos CLM (Constrained Local Models).
- 30

- En la fase de aprendizaje, los ASM solo modelan la forma (conjunto de puntos conectados entre sí) de un objeto, los AAM modelan la forma y la textura (píxeles contenidos dentro de la cubierta convexa de la forma) y los CLM modelan la forma y los parches de textura (píxeles contenidos en una zona de vecindad en cada punto de la forma).
- 35

- En la fase de detección, tanto ASM como CLM convergen la forma del objeto en función de las superficies de respuesta (que son unidimensionales para ASM y bidimensionales para CLM), la textura del modelo estadístico CLM es estática o está bien definida por predicciones. Mientras que para AAM la convergencia se realiza conjuntamente en la textura y la forma (con métodos de descenso de gradiente), la textura del modelo busca aproximarse lo mejor posible a la textura de la imagen sobre la que se realiza la detección.
- 40

- Los CLM tienen la ventaja de ser más resistentes a las ocultaciones y cambios de apariencia, pero son más sensibles a los mínimos locales. Los AAM son más robustos contra los mínimos locales, pero la convergencia de texturas es problemática en los métodos de descenso de gradiente (problema de triangulación de formas cuya función de deformación de la malla siempre debe crear una malla unida), y los AAM se adaptan menos a la variabilidad de los rostros.
- 45

- Los siguientes documentos presentan otros métodos para detectar puntos característicos de un rostro:
- 50

Zihian Xu et al. : "A hierarchical compositional model for face representation and sketching", 2008, ISSN: 1262-8828;

- 55 Zhang Li et al. : "Robust Face Alignment Based on Hierarchical Classifier Network", 2006, ISBN: 978-3-030-71592-2; y

Liang Lin et al. : "Face Alignment Via Component-Based Discriminative Search", 2008, ISBN: 978-3-540-74549-5.

60

Por lo tanto, estos diferentes algoritmos no son completamente satisfactorios.

Divulgación de la invención

- 65 Según un primer aspecto, la presente invención tiene como objetivo un método para detectar un conjunto predefinido de puntos característicos de un rostro a partir de una imagen de este rostro.

El método comprende un grupo de pasos 300-400 que consiste en hacer converger la forma y textura de una jerarquía de modelos estadísticos de partes del rostro sobre los datos reales proporcionados por la imagen del rostro.

5 En la literatura también llamamos reconstrucción facial a este proceso de detección o identificación de puntos característicos, ya que permite reconstruir la forma tridimensional y la posición del rostro y, por lo tanto, calcular varias dimensiones si es necesario, por ejemplo, en el contexto de probarse un par de gafas virtuales.

10 Preferiblemente, los modelos estadísticos utilizados son del tipo Modelos de Apariencia Activa por Parches (PAAM). Ventajosamente, estos se crean a partir de los mismos datos de aprendizaje (texturas y formas), solo que se utilizan diferentes subconjuntos de puntos que definen las formas.

15 Se trata, por tanto, de un algoritmo de reconstrucción facial, denominado en la siguiente Jerarquía de Modelos de Apariencia Activa por Parches jerárquicos ("Hierarchical Patches Active Appearance Models" en inglés o HPAAM).

20 El método comprende el uso de una jerarquía de modelos estadísticos de partes del rostro (HPAAM) para la detección de puntos que caracterizan rostros simplificados o partes del rostro (tales como, por ejemplo, los ojos y la boca) (paso 300) y luego una jerarquía de modelos estadísticos del rostro para la detección final (paso 400).

Preferiblemente, el grupo de etapas 300-400 utiliza un método de alineación de modelado que comprende dos fases:

25 - una fase de aprendizaje que consiste en la creación de un modelo estadístico y
- una fase de detección que consiste en hacer converger la forma y/o textura del modelo estadístico sobre datos reales.

30 Según una implementación preferida, la construcción de un modelo estadístico de forma y textura se realiza alineando y normalizando todas las formas y texturas aplicando luego un análisis de componentes principales sobre las formas alineadas y sobre las texturas alineadas.

35 El método de detección utiliza, en un caso particular, una etapa de minimización mediante un algoritmo compositivo inverso (ICA) para la reconstrucción del rostro haciendo converger parámetros de forma p , de pose q y de textura λ de cada modelo estadístico de parte del rostro.

40 En esta fase de detección utilizando un enfoque tipo Gauss Newton mediante el algoritmo ICA, se optimiza la textura, forma y pose del modelo PAAM para que la textura de cada uno de los parches converja hacia la textura de la imagen utilizada para la detección.

El modelo jerárquico es un conjunto compuesto por un modelo estadístico de partes del rostro (MSV1) y un modelo estadístico del rostro (MSV2).

Según diversas implementaciones posiblemente utilizadas conjuntamente:

45 - el método incluye una etapa 100 de preprocesamiento de la imagen para superar mejor las condiciones de iluminación de la captura, incluyendo este preprocesamiento la ecualización del histograma.
- el método comprende una etapa 200 de detección del rostro en la imagen, realizada utilizando un algoritmo del tipo de Viola y Jones.

Presentación de las figuras

Las características y ventajas de la invención se apreciarán mejor gracias a la descripción que sigue, descripción que recoge las características de la invención a través de un ejemplo de aplicación no limitativo.

La descripción se basa en las figuras adjuntas que representan:

Figura 1: un diagrama de flujo del método como se describe aquí
Figura 2: un ejemplo de un modelo estadístico de rostro.

Descripción detallada de una realización de la invención

En el contexto del ejemplo aquí descrito, el proceso según la invención (ilustrado Figura 1) está destinado a ser implementado en forma de software, por ejemplo, mediante un microordenador disponible comercialmente; conocido en sí mismo.

El método funciona con datos de entrada representativos de una imagen del rostro de un individuo.

El método y el dispositivo para adquirir la imagen del rostro como tal quedan fuera del alcance de la presente invención y, por lo tanto, no se detallan más aquí.

5

Modo de funcionamiento

El algoritmo de la solución de alineación de rostros propuesta consta de cuatro etapas:

10 Etapa 100: Preprocesamiento de la imagen para superar mejor las condiciones lumínicas de la captura. El preprocesamiento de imágenes utilizado es la ecualización de histograma.

15 Etapa 200: Detección de rostros en la imagen. La detección de rostros se realiza mediante un algoritmo ya conocido, del tipo de Viola y Jones

20 Etapa 300: Detección de partes del rostro (como los ojos y la boca) en la zona del rostro detectada en la etapa 200 anterior. La detección de partes del rostro consiste en hacer converger un modelo estadístico de rostro simplificado (MSV1) del tipo modelo de apariencia activa por parches (denominado PAAM por simplificación).

25 Cabe señalar que esta etapa 300 puede contener subetapas intermedias que no son esenciales, pero que facilitan la reconstrucción (es decir, la detección de los puntos característicos del rostro, permitiendo reconstruirla bajo cualquier ángulo) más robusta. Esta consiste en hacer converger en cada subetapa un modelo estadístico de tipo PAAM cada vez más rico o complejo en número de puntos y/o tamaño de parches. La última subetapa es la convergencia del modelo estadístico simplificado más complejo.

El objetivo de cada subetapa es hacer converger un algoritmo de tipo ICA (algoritmo composicional inverso), conocido per se, en un modelo estadístico antes que otro modelo estadístico más complejo.

30 Este algoritmo ICA permite encontrar iterativamente los parámetros de forma (p), pose (q) y textura (λ) que mejor se corresponden con la imagen procesada.

En efecto, esto permite, en particular, mejorar la estimación del parámetro de pose q , que es bastante impreciso utilizando la detección de rostros de Viola y Jones.

35

También cabe señalar que, en una variante de implementación, se puede considerar una inicialización más compleja de los modelos PAAM. Varias instancias o inicializaciones del modelo estadístico PAAM (del rostro simplificado MSV1 o de una subetapa) pueden iniciarse, por ejemplo, dentro del área del rostro encontrada en la etapa 200. Los puntos de las partes del rostro detectadas serán entonces los dados por el mejor ejemplo en términos de correlación o error de convergencia.

40

Etapa 400: Detección de puntos característicos del rostro. La detección de puntos característicos del rostro consiste en hacer converger un modelo estadístico de rostro (MSV2) del tipo PAAM. La Figura 2 ilustra un ejemplo de un modelo estadístico de rostro.

45

Los modelos estadísticos MSV1 y MSV2 se crean a partir de la misma base de aprendizaje, luego existen factores multiplicativos que permiten hacer coincidir los parámetros que rigen las formas correspondientes a los modelos estadísticos MSV1 y MSV2. Por ejemplo, el parámetro que modifica la forma de un movimiento del rostro de izquierda a derecha para el modelo estadístico MSV1 es -1 veces el del modelo estadístico MSV2.

50

La inicialización de la etapa 400 se realiza utilizando estos factores multiplicativos aplicados a los parámetros encontrados en la etapa 300. Los otros parámetros que no encuentran una coincidencia se pueden instanciar en varias posiciones (-0,5, 0 y 0,5 por ejemplo), y sólo la convergencia produce el error más pequeño. Las pruebas mostraron que era mejor considerar el error dado por la correlación de la textura del modelo y la textura de la imagen (ZNCC) en lugar de la diferencia al cuadrado de estas texturas (SSD).

55

El parámetro de pose (q) se inicializa usando la etapa 300.

Cabe señalar que esta etapa 400 también puede contener subetapas intermedias que no son esenciales, pero que hacen que la reconstrucción sea más robusta. Esta consiste en hacer converger en cada subetapa un modelo estadístico de rostros tipo PAAM cada vez más complejo en términos de tamaño de parche o resolución. La última subetapa es la convergencia del modelo estadístico de rostro más complejo.

60

Cabe destacar también que los parámetros de inicialización de las subetapas se pueden deducir de la subetapa anterior gracias a que se pueden crear los modelos estadísticos de rostros a partir de la misma base de aprendizaje.

65

Con respecto a los modelos estadísticos PAAM (utilizados en las etapas 300 y 400), el método de alineación por modelado estadístico tiene dos fases:

- una fase de aprendizaje que consiste en la creación del modelo estadístico y
- una fase de detección que consiste en hacer converger la forma y/o textura del modelo estadístico sobre datos reales.

La fase de aprendizaje del modelo estadístico PAAM por parches con apariencia activa (Patches Active Appearance Models), propuesta en el presente ejemplo de implementación de la invención, se diferencia de un algoritmo de modelos locales restringidos CLM en el hecho de que ahora se aprende una textura y cómo realizar la alineación de datos.

La fase de detección del modelo PAAM se realiza de forma similar a la de los AAM. Utilizamos una minimización de Gauss Newton conjuntamente sobre la textura, la forma y la pose del modelo, de modo que la textura de cada uno de los parches converja hacia la textura de la imagen utilizada para la detección. Una diferencia muy significativa proviene de la forma de modificar la forma (la forma de implementar las funciones de deformación o funciones 'Warp'). De hecho, los métodos de convergencia AAM requieren la composición de funciones de deformación de malla. Pero en el caso de las mallas, la composición en cada triángulo hace que se pierda la conectividad de la malla. Una malla es un conjunto de puntos conectados por triangulación.

La ventaja de este enfoque PAAM es que la resolución de la minimización no requiere aproximación debido a las funciones de deformación (o 'Warp') ya que ya no existe ninguna noción de conectividad de malla y podemos formalizar la solución del problema de forma matricial.

En una implementación preferida, en el presente método de alineación de caras se utiliza un modelo jerárquico (denominado HPAAM para simplificar en el resto de la descripción) que es la composición de modelos estadísticos PAAM, pudiendo crearse estos modelos a partir de los mismos datos de aprendizaje (texturas y formas).

Un modelo jerárquico se define como la composición de modelos estadísticos cuando estos modelos estadísticos se crean a partir de los mismos datos de entrenamiento (texturas y formas). Sólo se utilizan diferentes subconjuntos de puntos que definen la forma.

Se llama jerárquico porque la definición (en número de puntos) y por tanto el contenido semántico de los modelos involucrados aumenta gradualmente.

La construcción de un modelo estadístico de forma y textura se lleva a cabo, por su parte, alineando y normalizando todas las formas y texturas y luego aplicando un análisis en componente principal (observado por la simplificación ACP en el resto de la descripción) sobre las formas alineadas y las texturas alineadas. Los algoritmos de análisis de componentes principales son bien conocidos por los expertos en la técnica y, por lo tanto, no se detallan aquí.

Las formas (S) se parametrizan entonces de la siguiente manera: $S = \bar{S} + \Phi_S b_S$, en la que S denota la forma promedio y b_S los vectores propios dados por el análisis de componentes principales. De manera similar, para las texturas: $T = \bar{T} + \Phi_T b_T$

Los modelos estadísticos presentados, en este ejemplo, se crean a partir de alrededor de un centenar de texturas y formas de rostros.

Una textura es la concatenación de parches rectangulares centrados en cada punto de la forma.

El modelo jerárquico propuesto es un conjunto compuesto por un modelo estadístico de partes del rostro (MSV1) y un modelo estadístico del rostro (MSV2).

A continuación, se procede a una alineación de las formas mediante un análisis de Procrustes, según un método ya conocido.

La forma del modelo estadístico del rostro MSV2 se compone, en este ejemplo de implementación, de cincuenta y un puntos. Sin embargo, está claro que se pueden utilizar modelos que utilicen mayor o menor número de puntos.

También en este caso realizamos una alineación de las formas mediante un análisis de Procrustes, según un método ya conocido. La textura es la concatenación de los 51 parches rectangulares.

En la solución de alineación de rostros presentada aquí como ejemplo no limitante, se usa un HPAAM para la detección de partes del rostro (etapa 300) y luego un HPAAM para el rostro para la detección final (etapa 400).

Un modelo estadístico PAAM utiliza dos parámetros estandarizados para estimar: un primer parámetro p que define la forma y un segundo parámetro λ que define la textura. Para la variabilidad en la pose de las caras en la imagen (rotación, homotecia, traslación) un tercer parámetro q (parámetro de pose) debe ser estimado.

- 5 Para la parte HPAAM de las partes del rostro (paso 300), los parámetros p y λ del modelo estadístico PAAM MSV1 se inicializan a 0, mientras que el parámetro de pose q se estima usando la detección de rostros.

Asimismo, en el caso de que se utilicen subetapas intermedias, los parámetros p y λ de los siguientes modelos estadísticos PAAM (modelos más complejos) se inicializan a 0.

- 10 Cabe señalar que en una realización no preferida debido a que es menos robusta, se puede eliminar la etapa 300 de convergencia de un modelo estadístico PAAM de una cara simplificada MSV1. En este caso, luego se inician varias instancias o inicializaciones del parámetro de forma de la etapa 400. Los puntos faciales detectados serán entonces los dados por el mejor ejemplo en términos de error de correlación o convergencia.

- 15 El parámetro de textura λ se inicializa a 0 ya que las texturas de las dos etapas (correspondientes a los dos niveles de resolución) no están vinculadas. Finalmente, el parámetro de pose q se puede tomar directamente de la etapa PAAM anterior.

- 20 Según un segundo aspecto, la presente invención tiene como objetivo un método para rastrear un conjunto predefinido de puntos característicos de un rostro en un vídeo. En efecto, el método de alineación de rostros descrito anteriormente permite detectar puntos característicos de un rostro que se encuentran en la primera imagen que compone un vídeo. Esta etapa permite, en particular, encontrar los parámetros de pose, textura y forma que explican el rostro detectado. Para las siguientes imágenes que componen el vídeo, dicha detección
25 no es necesaria ya que siempre se trata de la misma identidad y las variaciones de forma y pose entre 2 imágenes consecutivas son mínimas. El seguimiento consiste, por tanto, en hacer converger un modelo estadístico del rostro (MSV2) de tipo PAAM sucesivamente para cada imagen. Los parámetros simplemente se inicializan utilizando los parámetros encontrados en la imagen anterior.

- 30 Para cada punto asociado al parche, es posible establecer correspondencias directas entre las imágenes y permitir así la reconstrucción 3D dispersa del modelo de rostro, así como su seguimiento 3D a lo largo de la secuencia de vídeo utilizando las técnicas clásicas de ajuste de haz conocidas por los expertos en la técnica.

Ventajas de la realización

- 35 Un algoritmo de alineación de rostro HPAAM permite detectar los puntos característicos de un rostro liberándose al máximo de las condiciones de captura de la imagen (inducidas por el sistema de adquisición o por la iluminación de la escena) y de las variaciones específicas del rostro (identidad, pose, expresión).

- 40 Cabe señalar que la elección de un modelo deformable por parche tiene la ventaja de crear menos dependencias de la iluminación, menos dependencias de las expresiones faciales y de evitar el problema de conectividad de la malla en comparación con un modelo completamente deformable.

- 45 La reconstrucción con parches propuesta ha demostrado su eficacia en bases de datos de rostros que contienen una gran variabilidad en identidades, poses e iluminaciones.

Variantes de realización

- 50 El alcance de la presente invención no se limita a los detalles de las realizaciones anteriores consideradas a modo de ejemplo, sino que por el contrario se extiende a modificaciones que están al alcance de los expertos en la técnica. El alcance de la protección lo proporcionan las reivindicaciones adjuntas.

REIVINDICACIONES

1. Método para detectar un conjunto predefinido de puntos característicos de un rostro a partir de una imagen de este rostro,
 5 en el que los puntos característicos se obtienen haciendo converger la forma y textura de una jerarquía de modelos estadísticos de partes del rostro, MSV1, y de una jerarquía de modelos estadísticos de rostro, MSV2, sobre los datos reales proporcionados por la imagen del rostro, los modelos estadísticos utilizados son del tipo Modelos de Apariencia Activa por Parches, PAAM,
 10 caracterizado porque comprende el uso de la jerarquía de modelos estadísticos de partes del rostro, MSV1, para la detección de partes del rostro en el área del rostro en la imagen (etapa 300), luego el uso de la jerarquía de modelos estadísticos de rostro, MSV2, para la detección final (paso 400) en el que las jerarquías de los modelos estadísticos MSV1 y MSV2 se crean a partir de la misma base de aprendizaje, factores multiplicativos correspondientes a parámetros que rigen las formas correspondientes a las jerarquías de modelos estadísticos MSV1 y MSV2, la inicialización de la detección final que se realiza utilizando dichos factores multiplicativos aplicados a los parámetros encontrados en la detección de las partes del rostro en el área del rostro en la imagen.
2. Método según la reivindicación 1, caracterizado porque la detección final comprende una pluralidad de subetapas intermedias, convergiendo en cada subetapa un modelo estadístico de rostro cada vez más complejo en términos de tamaño de parche o resolución.
3. Método según la reivindicación 2, caracterizado porque se considera un error dado por la correlación de la textura del modelo y la textura de la imagen para la convergencia de la jerarquía de modelos estadísticos de rostros para la detección final.
 25
4. Método según una cualquiera de las reivindicaciones anteriores, caracterizado porque la detección de partes del rostro comprende una pluralidad de subetapas intermedias, convergiendo en cada subetapa un modelo estadístico de tipo PAAM cada vez más rico en número de puntos y/o en tamaño de parches.
 30
5. Método según la reivindicación 4, caracterizado porque se utiliza un algoritmo compositivo inverso, ICA, para la convergencia del modelo estadístico durante cada subetapa de la detección de partes del rostro.
6. Método según una cualquiera de las reivindicaciones anteriores, caracterizado porque la detección de las partes del rostro y la detección final utilizan un método de alineación por modelado que comprende dos fases:
 35
 - una fase de aprendizaje que consiste en la creación de un modelo estadístico y
 - una fase de detección que consiste en hacer converger la forma y/o la textura del modelo estadístico sobre datos reales.
 40
7. Método según la reivindicación 6, caracterizado porque la construcción de un modelo estadístico de forma y textura se realiza alineando y normalizando todas las formas y texturas aplicando luego un análisis de componentes principales sobre las formas alineadas y sobre las texturas alineadas.
 45
8. Método según cualquiera de las reivindicaciones anteriores, caracterizado porque los modelos estadísticos se crean a partir de los mismos datos de aprendizaje, utilizándose únicamente diferentes subconjuntos de puntos que definen las formas.
9. Método según cualquiera de las reivindicaciones anteriores, caracterizado porque comprende una etapa de minimización mediante un algoritmo compositivo inverso, ICA, para la detección de puntos característicos del rostro haciendo converger parámetros de forma p , de pose q y de textura λ de cada modelo estadístico de parte de rostro.
 50
10. Método según cualquiera de las reivindicaciones 6 a 8, caracterizado porque en la fase de detección del modelo PAAM se utiliza conjuntamente una minimización de Gauss Newton sobre la textura, la forma y la pose del modelo, de modo que la textura de cada uno de los parches converge hacia la textura de la imagen utilizada para la detección.
 55
11. Método según cualquiera de las reivindicaciones anteriores, caracterizado porque comprende un pretratamiento de la imagen para superar mejor las condiciones de iluminación de la captura, comprendiendo este pretratamiento una ecualización del histograma.
 60
12. Método según cualquiera de las reivindicaciones anteriores, caracterizado porque comprende una detección del rostro en la imagen, realizada mediante un algoritmo del tipo de Viola y Jones.
 65
13. Método de seguimiento de un conjunto predefinido de puntos característicos de un rostro en un vídeo,

5

caracterizado porque dicho método comprende una detección de puntos característicos del rostro para la primera imagen que compone el vídeo, y que consiste en hacer converger la forma y textura de un modelo estadístico de partes del rostro sobre los datos reales proporcionados por las imágenes siguientes que componen el vídeo, siendo el modelo estadístico utilizado del tipo Modelo de Apariencia Activa por Parches, PAAM, implementando dicho método de seguimiento el método de detección de un conjunto predefinido de puntos según cualquiera de las reivindicaciones 1 a 12.

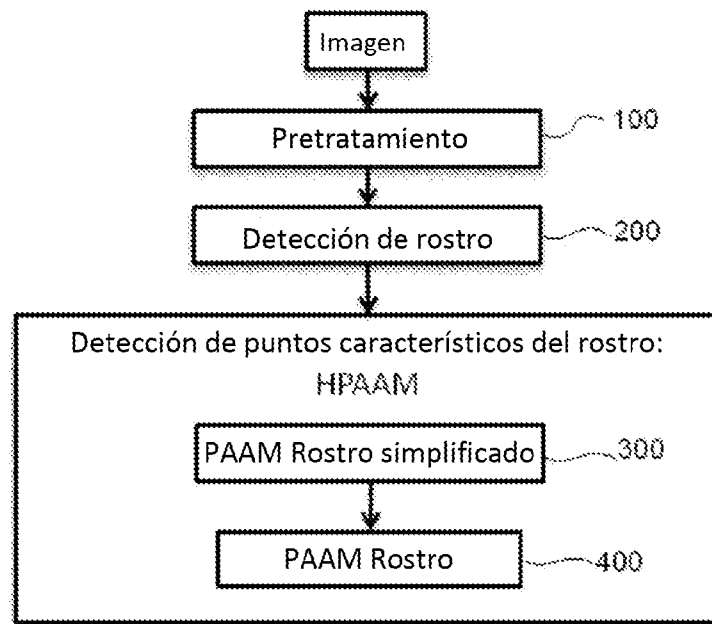


Figura 1

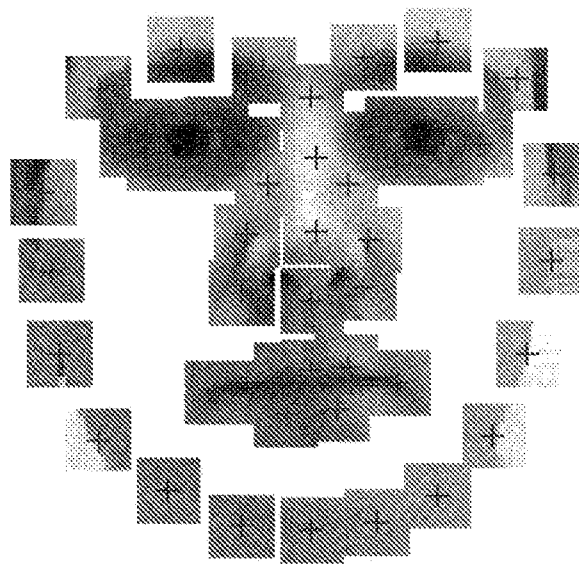


Figura 2