

Algoritmos de detección y seguimiento de rostro dirigidos a videojuegos

Juan Camilo Gómez Cuervo ^{1*}, Daniel Alberto Alarcón Velandia^{2**}

Resumen

Actualmente el desarrollo de aplicaciones de reconocimiento facial está en auge, hay un sin fin de estas en diversas áreas, pero en Colombia no se han desarrollado aplicaciones basadas en el movimiento facial encaminadas al desarrollo de videojuegos.

La no evolución de interfaces tradicionales como los mandos de videojuego (joysticks, teclado y mouse) que forman parte de la corriente principal del juego, hacen que existan problemas para el jugador, como: ergonomía, tamaño, batería, dureza de botones y cables. Lo más cercano a una evolución ha estado presente en cambios tales como: el número de los botones, aspectos de usabilidad y personalización.

La razón principal de apuntar a los videojuegos aprovechando la identificación y el seguimiento de personas en una imagen digital, es poder ofrecer una forma innovadora de jugabilidad e interacción en videojuegos, a través de la incorporación de un algoritmo de reconocimiento facial que pueda servir como mando para controlar acciones.

Por tanto, este trabajo se basó en analizar y probar varios algoritmos de detección y reconocimiento de código libre, justificando cual es el más apropiado para esta forma de interacción en entornos virtuales de videojuegos.

Abstract

Currently the development of facial recognition applications is booming, there is an endless number of these in various areas, but in Colombia have not been developed based on facial movement aimed at developing gaming applications.

Non evolution of traditional interfaces and controls video game (joysticks, keyboard and mouse) that are part of the mainstream of the game, cause there are problems for the player, such as ergonomics, size, battery, hard buttons and cables. The closest thing to an evolution has been in changes such as the number of buttons, aspects of usability and customization.

The main reason for targeting video games taking advantage of the identification and tracking of people in a digital image, is to offer an innovative way of gameplay and interaction in gaming, through the incorporation of a face recognition algorithm that can serve as command to control actions.

Therefore, this work was based on analysis and test various detection and recognition algorithms open source, justifying which is most appropriate for this form of interaction in virtual gaming environments.

Palabras Claves

Reconocimiento facial, videojuegos

¹ Programa de Ingeniería de Sistemas, Universidad Autónoma de Colombia, Bogotá, Colombia

² Programa de Ingeniería de Sistemas, Universidad Autónoma de Colombia, Bogotá, Colombia

*Contacto: jgomez.cuervo@fuac.edu.co

**Contacto: daniel.alarcon@fuac.edu.co

Índice

Introducción	1
1 Estado del arte	2
1.1 Detección facial	2
1.2 Técnicas	2
2 Algoritmos de reconocimiento facial	2
3 Metodología	3
4 Análisis algoritmos de reconocimiento	3
5 Pruebas	4
6 Resultados	5
Referencias	5

Introducción

La industria de los videojuegos va creciendo y variando cada vez más debido al avance acelerado de la tecnología, permitiendo así que se desarrollen videojuegos con una mayor inmersión. Esta evolución se ha visto en aspectos como: la animación, el sonido y la calidad gráfica. Pero respecto a los mandos de videojuegos (gamepad, teclado y mouse, entre otros), los cuales funcionan con dispositivos con botones o teclas, y que su fin es el de controlar los movimientos y acciones dentro del juego, no se ha visto una evolución hacia este tipo de interfaces, lo más cercano ha sido números de botones, peso y aspectos de usabilidad

El tener un mando no convencional sería un paso a la solución de inconvenientes que generan incomodidad al jugador, como problemas ergonómicos que producen lesiones en mano,

muñeca y antebrazo causando dolor cuando el movimiento es de larga duración y repetitivo, también existen problemas de tamaño que generan molestias a jugadores con dedos grandes, de baterías donde su vida útil es corta, dureza en botones que generan ampollas y dolor.

Si bien hoy en día existen mandos basados en biometría (PlayStation Move, Kinect), el tener un mando no tradicional menos costosa y alternativa a estos mandos los cuales realizan el seguimiento del cuerpo, pero tienen un precio elevado en el mercado, además un algoritmo de detección de rostro, se aplica con el fin de realizar el seguimiento al movimiento del rostro en base a la posición. Este trabajo se basó en analizar y probar varios algoritmos de detección y reconocimiento de código libre, justificando cual es el más apropiado para esta forma de interacción en entornos virtuales de videojuegos.

El proceso que se llevó a cabo fue probar a través de una serie de indicadores (tiempo de ejecución, estabilidad y rangos de detección) los algoritmos: Camshift (Continuously Adaptive Mean), Viola-Jones y Kanade-lucas y como resultado se condujo a preferir a camshift, ya que arrojó los mejores resultados convirtiéndolo en la mejor opción para este proyecto.

Actualmente y basándonos en el resultados de este trabajo se está desarrollando un prototipo de videojuego que será controlado por medio del algoritmo seleccionado.

1. Estado del arte

A continuación, se presenta el marco conceptual de la temática de este artículo.

1.1 Detección facial

En la computación, es el proceso donde el computador ubica los rostros presentes en una imagen digital o en un video, este reconocimiento es más veloz que la echa por el sistema visual humano[1]. Desde la perspectiva del computador, una imagen no es más que solo datos agrupados que se interpretan como información que contienen valores de color; por ejemplo, si se tiene la fotografía de algún familiar la computadora por lógicas razones no reaccionara de la misma forma que una persona, pero la computadora se puede entrenar de tal manera que esta pueda reconocer patrones de forma y de color, en este caso reconocer un rostro, a este proceso se le conoce como detección facial[2].

1.2 Técnicas

Se agrupan en:

- *Métodos basados en conocimiento:* Se basan usar las reglas que usan las personas para identificar en una imagen un rostro, dichas reglas se obtienen del conocimiento previo que se tenga del rostro la persona y describen las características y las relaciones que se tengan de un rostro. En un algoritmo esta técnica estaría dada por extraer características faciales en una imagen de entrada dadas por la distancia o posición entre componentes del rostro, como la simetría entre los ojos, la

distancia entre la nariz y la boca, el tamaño de los ojos, entre otros[3].

- *Métodos basados en características invariantes:* Basados en características invariantes según la situación y de bajo nivel, entre estos están:
 - Bordes: Como lo dice su nombre es extraer todas aquellas líneas externas o internas que limitan la composición del rostro, ya hay algoritmos basados en métodos como : el operador Canny, gradiente de Sobel y derivadas de funciones gaussianas[4].
 - Descriptores de textura: Las texturas del rostro son piel y cabello, este proceso se realiza por medio de redes neuronales[4].
 - Color: Las imágenes a color brindan mayor información que una imagen a blanco y negro como el detectar el color de la piel, su función principal es ayudar a determinar la raza de una persona[4].
- *Métodos basados en Moldes:* Se basan en aplicar una medida de distancia entre la imagen de entrada y el patrón a usar con el fin de modelar y agrupar características del rostro y establecer una relación entre regiones particulares que se encuentren en la imagen y poder reconocer el rostro por dichas regiones[3].
 - Patrones predefinidos: Divide el rostro en bloques donde cada uno de estos tiene información acerca de la luminosidad y nivel de bordes[1].
- *Métodos Basados en la Apariencia:* No necesitan conocimiento previo de la característica a ser detectada, acá se mencionan técnicas de clasificación como eigenfaces la cual necesita varias imágenes para instruirse y agrupar lo necesario para realizar la detección[1].

2. Algoritmos de reconocimiento facial

Hay variedad de algoritmos de reconocimiento y seguimiento facial, el problema que radica en algunos como: transformada de hough, cany, pca, AdaBoost, Haar-like features, SMQT Features, Elastic bunch graph matching etc. es que son algoritmos que hacen parte del reconocimiento de rostros pero que tienen un propósito distinto como detectar bordes o figuras geométricas, en cambio camshift, viola jones y kanade-lucas son algoritmos enfocados reconocimiento de rostros.

- **Viola-Jones:** La técnica de detección propuesta por Paul Viola y Michael Jones utiliza una serie de clasificadores, agrupados consecutivamente con características previamente definidas, con el fin de detectar el rostro en una imagen[5].

La metodología de este algoritmo es:

Imagen Integral: La imagen integral se define como una matriz de búsqueda de dos dimensiones, con el mismo tamaño que la imagen de entrada donde cada elemento

de la imagen integral contiene la suma de los píxeles que se encuentren en la región superior izquierda, para luego la imagen integral se represente como la integral doble de una imagen; primero a lo largo de las filas y después a lo largo de las columnas dividiéndola en cuatro búsquedas dentro de esa región[6].

La ecuación para hallar la imagen integral es:

$$ii(x,y) = \sum_{x' \leq x, y' \leq y} i(x',y') \quad (1)$$

De la ecuación (1) representan $ii(x,y)$ a la imagen integral y $i(x,y)$ es el valor de la imagen en unas coordenadas específicas. La sumatoria son los valores de píxeles en la imagen integral de las esquinas del rectángulo en una imagen de entrada[6].

Extracciones de características: En imágenes las características de cada objeto se extraen al aplicar ciertas funciones que permitan la representación y descripción de los objetos de interés de la imagen (patrones). La extracción de características es un paso en el reconocimiento de patrones en el cual las medidas u observaciones son procesadas para encontrar a tributos que puedan ser usados para asignar los objetos a determinada clase[7].

Clasificación: Los clasificadores son algoritmos preparados para aprender la distribución de datos a partir de entrenamientos con el fin de predecir el tipo al que pertenecen[8].

- **CAMSHIFT:** Este algoritmo es una adaptación del algoritmo Mean Shift, se basa en la probabilidad dinámica, que son cambios de tamaño y posición en objetos que se encuentren en movimiento. La imagen, que viene del modelo RGB es cambiada al modelo HSV (hue, saturation, value) con el fin de descartar los píxeles de bajo valor de saturación, muy oscuros o muy claros[9].

La metodología de este algoritmo es :

Inicialización: realiza un pre procesamiento de la imagen, el usuario selecciona la región de interés y allí se inicia búsqueda del objeto a seguir, luego se convierte de RGB a HSV para que la captura de los componentes del objeto se realice en forma análoga, se obtiene luego el histograma para realizar la detección del color el cálculo del histograma que se realiza mediante el canal hue del espacio HSV para reducir los tiempos del cálculo de ejecución del algoritmo[10].

Cálculo, comparación y actualización del histograma:

$$d(H_{n+1}, H_n) = \sqrt{1 - \frac{\sum_i^N \sqrt{H_n(i)H_{n+1}(i)}}{\sum_i^N H_{n+1}(i) \sum_i^N H_n(i)}} \quad (2)$$

La ecuación (2) H_n y H_{n+1} son el histograma anterior y actual, cada uno de estos tiene n contenedores con valores entre $[0, 1]$, y es interpretado como la distancia

entre los histogramas, entre más se acerca a 0 menos la distancia, este valor de distancia permite calcular el promedio del histograma y darle prioridad a los histogramas más cercanos a dicho promedio[10].

Seguimiento: Se calcula la ubicación del objeto en un frame, este procedimiento se realiza para el cálculo del backproject, donde el valor de cada pixel corresponde a la probabilidad de que el pixel sea del objeto[11].

- **KANADE-LUCAS** algoritmo detecta una serie de puntos en una secuencia de imágenes; estos puntos se les hace un seguimiento, los desplazamientos de los puntos pueden aparecer en diferentes macros correlativos, basándose en la traslación y la deformación lineal[11].

La metodología de este algoritmo es :

Búsqueda: Se buscan automáticamente parejas de puntos de control en el par de imágenes a comparar. Y la imagen es corregida geométricamente según una transformación polinomial y una interpolación de niveles de grises[12].

Compensación: Como las imágenes están bajo diferentes condiciones y de iluminación, éstas son compensadas radiométricamente[12].

Umbralización: Se obtiene la imagen diferencia mediante la diferencia de niveles de grises. Luego se eliminan los cambios poco significativos y se codifica con colores para representar la intensidad de los cambios detectados[12].

3. Metodología

La metodología que se llevó a cabo fue con el fin de poder seleccionar el algoritmo más apropiado, se dividió en las siguientes etapas:

1. Investigación de algoritmos de segmentación: Se investigó información acerca de los algoritmos más reconocidos por su uso en aplicaciones de procesamiento de imágenes enfocados al reconocimiento facial
2. Ventajas y desventajas: En base a los algoritmos seleccionados, investigación de condiciones favorables y desfavorables de cada uno
3. Pruebas: Ejecución en ambientes de prueba según los parámetros analizados
4. Resultados: Elección del algoritmo más óptimo.

4. Análisis algoritmos de reconocimiento

Se puede observar en los cuadros 1,2,3.

Cuadro 1. VIOLA JONES

VIOLA JONES	
Ventaja	Desventaja
1. La clasificación se realiza mediante características en vez de píxel a píxel, lo que permite una cierta abstracción del algoritmo respecto al resultado [13]	1. Tiempo de entrenamiento extremadamente larga [14].
2. Velocidad de detección es muy rápida [13]	2..Movimientos limitados de la cabeza [14]
3. Alta precisión hace exacta la detección de rostros [13]	3.No detecta rostros negros [14]

Cuadro 2. CAMSHIFT

CAMSHIFT	
Ventaja	Desventaja
1.Solidez al movimiento irregular objeto [15]	1. Los resultados de rastreo pueden ser muy fácilmente perturbados por objetos similares. [16]
2. Robustez ante la presencia de ruido de la imagen y la oclusión parcial [15]	2. Los resultados del rastreo son susceptibles a la iluminación cambios. [16]
3. Procesamiento rápido [15]	

Cuadro 3. KANADE-LUCAS

KANADE-LUCAS	
Ventaja	Desventaja
1.Determina la diferencia entre la intensidad de una subimagen en dos momentos: anterior y actual, con el fin de que el seguimiento no se realice por ser separado de la extracción de característica [17]	1.Si una característica se pierde en un marco posterior, el usuario puede solicitar opcionalmente el procedimiento para encontrar otro para mantener el número de características constantes [17]
2.Presenta inconvenientes al detectar una región en específica para el seguimiento [17]	

Código de prueba	Caso 1	Fecha	23/05/2015
Características	Caja negra		
OBJETIVO DE LA PRUEBA: Medir el tiempo de ejecución de cada algoritmo			
Resultado Esperado:	Al ejecutar cada algoritmo de una medida en unidad de tiempo		
Resultados obtenidos:	Cada iteración de ejecución arrojo un resultado en milisegundos		
Aprobado	Si		

Figura 1. Prueba en Tiempo de ejecución.

5. Pruebas

El proceso que se llevó a cabo fue ejecutar los algoritmos para obtener resultados para los siguientes parámetros

- **Tiempo de ejecución:** Para este indicador se tomaron 10 iteraciones (ejecuciones) por cada algoritmo para luego sacar un promedio ponderado de cada uno con el fin de medir el tiempo de respuesta para la detección del rostro, las condiciones con las que se realizó la prueba fueron: un solo rostro para todas las ejecuciones, con luz natural, un fondo neutro y pose definida (Figura 1, Cuadro 4).

- **Rangos de detección:** Cada algoritmo al ejecutarse muestra una ventana donde se ve el rostro del jugador, esta ventana contiene un plano cartesiano. Para este indicador se aprovechó el plano cartesiano tomando los ejes (x,y).

Dentro de la ventana se genera un patrón de reconocimiento con alguna forma geométrica, en el caso de los 3 algoritmos se generó una figura (círculo, rectángulo y puntos) que hace el seguimiento al movimiento del rostro, entonces, se midió el rango máximo que recorre este patrón dentro de la ventana tomando como punto de medida el plano cartesiano determinando así el máximo rango de detección en cada eje a través de la coordenadas, las demás condiciones en esta prueba fueron: un solo rostro para todas las ejecuciones, luz natural, un fondo neutro y una pose definida (Figura 2, Cuadro 5).

- **Estabilidad de detección:** Dentro de la ventana se genera un patrón de reconocimiento con alguna forma geométrica, en el caso de los 3 algoritmos se generó una figura (círculo, rectángulo y puntos) que hace el seguimiento al movimiento del rostro, para este indicador se realizaron movimientos en diferentes direcciones: arriba, abajo, derecha, izquierda, diagonal derecha superior e inferior y diagonal izquierda superior e inferior, con el fin de llegar al límite de la detección, las demás

Cuadro 4. Resultado en Tiempo de ejecución

ALGORITMO	Tiempo de ejecución		
	No iteración	Tiempo	Unidad
CAMSHIFT	1	1166	ms
	2	1114	ms
	3	1145	ms
	4	1316	ms
	5	1250	ms
	6	1100	ms
	7	1137	ms
	8	1136	ms
	9	1130	ms
	10	1142	ms
	Promedio	1163,6	ms
Viola-Jones	1	1285	ms
	2	1260	ms
	3	1252	ms
	4	1264	ms
	5	1256	ms
	6	1261	ms
	7	1262	ms
	8	1264	ms
	9	1278	ms
	10	1292	ms
	Promedio	1267,4	ms
KANADE-LUCAS	1	1126	ms
	2	1207	ms
	3	1152	ms
	4	1214	ms
	5	1182	ms
	6	1194	ms
	7	1149	ms
	8	1118	ms
	9	1156	ms
	10	1123	ms
	Promedio	1162,1	ms

Cuadro 5. Resultado en Rangos de detección.

ALGORITMO	Rangos de detección		
	Dirección	Coordenada en el plano	Eje
CAMSHIFT	Izquierda	55	x
	Derecha	460	x
	Arriba	101	y
	Abajo	216	y
VIOLA-JONES	Izquierda	11	x
	Derecha	368	x
	Arriba	234	y
	Abajo	288	y
KANADE-LUCAS	Izquierda	542	x
	Derecha	571	x
	Arriba	477	y
	Abajo	244	y

condiciones en esta prueba fueron: un solo rostro para todas las ejecuciones, luz natural, un fondo neutro y una pose definida (Figura 3, Cuadro 6).

- **Rostros simultáneos:** Se tomaron 3 rostros diferentes, con el fin de determinar la estabilidad del reconocimiento respecto patrón de seguimiento, para esto se ejecutó cada uno de los 3 algoritmos, cuando apareció el patrón de seguimiento, inmediatamente aparecieron los 2 rostros diferentes de forma gradual, para saber si influían en el reconocimiento y así determinar si afectaba el patrón de seguimiento.

Las características de cada rostro se basaron en los tipos de rostros (rostro ovalado, rostro alargado, rostro cuadrado, rostro redondo, rostro triangular, rostro diamante), y en tono de la piel (blanca, calidad, parda, morena, negra). La luminosidad para esta prueba fue neutra (cuadro 7).

- **Diferentes cámaras web:** Aprovechando que el reconocimiento hecho por cada algoritmo se da en pixeles, Se tomaron 4 diferentes tipos de cámara web para determinar si la calidad (resolución) influye o no y la calidad de la detección y posterior patrón de seguimiento (cuadro 8).

Código Caso de prueba	1	Fecha	23/05/2015
Características	Caja negra		
OBJETIVO DE LA PRUEBA: Medir el tiempo de ejecución de cada algoritmo			
Resultado Esperado:	Al ejecutar cada algoritmo de una medida en unidad de tiempo		
Resultados obtenidos:	Cada iteración de ejecución arrojo un resultado en milisegundos		
Aprobado	Si		

Figura 2. Prueba en Rangos de detección.

6. Resultados

Como resultado según el comportamiento de los algoritmos y la medición de cada indicador se eligió Camshift, ya que arrojo los mejores resultados convirtiéndolo en la mejor opción para realizar el videojuego, arrojo el menor tiempo de ejecución, y fue el más estable a pesar de que los resultados en el rango de detección fueron menores que en los demás algoritmos

Código Caso de prueba	3	Fecha	23/05/2015
Características	Caja negra		
OBJETIVO DE LA PRUEBA:			
PRUEBAS			
Resultado Esperado:	Al ejecutar cada algoritmo, se realizaron movimientos hasta perderse el patron de detección		
Resultados obtenidos:	Cada iteración de ejecución dio sus respectivas conclusiones cuando se intentó o se perdió el patron de ejecución		
Aprobado	Si		

Figura 3. Prueba en Estabilidad de detección.

Cuadro 6. Resultados en Estabilidad de detección

ALGORITMO	Estabilidad de detección	Calificación
CAMSHIFT	El área de seguimiento del rostro se realiza a través de un cirulo el cual nunca pierde la detección al objeto de interés debido a que realiza el reconocimiento basándose en patrones de color.	Buena
VIOLA-JONES	El área de seguimiento se realiza a través de un cuadro, a pesar de que intenta identificar objetos alrededor del rostro, el seguimiento sigue siendo bueno ya que no se pierde.	Regular
KANADE-LUCAS	El área de seguimiento se realiza a través de puntos, cuando se mueve mucho el rostro, los puntos desaparecen.	Mala

Cuadro 7. Pruebas en Rostros simultáneos

CAMSHIFT		
Tipo rostro	Color rostro	Resultado
Ovalado	Moreno	No afecto
Ovalado	Blanco	No afecto
Triangular	Trigueño	No afecto
VIOLA-JONES		
Tipo rostro	Color rostro	Resultado
Ovalado	Moreno	Afecto
Ovalado	Blanco	Afecto
Triangular	Trigueño	Afecto
KANADE-LUCAS		
Tipo rostro	Color rostro	Resultado
Ovalado	Moreno	No afecto
Ovalado	Blanco	No afecto
Triangular	Trigueño	No afecto

Cuadro 8. Pruebas con diferentes cámaras web

Marca y referencia	Resolución	velocidad de cuadro	Resultado
Ecotec 800	640X480	30 fps	No afecto
Microsoft Hd-3000	720 X1280	30 fps	No afecto
Genius Facecam 1005	1280X1040	30 fps	No afecto
Genius iSlim 1320	800x600	15 fps	No afecto

Referencias

- [1] Jorge Guevara-Díaz. Detección de rostros por medio de las wavelets de morlet.
- [2] Cómo utiliza google el reconocimiento de patrones – privacidad y condiciones – google. *Google*.
- [3] Carlos Marcelo Aravena Cereceda. Estudio de métodos para detección y seguimiento de rostros frontales en imágenes digitales: Transformada elíptica lineal y moldes pso. 2009.
- [4] Edwin Arturo Vega Aquino. Detección y seguimiento de rostros. 2011.
- [5] Ernesto Hernandez, Alejandro Cabrera Sarmiento, and Santiago Sanchez Solano. Implementación híbrida hardware software del algoritmo de detección de rostros de viola-jones sobre fpga. *Universidad Ciencia y Tecnología*, 16(63), 2013.
- [6] Francisco Martínez Bolaños. Using computer vision system for detecting tiredness. 2015.
- [7] Marta Lucía Guevara, Julián David Echeverry Correa, William Ardila Urueña, et al. Detección de rostros en imágenes digitales usando clasificadores en cascada. *Scientia*, 2008.
- [8] Cascade classification. *OpenCV 2.4.13.1 documentation*.

- [9] JA Corrales, P Gil, FA Candelas, and F Torres. Tracking based on hue-saturation features with a miniaturized active vision system. In *Proceedings Book of 40th International Symposium on Robotics*, page 107, 2009.
- [10] M Prieto, M Marufo, L Di Matteo, R Verrastro, A Hernández, J Gomez, and C Verrastro. Algoritmo de seguimiento de objetos en imágenes mediante reconstrucción iterativa de histograma en tiempo real.
- [11] Laura Torres Civit. Sistema de reconstrucción de caras 3d basado en modelos y multicámara. 2007.
- [12] V Arévalo, J González, and G Ambrosio. La librería de visión artificial opencv. aplicación a la docencia e investigación. *Base Informática*, 40:61–66, 2004.
- [13] Miguel Delgado Rodríguez. Extracción automática de caras en imágenes captadas con móviles android. 2012.
- [14] Ms Varsha Gupta and Mr Dipesh Sharma. A study of various face detection methods. *METHODS*, 3(5), 2014.
- [15] Zhiyu Zhou, Dichong Wu, Xiaolong Peng, Zefei Zhu, and Kaikai Luo. Object tracking based on camshift with multi-feature fusion. *Journal of Software*, 9(1):147–153, 2014.
- [16] Ryosuke Araki, Seiichi Gohshi, and Takeshi Ikenaga. Real-time both hands tracking using camshift with motion mask and probability reduction by motion prediction. In *Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC), 2012 Asia-Pacific*, pages 1–4. IEEE, 2012.
- [17] N Kasyanenko, A Kraft, and E Reithmeier. Comparison of three image processing algorithms for analysis of in-plane vibrating structures.