Aprendizaje Automático. Aplicaciones en Visión por Computadora

L. Lanzarini[1], C. Estrebou[1], F. Ronchetti[1], F. Quiroga[1], C. Luna[1], R. Antonio[1], L. La Frazia[1], A. Rosete[2]
[1] Instituto de Investigación en Informática LIDI, Facultad de Informática, UNLP, La Plata, Argentina. *
[2] Universidad Tecnológica de La Habana "José Antonio Echeverría" (CUJAE), La Habana, Cuba

*** Centro asociado de la Comisión de Investigaciones Científicas de la Pcia. De Bs. As. (CIC)**

{laural, cesarest, fronchetti, fquiroga}@lidi.info.unlp.edu.ar
{carla.lunagennari, lucholafrazia}@gmail.com, ramiro.antonio@outlook.com,
rosete@ceis.cujae.edu.cu

CONTEXTO

Esta presentación corresponde al proyecto "Sistemas Inteligentes. Aplicaciones en Reconocimiento de Patrones, Minería de Datos y Big Data." (Periodo 2018–2021) del Instituto de Investigación en Informática LIDI.

RESUMEN

Esta línea de investigación se centra en el estudio y desarrollo de Sistemas Inteligentes para la resolución de problemas de reconocimiento de patrones en imágenes y video, utilizando técnicas de Aprendizaje Automático. El trabajo presentado describe diferentes casos de aplicación en visión por computador de técnicas tanto supervisadas como no supervisadas.

Uno de los principales problemas desarrollados es el reconocimiento de lengua de señas. Este es un caso que presenta diversas aristas a atacar como el reconocimiento del intérprete, la segmentación de manos, la clasificación de diferentes configuraciones y de un gesto dinámico, entre otros problemas.

Con respecto a la segmentación de manos se realizaron diferentes trabajos, tanto utilizando marcadores de colores, redes neuronales capaces de reconocer el color de la piel de una persona, como así también redes convolucionales.

Por otro lado, para llevar a cabo la clasificación de diferentes gestos dinámicos, incluyendo la lengua de señas, se realizó un clasificador dinámico capaz de identificar acciones humanas que faciliten la interfaz hombre/máquina.

En el área del procesamiento de video se está comenzando a investigar sobre detectores de peatones y automóviles para utilizar con cámaras instaladas en la vía pública.

Adicionalmente, se están realizando trabajos de clasificación de imágenes de especies de serpientes utilizando técnicas clásicas del aprendizaje automático

Palabras clave: Aprendizaje Automático, Visión por Computadoras, Redes Neuronales, Reconocimiento de Patentes, Lengua de Señas, Clasificación de Serpientes, Reconocimiento de Peatones y Autos.

1. INTRODUCCION

El Instituto de Investigación en Informática LIDI tiene una larga trayectoria en el estudio, investigación y desarrollo de Sistemas Inteligentes basados en distintos métodos de Aprendizaje Automático. Los resultados obtenidos han sido medidos en la solución de problemas pertenecientes a distintas áreas.

En el III LIDI, desde hace varios años se viene trabajando en el procesamiento de señales de audio y video. Como resultado de estas investigaciones se han diseñado e implementado técnicas originales aplicables al reconocimiento tanto de gestos dinámicos como de detección de patrones en videos en diferentes problemas. En relación con esta línea, actualmente se están desarrollando los siguientes temas:

1.1. Reconocimiento de la lengua de señas

El reconocimiento de la lengua de señas es un campo de investigación relativamente nuevo cuyo objetivo final es traducir de la lengua de señas a una lengua escrita. Esto implica poder tomar un video en donde una persona habla en lengua de señas, y reconocer la posición de su cuerpo, su cara y sus manos, la expresión de su rostro, la forma de sus manos y también la de sus labios si la seña requiere pronunciar la palabra para desambiguar. Con esa información, se debe reconocer la seña realizada, para luego a partir de una secuencia de señas generar una traducción a una lengua escrita [6].

En esta área, se publicó un método probabilístico para clasificar señas en videos segmentados que abarca todas las etapas del reconocimiento [1, 2]. Este método probabilístico utiliza tres componentes esenciales en una seña: la posición de las manos, la forma y el movimiento. Cada componente es primero analizado por separado por sub-clasificadores para luego ser unificado en un clasificador global.

Este método no utiliza la información secuencial de la seña, es decir, no utiliza la información temporal. No obstante, los resultados de los experimentos muestran que aún con esa dificultad se pueden clasificar correctamente el 96% de las señas del conjunto de prueba.

Para llevar a cabo la validación de los métodos desarrollados, se utilizaron dos conjuntos de datos recolectados en el III-LIDI: LSA16 y LSA64. El primero, LSA16, contiene 800 imágenes con 16 clases de formas de mano y el segundo, LSA64, está formado por 3200 videos de 64 clases de señas dinámicas. Los detalles de la base de señas dinámicas LSA64 se encuentran publicados en [3].

Como resultado de esta línea de investigación se culminó exitosamente una tesis doctoral. Actualmente se están ampliando los métodos desarrollados para poder realizar una implementación de los métodos en un entorno real, lo que contribuiría con la comunidad para facilitar la traducción entre la lengua de señas y el castellano.

1.2. Clasificación de formas de manos

Siguiendo con la temática de la sección anterior, una de las líneas investigadas es la clasificación de formas de manos. Las lenguas de señas utilizan un conjunto finito de formas de mano, que, en combinación con movimientos de las manos y el cuerpo, y expresiones faciales, se utilizan para señar [1].

En base a estudios previos [1,5], una etapa fundamental en el reconocimiento de la lengua de señas es la clasificación de estas formas de mano, y por ende un área prioritaria para mejorar el reconocimiento.

A su vez, las redes neuronales convolucionales, que han establecido nuevos estados del arte en casi todas las aplicaciones de visión por computadora, son idóneas para esta tarea. No obstante, no existe una revisión sistemática de la aplicabilidad de los diversos modelos de redes convolucionales a la clasificación de las formas de mano.

Por ende, se realizó una evaluación de desempeño de distintos modelos de redes convolucionales (LeNet, Inception, VGG, ResNets y AllConvolutional) con dos bases de datos de formas de mano (LSA16 [2] y RWTH-PHOENIX-Weather [5]).

Método	LSA16	RWTH
DeepHand (VGG)[5]	-	85.50
ProbSom[2]	92.30	-
Feedforward	86.58	60.27
LeNet	95.78	81.19
AllConvolutional	94.56	80.29
VGG	95.92	82.88
ResNet	93.49	80.89
Inception	91.98	75.33
Inception+SVM (pretrained)	93.67	78.12
Inception+NN (pretrained)	80.62	75.97

Tabla 1: Desempeño de cada método en LSA16 y RWTH-Phoenix-Weather. Los números representan porcentajes de ejemplos clasificados correctamente.

Tal como se muestra en la *Tabla 1*, se encontró que, salvo por Inception, todos los modelos consiguen un buen desempeño, incluso superando al estado del arte en LSA16. Además, se pudo observar que las redes convolucionales mejoran su desempeño al pre-segmentar la forma de la mano removiendo el fondo.

1.3. Localización de partes del cuerpo

Una etapa fundamental de todo el proceso de reconocimiento de un gesto en una imagen es la segmentación de las manos. En este sentido se están explorando distintas alternativas de segmentación de manos por el color de piel (sin marcadores) utilizando redes neuronales. En esta área se ha realizado un estudio del comportamiento de las redes neuronales RCE para segmentar la piel por color y su efectividad en diferentes sistemas de representación de color. Como resultado se llegó a la conclusión de que tanto el tiempo de cómputo como la efectividad de la segmentación son similares sin importar el sistema de representación de color elegido.

Este algoritmo de segmentación basado en la red RCE se ha utilizado en un prototipo de hardware y software que reconoce gestos dinámicos simples, en tiempo real, para controlar dispositivos electrónicos, en particular el control de un TV [7].

1.4. Detección de vehículos y peatones.

Continuando la línea de trabajos anteriores de reconocimiento de patentes, localización de objetos, y reconocimiento de acciones, se está desarrollando un proyecto en el que participan alumnos de grado de la Facultad de Informática para reconocer vehículos y peatones en cámaras de vigilancia.

El objetivo del proyecto es desarrollar la tecnología de base para un sistema de software que podría ser utilizado por los municipios para monitorear la entrada y salida de vehículos, controlar la velocidad de los mismos, y detectar posibles accidentes viales.

Con ese fin, se están evaluando distintos detectores de objetos, ya sean basados tanto en técnicas tradicionales como HOG y SVM, o en redes convolucionales como YOLO y R-CNN, utilizando bases de datos estándar como la Daimler Pedestrian Dataset o la Caltech Cars dataset.

1.5. Clasificación de especies de serpiente

En el marco de reconocimiento de objetos presentes en una imagen se está trabajando en conjunto con profesionales del CEPAVE[1] con el objetivo de desarrollar e implementar un algoritmo capaz de identificar la especie a la que pertenece una serpiente a partir de la imagen de un ejemplar. Interesan especialmente las especies de serpientes de la provincia de Buenos Aires y en particular las del partido de La Plata. Esto, si bien reduce considerablemente el número de opciones, no invalida la importancia y aplicabilidad de las investigaciones realizadas.

Cabe destacar la importancia de esta aplicación ya que permite determinar la peligrosidad de una serpiente. Por un lado, permite tomar decisiones rápidas ante un incidente y por otro lado contribuye a la conservación de las especies ya que las personas ante la duda de su peligrosidad deciden matarlas.

En esta aplicación en desarrollo, se han utilizado los algoritmos SIFT, SURF y ORB para extraer los descriptores locales de la imagen de una serpiente. Se han analizado distintas métricas de comparación de características para determinar la correspondencia entre los descriptores. Los resultados preliminares arrojan una tasa de acierto superior al 70% pero aún queda trabajo por realizar.

2. LINEAS DE INVESTIGACIÓN Y DESARROLLO

- Estudio de técnicas de segmentación de objetos en movimiento presentes en un

[1] CEPAVE Centro de Estudios Parasitológicos y de Vectores - Conicet La Plata – U.N.L.P.

video.

- Representación y clasificación de configuraciones de manos para el lenguaje de señas.
- Clasificación de señas dinámicas.
- Estudio y análisis de las distintas representaciones de color.
- Redes neuronales competitivas dinámicas. Redes neuronales RCE.
- Detección y extracción de características. Puntos de interés y descriptores de regiones.
- Estudios de performance de los algoritmos desarrollados

3. RESULTADOS OBTENIDOS/ESPERADOS

- Desarrollo de un modelo de clasificación de señas segmentadas y comparación de su desempeño con otros modelos del estado del arte.
- Evaluación de desempeño de varias arquitecturas de redes convolucionales del estado del arte para la clasificación de formas de manos.
- Desarrollo de un método de clasificación de especies de serpientes basados en descriptores SIFT, SURF y ORB.
- Detección y clasificación de peatones y vehículos mediante cámaras RGB.

4. FORMACIÓN DE RECURSOS HUMANOS

El grupo de trabajo de la línea de I/D aquí presentada está formado por: 2 profesores con dedicación exclusiva, 2 becarios de investigación UNLP con dedicación docente, 1 becario CIN, 2 tesistas y 1 profesor extranjero.

Dentro de los temas involucrados en esta línea de investigación, en el último año se han finalizado 2 tesis de doctorado, y 2 tesinas de grado de Licenciatura.

Actualmente se están desarrollando 1 tesis de doctorado, 1 tesis de especialista y 3 tesinas de grado de Licenciatura. También participan en el desarrollo de las tareas becarios y pasantes del III-LIDI.

5. REFERENCIAS

[1] Ronchetti F., Quiroga F., Estrebou C., Lanzarini L., Rosete A. *Sign languague recognition without frame-sequencing constraints: A proof of concept on the argentinian sign language*. Publicado en Ibero-American Conference on Artificial Intelligence IBERAMIA 2016 (pp. 338-349)

[2] Ronchetti, F., Quiroga, F., Estrebou, C.A., Lanzarini. *Handshape recognition for Argentinian Sign Language using ProbSom*. Journal of Computer Science & Technology, vol. 16, N° 1, págs. 1-5, ISSN 1666-6038, 2016.

[3] Ronchetti, F., Quiroga, F., Estrebou, C.A., Lanzarini, L.C., Rosete, A. . *LSA64: An Argentinian Sign Language Dataset*, publicado en el XXII Congreso Argentino de Ciencias de la Computación (CACIC 2016) (pp. 794-803).

[4] Quiroga, F., Antonio, R., Ronchetti, R., Lanzarini, L., Rosete, A. *A Study of Convolutional Architectures for Handshape Recognition applied to Sign Language,* publicado en el XXIII Congreso Argentino de Ciencias de la Computación (CACIC 2017) (pp. 13-22).

[5] Koller, O., Ney, H., Bowden, R. . *Deep Hand: How to Train a CNN on 1 Million Hand Images When Your Data Is Continuous and Weakly Labelled.* Computer Vision and Pattern Recognition Conference (CVPR 2016) (pp. 3793-3802).

[6] Cooper H., Holt B., and Bowden R. *Sign language recognition.* In Visual Analysis of Humans: Looking at People, chapter 27, Springer, 2011. (pp539- 562).

[7] Luna Gennari C., Estrebou C., Lanzarini, L. *Reconocimiento de gestos aplicado al control de dispositivos,* publicado en el XXIII Congreso Argentino de Ciencias de la Computación (CACIC 2017) (pp.1040-1049).