

Sistema de visión artificial para un robot bípedo
jugador de fútbol

Ruth Getzemaní Moreno Cedano

Índice general

1. Introducción	5
1.1. Motivación	6
1.2. Planteamiento del problema	6
1.3. Hipótesis	6
1.4. Objetivos	7
1.5. Descripción del documento	7
2. Theoretical Framework	9
2.1. Transformers	9
2.2.	9
2.3. Vision Transformers	9
2.4. Competencia Robocup Humanoid kid size	9
3. Sistema de reconocimiento de porterías	11
3.1. Filtros Gaussianos	11
3.2. Detección de bordes	13
3.2.1. Métodos basados en gradientes	13
3.3. Detección de contornos	13
3.4. Caracterización de formas	13
3.5. Estimación de orientación	13
4. Reconocimiento de jugadores	15
4.1. Redes neuronales artificiales	15
4.2. Entrenamiento	15
4.3. Redes neuronales convolucionales	15
4.4. Transferencia de aprendizaje	15
5. Implementación	17
5.1. Robots nimbrop y Darwin-op	17
5.2. Simulador Gazebo	17
5.3. Plataforma ROS	17
5.4. Tarjeta de desarrollo nvidia jetson	17

6. Resultados	19
6.1. Detección de porterías	19
6.1.1. Resultados en simulación	19
6.1.2. Resultados en los robots reales	19
6.2. Detección de otros jugadores	19
6.2.1. Resultados en simulación	19
6.2.2. Resultados en los robots reales	19
7. Discusión	21
7.1. Conclusiones	21
7.2. Trabajo futuro	21

Capítulo 1

Introducción

La robótica ha sido testigo de avances significativos a lo largo de los años. Comenzó con robots industriales que realizaban tareas repetitivas en ambientes muy controlados. Hoy, los robots han evolucionado a ser más inteligentes y versátiles[?], lo que los hace ser cada vez más utilizados en ambientes de trabajo y vida cotidiana.[?]

Una de las clasificaciones que destaca son los robots móviles debido a sus habilidades y aplicaciones. En este mismo contexto, uno de los más importantes son los robots humanoides.[?] La importancia de esta clasificación recae en sus características, que pueden ser reducidas a 3; los robots humanoides son capaces de desenvolverse en diversos ambientes, esto permite que no sea necesario alterar el ambiente de trabajo humano y resultaría más económico modificar el robot que el ambiente completo, pueden usar herramientas como los humanos y su forma es similar.[?] Lograr que un robot desempeñe las actividades que realiza un ser humano, como caminar la percepción de objetos ha significado un reto, ya que, para poder decir que una máquina piensa como un humano, debe haber una manera de poder determinarlo.

La inteligencia artificial como un nuevo campo de estudio en la ciencia e ingeniería, se ha desarrollado en cuatro aproximaciones divididas en actuar y pensar humana, y racionalmente. La tarea de percibir y manipular objetos se encuentra dentro del actuar humanamente, donde se involucran las dos disciplinas de interés en este documento: robótica y visión computacional. [?]

La visión es el sentido más poderoso, ya que a través de él se obtiene la información del ambiente y con ello se puede interactuar inteligentemente con el entorno.[?] La visión computacional es la extracción automatizada de información que provee una imagen. Dicha información puede tratarse de modelos 3D, la posición de una cámara, reconocimiento y detección de objetos.[?] Con el paso del tiempo los robots móviles tendrán que explorar cada vez distancias más grandes, por lo que será primordial contar con sensores avanzados y buena capacidad de percepción.

1.1. Motivación

Una de las tareas más importantes de un sistema autónomo de cualquier tipo es adquirir conocimiento de su entorno[?]. Por ende, la visión artificial es una de las habilidades más requeridas en los robots móviles como robots de servicio, de rescate o conductores.

Para fomentar el desarrollo de robots humanoides existen competencias como la RoboCup. En la categoría de humanoid kid size se busca que un robot juegue fútbol en un entorno dinámico, para lo que se requieren varias habilidades. Entre muchas otras cosas, el robot debe percibir y comprender su entorno visual para jugar de una manera efectiva. Con estas competencias lo que se busca es desafiar, probar e impulsar la investigación constante en nuevos algoritmos cada vez más eficientes. [?]

Este tipo de desafíos no dejan de ser parte de una investigación activa, ya que el campo de la visión computacional sigue representando un problema abierto. Puesto que se parte de información insuficiente para dar solución a incógnitas que se quieren resolver, se usan modelos probabilísticos o inteligencia artificial, con el objetivo de eliminar la ambigüedad entre posibles respuestas. Sin embargo, esto sigue siendo muy propenso a fallar.

En este contexto, es conveniente partir del problema hacia la técnica adecuada, es decir, darle un enfoque ingenieril. [?]

1.2. Planteamiento del problema

Al tratarse de un partido de fútbol, es fundamental que el robot pueda cumplir tareas como la detección y localización de la portería, así como detectar a los jugadores oponentes para, esquivarlos y llegar al objetivo. Para abordar estos desafíos, es necesario desarrollar un sistema de visión utilizando exclusivamente cámaras RGB, en conformidad con las regulaciones establecidas por la RoboCup.

Esta restricción busca equiparar las capacidades del robot humanoide con las de un jugador humano. Sin embargo, el uso exclusivo de cámaras RGB plantea sus propios desafíos, como lidiar con variaciones en la iluminación, especialmente en la pelota blanca, la presencia de objetos de colores similares en el entorno y la dificultad para estimar la profundidad. Esto mismo debe ser probado en simulación y en el robot real.

1.3. Hipótesis

El presente trabajo está desarrollado con base en las siguientes hipótesis:

- Al implementar un sistema de visión basado en cámaras RGB se enfrentan desafíos en condiciones de iluminación variable o al haber inferencias visuales significativas, pero estrategias específicas de procesamiento de imágenes permiten contrarrestar estos efectos adversos.

- Es posible asemejar las capacidades perceptuales humanas con un sistema de visión eficaz y cámaras RGB.

1.4. Objetivos

Objetivo general

Desarrollar un sistema de visión artificial para identificar porterías y jugadores en un partido de fútbol con otros robots bípedos.

Objetivos particulares

- Aplicar características geométricas para identificar y localizar porterías.
- Aplicar redes neuronales artificiales para la detección de otros humanoides en la cancha.
- Evaluar el desempeño en un ambiente simulado.
- Evaluar el desempeño en un ambiente real.

1.5. Descripción del documento

Este trabajo se presenta de la siguiente forma: en el capítulo 2 se hace un breve desarrollo acerca de los robots bípedos y su relevancia alrededor del mundo. Así mismo se abordan algunos conceptos de visión artificial necesarios para la utilización de algunos modelos más adelante. También se desglosa el tema de redes neuronales, sus características y tipos ya que serán de relevancia durante una de los principales capítulos de este documento. La última parte trata sobre la competencia Robocup enfocada en la liga de interés; Humanoid kid size. En el capítulo 3 se explica todo lo relacionado con la detección de porterías. Primero se toca el tema de filtros gaussianos como parte del procesamiento de la imagen, donde se explican los puntos claves y el efecto final en la imagen. Continuando con lo anterior se presenta la detección de bordes,

Este trabajo se presenta de la siguiente forma: en el capítulo 2 se hace un breve desarrollo acerca de los robots bípedos y su relevancia alrededor del mundo. Así mismo se abordan algunos conceptos de visión artificial necesarios para la utilización de algunos modelos más adelante. También se desglosa el tema de redes neuronales, sus características y tipos ya que serán de relevancia durante una de los principales capítulos de este documento. La última parte trata sobre la competencia Robocup enfocada en la liga de interés; Humanoid kid size. En el capítulo 3 se explica todo lo relacionado con la detección de porterías. Primero se toca el tema de filtros gaussianos como parte del procesamiento de la imagen, donde se explican los puntos claves y el efecto final en la imagen. Continuando con lo anterior se presenta la detección de bordes,

Capítulo 2

Theoretical Framework

2.1. Transformers

A *Transformer* is defined as a simple network architecture [3] based on attention mechanisms for diverting attention to the most important regions of an image and disregarding irrelevant parts [1] and have a canonical implementation called multi-head attention which is optimized for parallelization. [2] Due to their simple design, Transformers require minimal inductive biases and are naturally suited as set-functions. Also allows processing multiple modalities (images, videos, text, speech). Unlike models that use convolutional neural networks as basic building block,in Transformer, the number of operations required to relate signals from two arbitrary input or output positions is reduced to a constant number of operations. [3]

(a)

2.2.

- Detección y extracción de características

2.3. Vision Transformers

2.4. Competencia Robocup Humanoid kid size

Capítulo 3

Sistema de reconocimiento de porterías

3.1. Filtros Gaussianos

Las imágenes típicamente tienen la propiedad de que el valor de un pixel usualmente es similar al de su vecino. Así mismo es común que contengan componentes de ruido, es decir, variaciones no deseadas en la intensidad de los pixeles, como números aleatorios muy pequeños o pixeles "muertos". Es natural el tratar de reducir los efectos de este ruido reemplazando cada pixel con un promedio ponderado en el valor de los pixeles vecinos, este proceso se conoce como "difuminado." o "desenfoque". El modelo más común para representarlo es el ruido gaussiano. Esto se debe a que sigue una distribución normal, lo que facilita su análisis matemático. Al patrón de pesos usado por un filtro lineal (como el filtro gaussiano) se le conoce usualmente como el "kernel" del filtro. Y el proceso de aplicar el filtro es referido como convolución.[?] Entonces, para difuminar o desenfocar una imagen con un filtro gaussiano, esta se convoluciona con el modelo del kernel Gaussiano simétrico de 2 dimensiones,

$$f(m, n) = \frac{1}{2\pi\sigma^2} \exp\left[-\frac{m^2 + n^2}{2\sigma^2}\right] \quad (3.1)$$

Al final se obtiene una imagen resultante donde cada pixel está representado por una suma ponderada de los pixeles vecinos a él. Estos pesos o ponderaciones de estos pixeles dependerán del valor de la *desviación estándar* o sigma (σ), del modelo gaussiano.[?] Esto se puede entender ya que el análisis se basa en la curva gaussiana, cuya desviación estándar indica la variabilidad respecto al promedio o la media general.

- Para una desviación estándar muy pequeña, el efecto de difuminado o desenfoque será menor, ya que los pesos para todos los pixeles que no estén en el centro serán muy pequeños.

- Para una desviación estándar mayor, los pixeles vecinos tendrán pesos más grandes en el promedio ponderado. Esto resulta en una buena estimación del valor en el pixel central. Un kernel con una desviación estándar grande, sí desaparecerá el ruido en gran medida, pero causará pérdidas de detalles en la imagen.[?]

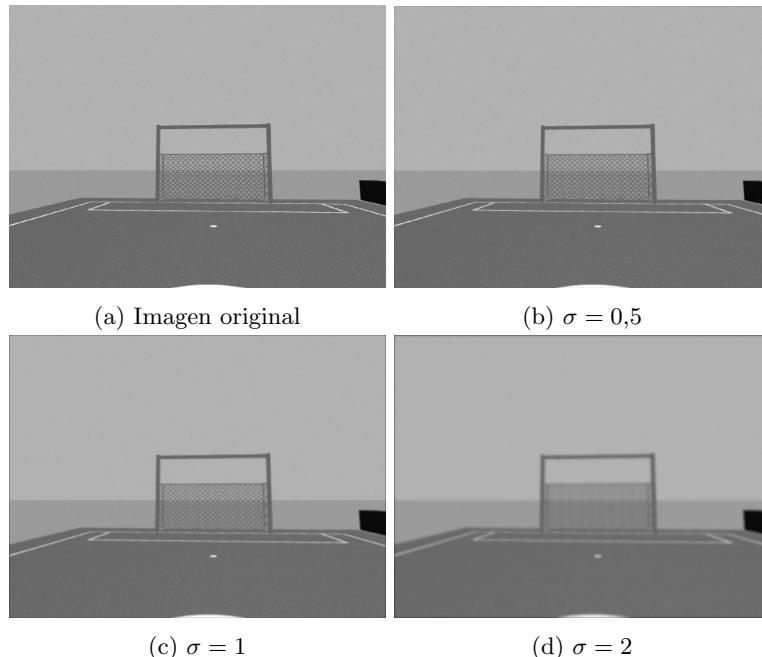


Figura 3.1: Filtro gaussiano aplicado en MATLAB variando el valor de sigma.

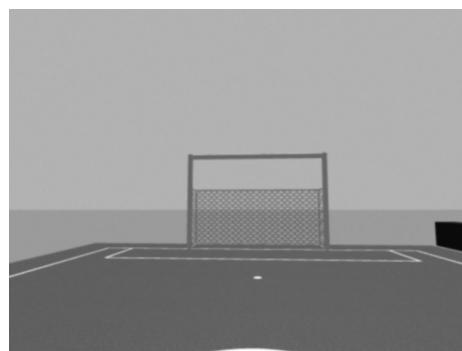


Figura 3.2: Al emplear la función `cv2.GaussianBlur` de OpenCV con un tamaño de 7x7 y un valor de $\sigma = 1$, se obtiene como resultado la figura mostrada. Estos parámetros permiten suavizar la imagen de manera efectiva, minimizando el desenfoque y preservando los bordes de la portería.

3.2. Detección de bordes

La detección de bordes es importante en el área de procesamiento de imágenes, ya que facilita diversas tareas. Consiste en el análisis de los cambios bruscos en la intensidad de los píxeles para obtener información precisa sobre las regiones de interés.[?] Dada una imagen, como la Figura 3.1, ¿cómo se pueden encontrar los bordes de interés o destacados?. Si alguien pidiera dibujar los bordes de los objetos, ¿cuáles serían trazados?. Bien, aunque ya se mencionó de forma cualitativa cuándo ocurren los bordes, lograr una segmentación de la imagen en regiones coherentes es una tarea complicada, así que, frecuentemente se prefiere detectar los bordes usando solamente información local.[?] Desde el punto de vista matemático, un borde puede definirse como una ubicación con un cambio rápido en la intensidad o el color.

3.2.1. Métodos basados en gradientes

Una forma matemática de definir la pendiente y dirección de una superficie es con su gradiente,

$$J(x) = \nabla I(x) = \left(\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y} \right)(x). \quad (3.2)$$

El vector gradiente apunta en la dirección de mayor ascenso en la intensidad de la imagen. Su magnitud indica la fuerza de la variación, mientras que su orientación es perpendicular al contorno local. Uno de los detectores más eficaces y ampliamente utilizados es el detector de bordes de Canny, que ofrece un enfoque de múltiples etapas para detectar bordes de forma resistente al ruido y captura contornos significativos. El algoritmo de detector de bordes de Canny comienza con el suavizado de la imagen con un filtro Gaussiano para suprimir el ruido y reducir el contenido de alta frecuencia. Despues de este paso, la imagen se convoluciona con filtros de derivada horizontal y vertical

3.3. Detección de contornos

3.4. Caracterización de formas

Los momentos de Hu,

3.5. Estimación de orientación

Capítulo 4

Reconocimiento de jugadores

- 4.1. Redes neuronales artificiales**
- 4.2. Entrenamiento**
- 4.3. Redes neuronales convolucionales**
- 4.4. Transferencia de aprendizaje**

Capítulo 5

Implementación

- 5.1. Robots nimbroid y Darwin-op
- 5.2. Simulador Gazebo
- 5.3. Plataforma ROS
- 5.4. Tarjeta de desarrollo nvidia jetson

Capítulo 6

Resultados

- 6.1. Detección de porterías**
 - 6.1.1. Resultados en simulación**
 - 6.1.2. Resultados en los robots reales**
- 6.2. Detección de otros jugadores**
 - 6.2.1. Resultados en simulación**
 - 6.2.2. Resultados en los robots reales**

Capítulo 7

Discusión

7.1. Conclusiones

7.2. Trabajo futuro

Bibliografía

- [1] M.-H. Guo, T.-X. Xu, J.-J. Liu, Z.-N. Liu, P.-T. Jiang, T.-J. Mu, S.-H. Zhang, R. R. Martin, M.-M. Cheng, and S.-M. Hu. Attention mechanisms in computer vision: A survey. *Computational visual media*, 8(3):331–368, 2022.
- [2] S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan, and M. Shah. Transformers in vision: A survey. *ACM computing surveys (CSUR)*, 54(10s):1–41, 2022.
- [3] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.