

# Proyecto Visión por computador.

Universidad Nacional de Colombia - Sede Manizales.  
 Departamento de ingeniería eléctrica, electrónica y computación.  
 Grupo de Control y Procesamiento Digital de Señales  
 Pavón Gómez Nicolás - 816548.  
 Fecha: 24/06/2020

**Resumen**—Los proyectos de visión por computador son muy útiles e implementados hoy en día, en las industrias es muy cotidiano ver a los operarios de planta realizar movimientos incorrectos, o tienen mala manipulación de los implementos, que impactan a largo plazo en su salud. Mediante las técnicas de visión por computador se implementa el algoritmo de seguimiento de la pose humana (Human Pose Estimation), con el fin de verificar y de ser necesario corregir los movimientos realizados por el personal de la empresa.

**Palabras Clave**—Pose, Visión, Aprendizaje profundo, inteligencia Artificial

## I. INTRODUCCIÓN

La visión por computador o visión artificial es el conjunto de herramientas y métodos que permiten obtener, procesar y analizar imágenes del mundo real con la finalidad de que puedan ser tratadas por un ordenador. Esto permite automatizar una amplia gama de tareas al aportar a las máquinas la información que necesitan para la toma de decisiones correctas en cada una de las tareas en las que han sido asignadas.

### I-A. Análisis de Video.

Así mismo, al tener la capacidad de identificar cada uno de los objetos, esta capacidad se puede utilizar para llevar a cabo un análisis de video. Esto es especialmente útil a la hora de aplicarlo en estrategias de seguridad y control, como la identificación de patrones concretos y búsqueda de los mismos dentro de un sistema de video.

### I-B. Detección de Objetos.

Estos aprendizajes permite la detección de objetos ya que, cuando el ordenador recibe la información visual, puede relacionarla con los patrones aprendidos con anterioridad y, de esta forma, identificar los objetos y diferenciarlos.

Human Pose Estimation es el problema central para entender a las personas en imágenes y videos. Esencialmente siguiendo sus movimientos. El problema es mucho más fácil para hacer el seguimiento a una sola persona, pero se complica un poco más al tener varias personas en una imagen para ver sus movimientos. Para hacer la estimación de pose para cada persona, comúnmente se utiliza un marco de dos pasos utilizando un detector humano. El problema de la estimación de la pose humana puede definirse como las técnicas de visión por computador que predicen la ubicación de varios puntos clave humanos (articulaciones y puntos de referencia) como codos, rodillas, cuello, hombro, caderas, pecho, etc.

### I-C. Ventajas.

- Posicionamiento
- Identificación
- Verificación
- Medición
- Detección de defectos
- Soluciones de visión

## II. METODOLOGÍA.

El modelo tiene de entrada una imagen a color de tamaño  $H \times W$  y produce, como salida, una vector de matrices que consiste en los mapas de confianza de Keypoints(Puntos claves) y Part Affinity Heatmaps(PAF) para cada par de puntos clave. La arquitectura de la red consta de dos etapas:

### Etapa 0.

Las primeras 10 capas del VGGnet (Visual Geometry Group) se utilizan con el fin de crear mapas de características para la imagen de entrada.

### Etapa 1.

Se utiliza un CNN multietapas de dos ramas, de la siguiente forma:

- La primera rama predice un conjunto de Mapas de Confianza en 2D dando las ubicaciones de partes del cuerpo (codo, rodilla, etc.). Un Mapa de Confianza es una imagen en escala de grises que tiene un alto valor en los lugares en que la probabilidad para la parte del cuerpo es alta.
- La segunda rama predice un conjunto de campos vectoriales 2D de Afinidades de Partes (PAF), que codifican el grado de asociación entre partes (puntos clave).

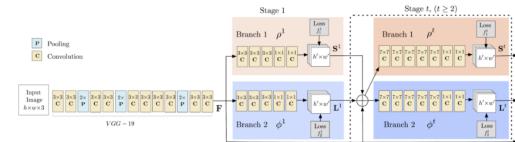


Figura 1. Arquitectura del modelo para la estimación de la pose para Múltiples Personas.

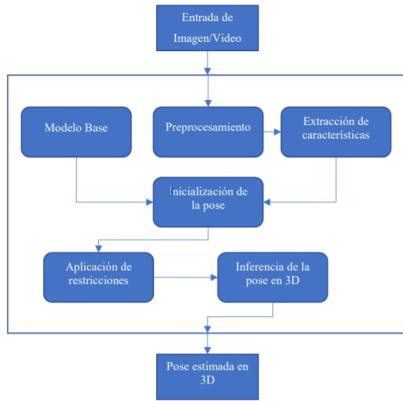


Figura 2. Pasos comunes a los sistemas de estimación de la pose.

La primera etapa toma la imagen de entrada y predice las posibles ubicaciones de cada punto clave en la imagen con una puntuación de confianza (llamado el mapa de confianza).

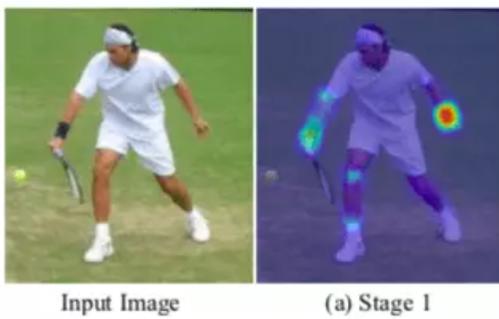


Figura 3. Primer etapa

Se toma la imagen de entrada y genera un mapa de confianza para cada punto clave. En la imagen, se ve el mapa de confianza después de la primera etapa para el codo derecho. La parte resaltada tiene una alta probabilidad de ser el codo derecho.

Cada etapa posterior toma no sólo los datos de imagen, sino también el mapa de confianza de la etapa anterior para utilizar el contexto de la zona cercana de la etapa anterior. Esto mejora la predicción después de cada paso, el mapa de confianza es más preciso después de pasar a través de las 4 etapas.

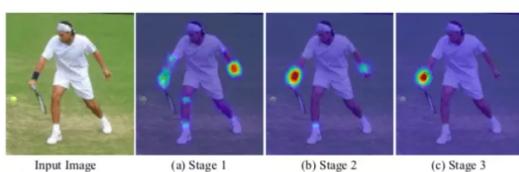


Figura 4. Segunda etapa

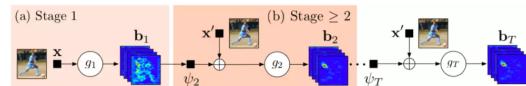


Figura 5. Segunda etapa

La estimación de la pose humana multipersona tiene muchos desafíos adicionales, tales como un número desconocido de personas en la imagen, ocultación, variación en la escala de personas. Utilizando la confianza, sólo puede generar todos los puntos clave como se muestra en la siguiente imagen:

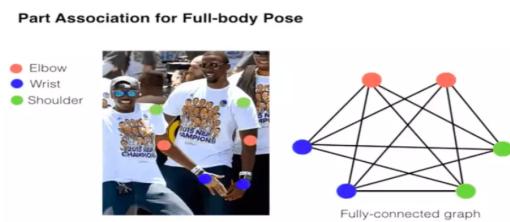


Figura 6. En la estimación de la pose para múltiples personas, las partes del cuerpo que se superponen presentan el problema de identificar cada parte del cuerpo por separado.

Así, se introduce la idea de los mapas de afinidad parcial. Se utiliza el hecho de que ciertas articulaciones se unen mediante las extremidades por lo que otra rama CNN se introduce con el fin de predecir campos vectoriales 2D llamada mapas de afinidad parte (PAF). Los PAF codifican la ubicación y orientación de las extremidades.

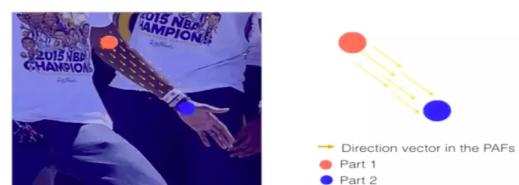


Figura 7. PAF.

Así, se puede utilizar la dirección de los mapas de afinidad para predecir con precisión el problema de estimación de la pose humana para múltiples personas. Ahora, en el paso final, podemos conectar estos puntos usando el algoritmo voraz para generar los puntos clave de la pose para toda la gente en la imagen. La conexión para un tipo de miembros humanos se determina primero y luego se repite para otros tipos de miembros.

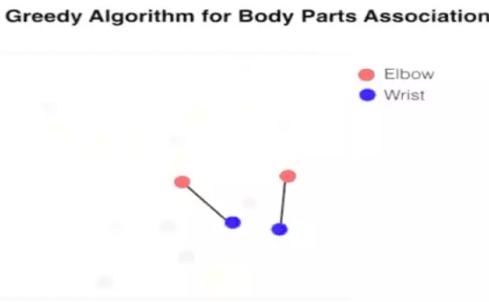


Figura 8. Algoritmo voraz

Un tipo de extremidades se conectan primero y luego esto se repite para otros tipos de extremidades.

Entonces, usamos dos ramas de la arquitectura de la Red Neuronal Convolucional.

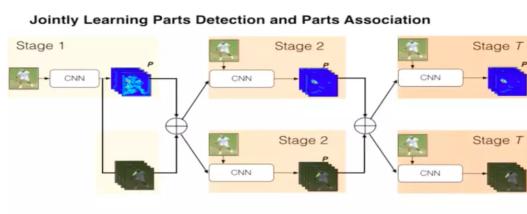


Figura 9. Algoritmo voraz

En general, para realizar la estimación de la pose humana se deben tener en cuenta tres pasos claves:

- **Localización de puntos clave:** Una rama de la red es responsable de predecir todos los puntos clave con una puntuación de confianza. Esto se llama Mapa de confianza.
- **Campos de afinidad:** Otra rama de la red predice un campo de vectores 2D que predice los datos de asociación de puntos clave.
- **Algoritmo voraz:** Ahora, conectamos todos los puntos clave usando el algoritmo.

### III. SETUP EXPERIMENTAL.

### IV. RESULTADOS.

Implementando el algoritmo en la CPU del computador portátil, podemos ver los pantallazos del video resultante para la estimación de la pose humana.

#### Video tomado de internet.



Figura 10. Resultado

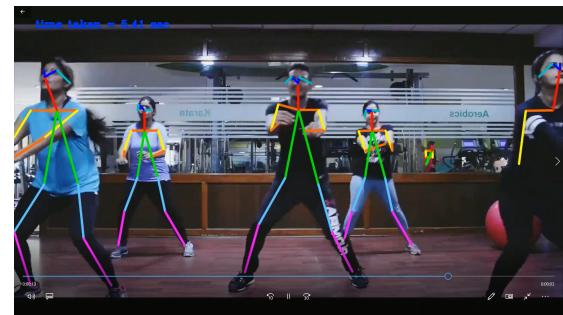


Figura 11. Resultado

#### Video tomado de la cámara del computador portátil.



Figura 12. Resultado

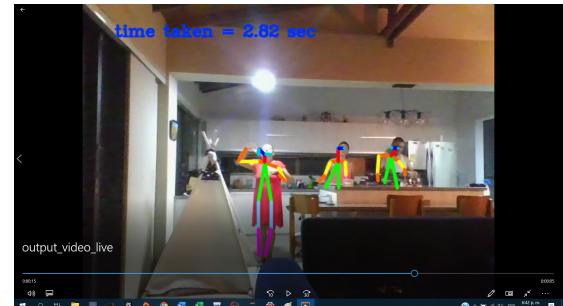


Figura 13. Resultado

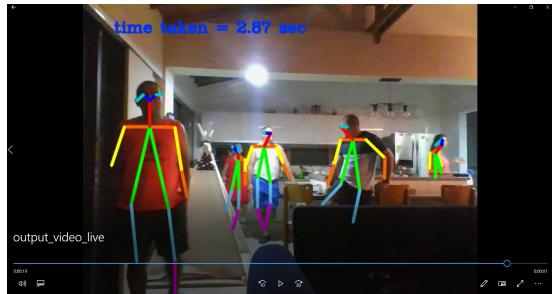


Figura 14. Resultado

#### IV-A. GPU

#### V. CONCLUSIONES

- La estimación de la pose humana, nos permite identificar anomalías en los movimientos, con el fin de evitar problemas de salud.
- Las grandes empresas pueden utilizar este tipo de técnicas para el cuidado de sus empleados.
- Hoy en día las técnicas de visión por computador, son muy útiles; por ejemplo hacer control del distanciamiento social y el uso del tapabocas.