

Visión por computador: identificación, clasificación y seguimiento de objetos.

Cantero Alonso, Alcides David; Martínez Jara, Eustaquio Alcides
Faculta Politécnica - Universidad Nacional del Este

Resumen

El reconocimiento es una función básica, primordial y compleja de la visión por computador, por medio de la misma el sistema es capaz de aprender a reconocer las formas para posteriormente clasificarlas de forma correcta. Este trabajo propone el desarrollar e implementar un sistema de visión por computador, que realice la identificación de cierto grupo de objetos previamente "enseñados" al sistema, para su posterior clasificación y seguimiento. El sistema es desarrollado utilizando técnicas de procesamiento digital de imágenes y redes neuronales artificiales en distintas etapas conforme a la arquitectura previamente diseñada. La primera etapa consiste en el entrenamiento de la red neuronal artificial, para la cual previamente son necesarios extraer los patrones de los objetos en estudio, siendo este proceso realizado por medio de las técnicas de procesamiento digital de imágenes, entre las que se destaca la utilización de la técnica de segmentación denominada desplazamiento de la media (*Mean-Shift*). El tipo de red neuronal utilizado es el perceptrón multicapa, contando el mismo con una configuración de dos capas ocultas, a más de las capas de entrada y de salida, y la utilización del algoritmo de Levenberg-Marquardt como algoritmo de entrenamiento. En una segunda etapa se realizan tanto la clasificación como el seguimiento de los objetos en estudio. Los resultados obtenidos, al realizar las pruebas con dos conjuntos de objetos diferentes, son realmente satisfactorios. En cuanto a la identificación y clasificación, el análisis de dichos resultados arroja una precisión de la red neuronal artificial por encima del 98%, mientras que para el seguimiento se obtienen resultados muy buenos para escenas en las cuales no existen oclusiones.

Palabras claves: Visión por computador, Redes Neuronales Artificiales, Procesamiento Digital de Imágenes, Patrones, Algoritmos.

1. Introducción

La percepción visual es la actividad que la mayoría de los seres humanos es capaz de realizar desde el momento en que despierta hasta el momento en el que vuelve a dormir.

Uno de los mayores retos que el hombre se propuso es el de comprender el funcionamiento de los órganos de los sentidos, de manera a posteriormente tratar de imitar dicho funcionamiento por medio de las tecnologías existentes buscando así la solución de problemas del mundo real.

La visión por computador (Sucar & Gómez, 2008), surge inspirada en el sistema visual humano, el cual se podría decir que es la mayor fuente de informaciones para las personas y sugiere el tratamiento de las informaciones por medio de diferentes técnicas, las cuales son dependientes de las metas buscadas. La meta de la visión por computador es la de modelar y automatizar el proceso de reconocimiento visual, es decir, distinguir entre objetos. El reconocimiento de objetos es una de las habilidades más fascinantes que el ser humano posee desde niño, pudiendo con un simple esbozo del objeto definir su identidad o la categoría a la cual pertenece el mismo a pesar de las variaciones en la apariencia que pueden deberse a distintos motivos.

1.1. Importancia

La visión por computador (Sucar & Gómez, 2008), es un área relativamente nueva comparada con otras y a pesar de ser muy compleja y difícil se encuentra en constante desarrollo con el surgimiento de nuevas tecnologías con el paso de los años. El reconocimiento es una función básica, primordial y compleja de la visión por computador, por medio de la misma el sistema es capaz de aprender a reconocer las formas para posteriormente clasificarlas de forma correcta (Munoz,

2006), mientras que el seguimiento de los objetos permite analizar y observar el comportamiento o las trayectorias de los objetos en estudio.

1.2. Objetivos

- Procesar las imágenes captadas por un sensor óptico.
- Obtener patrones de los objetos a identificar.
- Estudiar y analizar modelos de Redes Neuronales de modo a utilizar el más apropiado para el problema planteado.
- Entrenar a la Red Neuronal con los patrones obtenidos, con sus respectivas salidas.
- Identificar los objetos captados por el sensor óptico.
- Clasificar los objetos captados por el sensor óptico.
- Realizar un seguimiento de los movimientos del objeto en estudio.

1.3. Visión por computador

La visión por computador, también denominada visión artificial (Platero, 2013), es considerada una rama de la Inteligencia Artificial clásica (Schierwagen, 2013). Esta incluye la serie de fases necesarias para dotar al ordenador de la capacidad de percibir y comprender una imagen de forma a tratar de imitar el proceso que realizan los seres humanos. Si bien la arquitectura y las etapas de un sistema de visión son dependientes de su aplicación, existen algunas etapas que se encuentran en la mayoría de las aplicaciones, las mismas se muestran en la Figura 1.

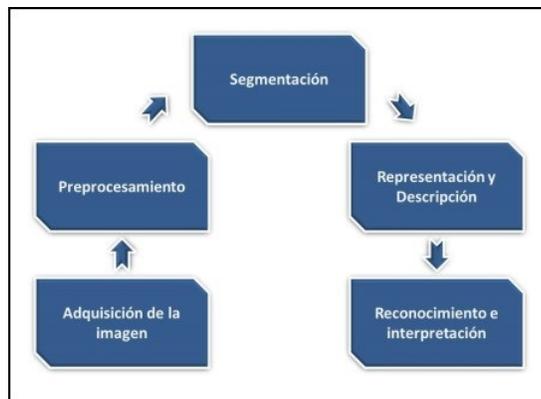


Figura 1. Etapas de un Sistema de Visión Artificial

Los sistemas de visión artificial poseen diferentes tipos de aplicaciones: de medición, de detección de fallas, de corrección y de reconocimiento; las cuales pueden ser utilizadas en áreas como la industria, la medicina y la robótica.

1.3.1. Visión para el reconocimiento de objetos

Es posible considerar dos tipos de reconocimiento (Grauman & Leibe, 2013) si en la imagen se busca un objeto particular o conocido se realiza una detección del objeto, sin embargo si se busca reconocer diferentes instancias de una categoría genérica se realiza el reconocimiento de la instancia. El último tipo es el proceso que divide los objetos en diferentes clases, por ejemplo, reconocer dos marcas y modelos de vehículos diferentes, y según ciertas características compartidas ubicarlos en la clase autos. El reconocimiento de objetos está basado en asignar dichas clases a los diferentes objetos, y la herramienta que realiza este proceso se denomina clasificador (Sonka, Vaclav, & Boyle, 2008).

1.3.2. Reconocimiento de Patrones

Una característica es la unidad elemental utilizada en la representación de los objetos (Marr & Nishihara, 1978). Estos objetos están representados por regiones en una imagen segmentada, los mismos pueden ser divididos en diferentes conjuntos, que desde el punto de la clasificación, se da según ciertas características comunes que tienen dichos objetos entre sí. Algunas de ellas son medidas de la intensidad, medidas del color, esquinas, la textura, entre otros. Estas características son descripciones numéricas elementales utilizadas por la descripción de objetos estadísticos. El objeto medido según sus características es denominado **patrón**.

El patrón $x = (x_1, x_2, x_3, \dots, x_n)$ que describe a un objeto es un vector de características, y el conjunto de todas las formas de patrones posibles es denominado espacio de características X . Los clasificadores son incapaces de reconocer objetos, pero sí lo pueden hacer con los patrones que se les presenta. Los pasos seguidos en el reconocimiento de patrones se muestran en la Figura 2.

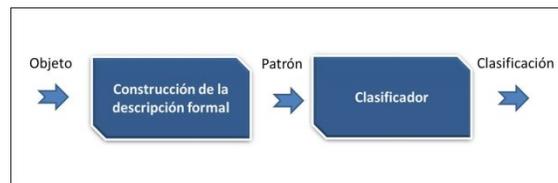


Figura 2 Pasos para el reconocimiento de patrones.

1.3.3. Seguimiento de objetos en tiempo real

Gran parte de las técnicas o algoritmos de seguimiento en escenas dinámicas están basadas en la detección de cambios en una secuencia de imágenes (Yilmaz, Javed & Shah, 2006). De manera simple se puede definir que el seguimiento es el problema de estimar la trayectoria de un objeto en el plano de la imagen mientras se mueve en la escena, para eso el sistema seguidor asigna etiquetas fijas al objeto o los objetos a seguir durante la secuencia de imágenes. Las dificultades en el seguimiento incluyen (Yilmaz, Javed & Shah, 2006) cambios abruptos en el movimiento, el cambio en las apariencias de patrones, tanto de la escena como del objeto mismo y oclusiones entre objetos.

1.4. Procesamiento Digital de Imágenes

El procesamiento digital de imágenes se describe como el conjunto de técnicas que tienen en general tres grandes objetivos (Sucar & Gómez, 2008): mejorar, comprender y extraer medidas de una imagen digital de manera a mejorar el proceso de interpretación visual de la misma por parte de las personas o por el proceso de interpretación de datos por parte de una máquina autónoma en una fase siguiente (Fig. 3).

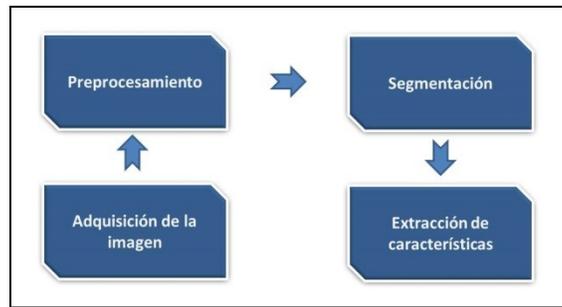


Figura 3. Fases del PDI

En la fase de adquisición de la imagen se adquiere la imagen digital, pudiendo ser adquirida directamente en ese formato o realizando el muestreo y cuantificación de la misma. El preprocesamiento permite obtener imágenes con el mínimo de ruidos no deseados de manera a facilitar el trabajo en las siguientes etapas. En esta etapa se llevan a cabo las operaciones sobre las imágenes.

La segmentación es el proceso por el cual se aíslan los objetos de la escena de manera a ser analizados individualmente. Por último en la fase de extracción de características se convierten los datos como bordes o puntos de la imagen de manera a que sean legibles por el ordenador.

1.5. Redes Neuronales Artificiales

Las Redes Neuronales artificiales, son aquellas utilizadas con el fin de imitar el comportamiento del cerebro humano, otorgándole a un objeto no pensante una aproximación a la capacidad de procesamiento de información que tienen las personas. Pueden ser utilizadas para la solución de problemas de manera individual o en combinación con otras técnicas.

Son capaces de resolver problemas de identificación, optimización, predicción y clasificación. Los elementos con los cuales consta una red neuronal artificial son: la **neurona**, donde se llevan a cabo los cálculos a realizar, casi siempre una suma ponderada de las **entradas** proporcionadas, al que se le aplica una **función de activación** que genera una **salida**. Esta estructura básica fue propuesta por McCulloch y Pitts (Haykin, 1999) (Fig. 4).

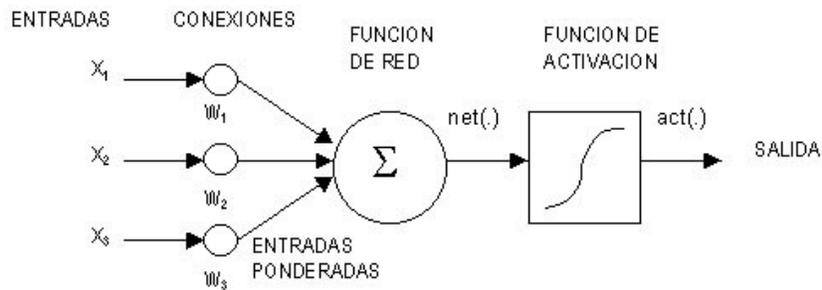


Figura 4. Estructura básica de una Red Neuronal.

Existen dos fases de una red neuronal, la **fase de entrenamiento**, en la se utiliza un conjunto de datos de entrada para determinar los pesos que definen a la red. Estos se calculan de forma iterativa con el objetivo de ir minimizando el error hasta obtener una salida próxima a la deseada. La otra fase, es denominada **fase de prueba o recuperación de la información**, en la cual se utiliza un distinto conjunto de datos, de los cuales se obtiene sus salidas.

Existen varios modelos de redes neuronales artificiales, entre las cuales se pueden citar: Perceptrón, Perceptrón multicapa, Redes de Hopfield, Redes de neuronas de base radial, Backpropagation, entre otros.

- **Perceptrón Multicapa**

El perceptrón multicapa o *mip* es una red neuronal supervisada que puede tomar como entrada valores continuos o discretos (binarios), normalmente constituida por un conjunto de unidades sensoriales (capa de entrada), una o más capas ocultas y una capa de salida como se muestra en la Figura 5.

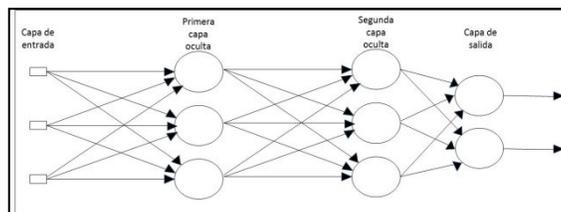


Figura 5. Estructura de un Perceptrón Multicapa.

Al poseer la red capas intermedias, el desafío fue encontrar un algoritmo de aprendizaje para la actualización de los pesos. Sin dudas la aparición del algoritmo de retropropagación, en inglés *backpropagation*, fue la piedra angular para el resurgimiento del interés por las redes neuronales, siendo desde entonces uno de los algoritmos más estudiados y utilizados para el proceso de aprendizaje (Rojas, 1996).

2. Materiales y Métodos

2.1. Recursos utilizados

Hardware

- Cámara web
- Notebook
- Memoria 16GB
- Procesador Intel i7

Software

- Matlab R2012B

Otros

- Artefacto lumínico de manera a mejorar la adquisición de las imágenes.

2.2. Arquitectura del Sistema

La arquitectura del sistema se define como en la Figura 6.

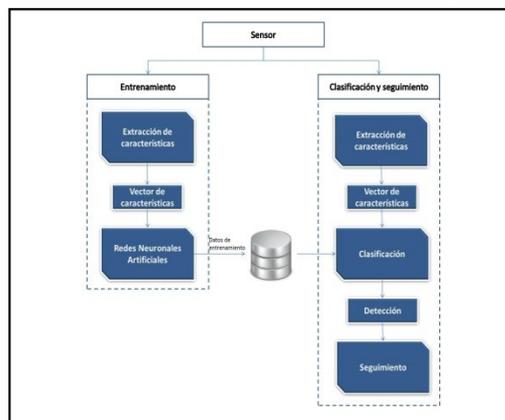


Figura 6. Arquitectura del Sistema

2.2.1. Módulo de extracción de características

El objetivo final de este módulo es obtener todas aquellas características más representativas de los objetos a clasificar de manera a almacenarlas y utilizarlas como datos de entrada del clasificador a utilizar (Fig 7).

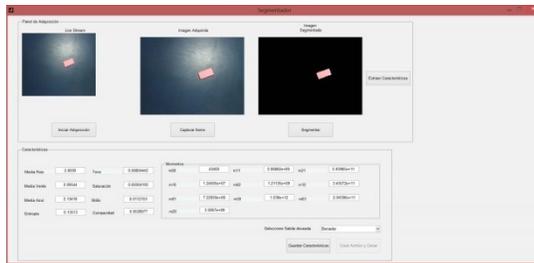


Figura 7. Módulo de extracción de características.

De manera a lograr dicho objetivo se realizan las tareas que abarcan operaciones de procesamiento digital de imágenes. Para la segmentación del objeto fue utilizada la técnica del desplazamiento de la media.

En total son extraídas 18 características distintivas de los objetos que son guardadas en una matriz, con sus respectivas salidas en otra matriz.

- El valor medio del componente Rojo.
- El valor medio del componente Verde.
- El valor medio del componente Azul.
- El tono.
- La saturación.
- El brillo.
- La compacidad.
- La entropía.
- Los momentos del objeto, que totalizan 10.

2.2.2. Módulo de entrenamiento

Las 18 características extraídas en la etapa anterior son utilizadas en esta etapa de manera a ser presentadas al clasificador a ser utilizado, en este caso una red neuronal artificial (Fig. 8).

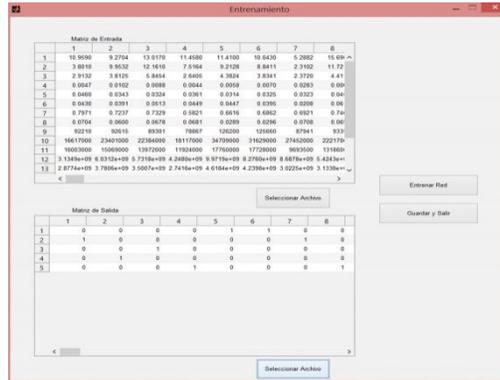


Figura 8. Módulo de entrenamiento.

El tipo de RNA elegida para utilizarse en la aplicación es el Perceptrón multicapa (Fig. 9), debido a su capacidad de resolver con éxito un gran número de problemas. La función de entrenamiento elegida es la que utiliza el método Levenberg-Marquardt, teniendo el error medio cuadrático como parámetro de rendimiento.

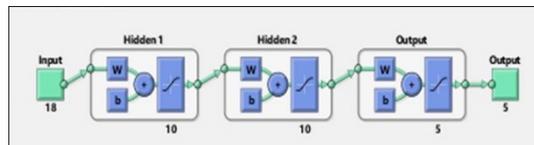


Figura 9. Arquitectura de la RNA utilizada.

Para el entrenamiento de la RNA, se asignan el 60% de las muestras para el conjunto de entrenamiento y el 40% para el conjunto de test de manera a utilizarlos en la etapa de reconocimiento.

2.2.3. Módulo de clasificación y seguimiento

Ejecuta todo el proceso realizado en las etapas anteriores, pero con la red ya entrenada de manera a clasificar los objetos que se encuentran en la escena enviada por la cámara web y luego realizar el seguimiento (Figs. 10 y 11).



Figura 10. Módulo de clasificación y seguimiento

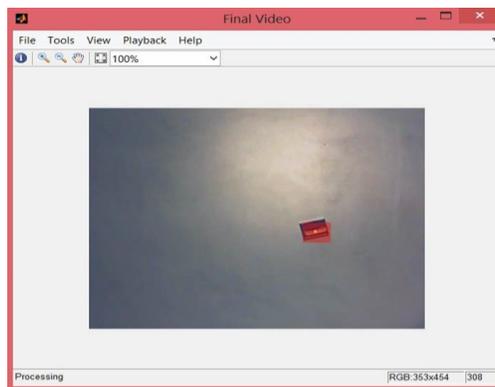


Figura 11. Ventana de seguimiento del objeto en tiempo real

3. Resultados obtenidos

De manera a comprobar el funcionamiento del sistema se realizaron pruebas con dos conjuntos de objetos, divididos en 5 clases.

El primer conjunto de objetos, denominado C1, lo conforman las clases:

- Lápiz.
- Disco Compacto (CD)
- Cuaderno.
- Sacapuntas.
- Borrador.

Mientras que para las siguientes pruebas se utilizó, el conjunto de objetos C2, el cual está conformado por 5 clases de frutas:

- Banana.

- Manzana roja.
- Manzana verde.
- Pera.
- Naranja.

Como criterio de parada de la red neuronal se utiliza el error medio cuadrado con un valor de 1×10^{-9} .

3.1. Reconocimiento del conjunto C1

El conjunto se encuentra conformado por 105 muestras, las cuales fueron divididas: 60% para el conjunto de entrenamiento y 40% para el conjunto de test, es decir 63 y 42 muestras respectivamente.

El proceso de entrenamiento se detiene luego de 26 iteraciones cuando el parámetro del error es minimizado al valor deseado (Fig. 12).

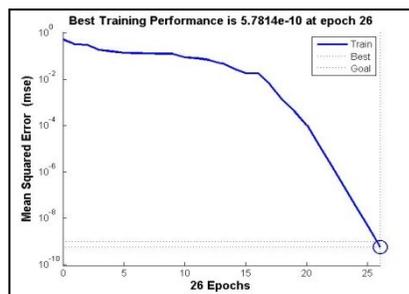


Figura 12. Minimización del error durante el proceso de entrenamiento

La matriz de Confusión es una matriz usada para realizar el análisis de los resultados de las pruebas de identificación y clasificación la cual se muestra en la figura 13, ella indica el número de aciertos (color verde) y errores (color rojo) en la clasificación de los objetos para el conjunto C1 de test, teniendo en cuenta estas consideraciones:

- Clase 1: Lápiz.
- Clase 2: Disco Compacto.
- Clase 3: Cuaderno.
- Clase 4: Sacapuntas.

- Clase 5: Borrador.

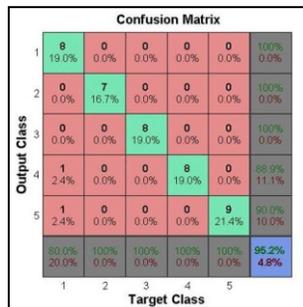


Figura 13. Matriz de confusión del conjunto C1

3.2. Reconocimiento del Conjunto C2

El conjunto posee 114 muestras, las cuales fueron divididas en un 60% para el proceso de entrenamiento y un 40% para el conjunto de test, es decir 68 y 46 muestras respectivamente. El proceso de entrenamiento se detiene luego de 16 iteraciones (épocas), cuando el parámetro del error llega al mínimo deseado (Fig. 14).



Figura 14. Minimización del error durante el proceso de entrenamiento

La figura 15 indica el número de aciertos (color verde) y errores (color rojo) en la clasificación de los objetos para el conjunto de test, teniendo en cuenta estas consideraciones:

- Clase 1: Banana.
- Clase 2: Manzana Roja.

- Clase 3: Manzana Verde.
- Clase 4: Pera.
- Clase 5: Naranja.

Output Class	1	2	3	4	5	
1	10 21.7%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
2	0 0.0%	8 17.4%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
3	0 0.0%	0 0.0%	7 15.2%	0 0.0%	0 0.0%	100% 0.0%
4	0 0.0%	0 0.0%	2 4.3%	8 17.4%	0 0.0%	80.0% 20.0%
5	0 0.0%	0 0.0%	0 0.0%	0 0.0%	11 23.9%	100% 0.0%
	100% 0.0%	100% 0.0%	77.8% 22.2%	100% 0.0%	100% 0.0%	95.7% 4.3%
	1	2	3	4	5	
	Target Class					

Figura 15. Matriz de confusión del conjunto C2

3.3. Análisis de los resultados

Para realizar el análisis de los resultados de las pruebas de identificación y clasificación se utilizaron matrices de confusión, presentadas en las secciones anteriores, de manera a obtener la **sensibilidad**, la **razón de falsos positivos**, la **especificidad** y la **precisión** de la red neuronal artificial.

A partir de dichas matrices son calculados el número total de verdaderos positivos, falsos negativos, falsos positivos y verdaderos negativos, los cuales son necesarios para realizar la evaluación de la red.

Para el **CONJUNTO C1**, la red arroja 40 verdaderos positivos, 2 falsos positivos, 2 falsos negativos y 166 verdaderos negativos. Los resultados según los criterios presentados anteriormente se presentan en la Tabla 1.

Criterio	Fórmula	Porcentaje
Sensibilidad	$VPR = \frac{40}{(40 + 2)} = 0,952$	95,2%
Razón de Falsos Positivos	$FPR = \frac{2}{(2 + 166)} = 0,012$	1,2%
Especificidad	$SPC = 1 - 0,012 = 0,988$	98,8%
Precisión	$ACC = \frac{40 + 166}{42 + 168} = 0,981$	98,1%

Tabla 1. Análisis de los resultados para el conjunto C1

Para el **CONJUNTO C2**, la red arroja 40 verdaderos positivos, 2 falsos positivos, 2 falsos negativos y 166 verdaderos negativos. Los resultados según los criterios presentados anteriormente se presentan en la Tabla 2.

Criterio	Fórmula	Porcentaje
Sensibilidad	$VPR = \frac{44}{(44 + 2)} = 0,957$	95,7%
Razón de Falsos Positivos	$FPR = \frac{2}{(2 + 182)} = 0,011$	1,1%
Especificidad	$SPC = 1 - 0,011 = 0,989$	98,9%
Precisión	$ACC = \frac{44 + 182}{46 + 184} = 0,983$	98,3%

Tabla 2. Análisis de los resultados para el conjunto C2

Pruebas del seguimiento

El algoritmo de seguimiento utilizando la técnica del desplazamiento de la media, presenta buenos resultados cuando la escena es limpia, sin oclusiones del objeto.

La oclusión se entiende como la desaparición parcial o total del objeto en análisis. En los casos en los cuales existen oclusiones del objeto o cuando el objeto se encuentra en parte fuera de la escena el seguimiento se pierde.

Para movimientos rápidos del objeto el algoritmo tarda un tiempo en volver a seguir al objeto, aunque lo vuelva a encontrar (Fig. 16).

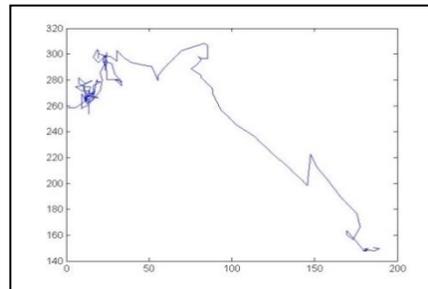


Figura 16. Trayectoria del objeto seguido

4. Conclusiones

Se ha desarrollado una aplicación de visión por computador, con la capacidad de identificar, clasificar y seguir ciertos objetos. Luego del análisis de los resultados, es posible manifestar que el sistema posee una eficiencia y una eficacia muy alta.

En cuanto a la implementación de la red neuronal artificial, con los valores obtenidos para ambos conjuntos de datos, es posible expresar que la misma funciona de manera bastante precisa para la clasificación de los objetos presentados, teniendo en cuenta que obtener clasificadores con una eficacia del 100% es muy difícil de lograr.

Un aspecto a mejorar es el módulo de seguimiento, principalmente para el problema de la oclusión del objeto lo que generaba la pérdida del seguimiento por parte del sistema. Otro de los mayores problemas presentados fue el de la iluminación de la escena, ya que un cambio mínimo en la misma afecta críticamente a todo el proceso.

5. Referencias

- Grauman, K. & Leibe, B. (2013). Visual Object Recognition. Recuperado de <http://cs.gmu.edu/~kosecka/cs482/grauman-recognition-draft-27-01-11.pdf>
- Haykin, S. (1999). Neural Networks: A comprehensive Foundation. New Jersey: Prentice Hall International, Inc.
- Marr, D. & Nishihara, H. K. (1978). Representation and Recognition of the Spatial Organization of Three-Dimensional Shapes. *Proceedings of the Royal Society of London*, 200(1140).
- Munoz, D. J. (2006). Proceso de reconocimiento de objetos, asistido por computador (Visión artificial), aplicando gases neuronales y técnicas de minería de datos. *Scientia et Technica*, 30.
- Platero, C. (2013). Introducción a la Visión Artificial. Recuperado de http://www.elai.upm.es/webantigua/spain/Asignaturas/MIP_VisionArtificial/ApuntesVA/cap1IntroVA.pdf
- Rojas, R. (1996). Neural Networks. A systematic Introduction. Berlin: Springer.
- Schierwagen, A. (2013). Vision as computation, or: Does a computer vision system really assign meaning to images? Recuperado de <http://www.informatik.uni-leipzig.de/~schierwa/vision.pdf>

Sonka, M., Vaclav, H. & Boyle, R. (2008). *Image Processing, Analysis, and Machine Vision*. Toronto: Thomson.

Sucar, L. E. & Gómez, G. (2008). Instituto Nacional de astrofísica, óptica y electrónica. Recuperado de <http://ccc.inaoep.mx/~esucar/Libros/vision-sucar-gomez.pdf>

Yilmaz, A., Javed, O. & Shah, M. (2006). Object Tracking: A Survey. *ACM Computing Surveys*, 38(4), 45.