



REVISTA ARISTOS

Ciencia Básica y Aplicada

VOLUMEN 7

NUMERO 14

AÑO 2019

CONGRESO INTERNACIONAL Ci2T 2019



NUMERO ESPECIAL

ISSN 2007-9478



UNIVERSIDAD AUTONOMA DE BAJA CALIFORNIA

ALGORITMO DE PROCESAMIENTO DE IMAGENES PARA SEÑALES DE TRANSITO

Lopez-Miguel (CITEDI-IPN) -1ro*, Rubio-Yoshio (CITEDI-IPN) -2do*, Sánchez-Moisés (CITEDI-IPN) – 3ro*, Orozco-Rosas - Ulises (CETYS Universidad) – 4to**

*Instituto Politécnico Nacional (CITEDI-IPN), Departamento de cómputo inteligente de alto rendimiento, Instituto Politécnico Nacional 1310, Nueva Tijuana, Tijuana, B.C.

{milopez, rrubio, msanchez1@citedi.mx}

** CETYS Universidad, Centro de Innovación y Diseño (CEID), Av. CETYS Universidad No. 4. El Lago, Tijuana B.C. ulises.orozco@cetys.mx

Resumen

Este trabajo aborda el problema de clasificar las señales de tránsito relacionadas con límites de velocidad. Se emplea un algoritmo de preprocesamiento para adecuar y mejorar las características relevantes de la imagen, para este caso, los números que representan los límites de velocidad. El preprocesamiento inicia realizando un escalamiento de la imagen para ajustar al tamaño que se requiere en la entrada de la red neuronal convolucional. Además, se mejora la propiedad del contraste empleando el algoritmo de CLAHE para ajustar esta propiedad, debido a que este parámetro influye al momento de clasificar las señales. También se mejoran las propiedades donde se encuentran los números en las señales de tránsito, resaltando los números como las características más

relevantes de la señal utilizando las operaciones morfológicas de erosión para eliminar ruido de la imagen y de dilatación para amplificar las propiedades del objeto de interés. En la etapa de procesamiento se utiliza una red neuronal convolucional para la clasificación de las señales. Al analizar los resultados obtenidos se observa una mejoría cuando se utiliza la etapa de preprocesamiento. Se puede concluir que el preprocesamiento de imágenes es una etapa fundamental para acondicionar las entradas para un clasificador.

Palabras Clave: Aprendizaje Automático, Aprendizaje Profundo, Redes Neuronales Convolucionales, CLAHE, Operación Morfológica.

1. Introducción

La visión por computadora se utiliza cada vez más en el campo del transporte inteligente, donde uno de los problemas a resolver es el reconocimiento de señales de tránsito. Este problema consta de distintas etapas como es la detección y clasificación de las señales para realizar alguna tarea en específico. Generalmente estos sistemas están basados en la detección de una región de interés (ROI, *Region of Interest*) de la señal de tránsito, basándose en las características de su color y forma geométrica. Existen algoritmos de preprocesamiento para mejorar la calidad de la imagen ajustando el contraste para obtener en la siguiente etapa una mejor eficiencia como la equalización del histograma (HE, *Histogram Equalization*) y la equalización adaptativa del histograma (AHE, *Adaptive Histogram Equalization*) [1]. Otro factor relevante para la etapa de preprocesamiento de imágenes es ajustar el contraste para mejorar las definiciones de bordes de cada región de una imagen y tener una mejor extracción de las características. La manipulación del ajuste del contraste de una imagen en forma no uniforme es un factor importante para que se obtenga una mejor detección de las características de las imágenes [1].

En las etapas de preprocesamiento el método adaptativo para el control del contraste de una imagen (CLAHE, *Contrast Limited Adaptive Histogram*) tiene gran relevancia para mejorar la extracción de características. Otros métodos que considerar son los operadores de morfología como son la erosión y la dilatación, para disminuir el ruido en el caso de la erosión y aumentar el número de píxeles de la imagen con el operador de dilatación [2].

2. Metodología

El presente trabajo se divide en dos etapas: preprocesamiento y procesamiento de las imágenes de señales de tránsito. En la etapa de preprocesamiento, se selecciona una imagen para dividirla a la mitad de forma manual, separando los números de la señal de límite de velocidad, de esta manera se ingresa uno de los números de la señal de tránsito a la red. Antes de ingresar el número se hace un ajuste del tamaño a cada número utilizando un escalamiento y adecuarlo a la entrada de la red, también se mejora el contraste y se amplifican las características. En la etapa de procesamiento se

utiliza una red neuronal convolucional (CNN, *Convolutional Neural Network*) preentrenada con la base de datos MNIST (*Modified National Institute of Standards and Technology database*). Las imágenes de entrada son de 28x28 píxeles y son puestas sobre una máscara de 32x32 píxeles para evitar que la red omita los bordes de los dígitos en la etapa de entrenamiento. Los resultados se evalúan utilizando una matriz de confusión para conocer el número de aciertos en la predicción de la red.

2.1 Algoritmo de preprocesamiento

El algoritmo se enfoca principalmente en el mejoramiento del contraste y la amplificación de las características de la imagen, para obtener una mejor calidad de la imagen original.



Figura 1. Imagen Original.

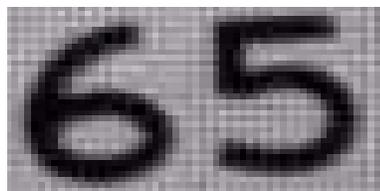


Figura 2. Segmentación manual.

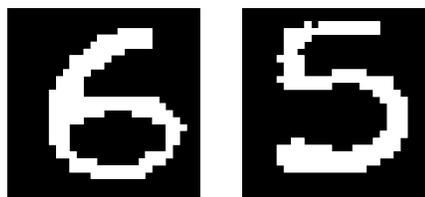


Figura 3. Imagen preprocesada.

Como se muestra en las Figuras 1 a la 3, se realiza una segmentación manual de la imagen original, pero al momento de realizar el enfoque en la región de interés, la resolución de la imagen que se obtiene es limitada de tal forma que el algoritmo de preprocesamiento mejora las propiedades, para no perder las características principales de la imagen.

2.1.2 Ecuación adaptativa limitada del histograma (CLAHE)

Alguno de los problemas al trabajar con imágenes es su visibilidad, la cual es dependiente de cómo se realizó la adquisición de la imagen, para mejorar la imagen se utiliza la ecualización del histograma adaptativo (AHE, *Adaptive Histogram Equalization*), el problema es que el método realiza una ecualización de histograma normal diferente, en la cual se toman distintas regiones de la imagen y en éstas se redistribuye un valor de contraste, en cambio cuando se emplea CLAHE, la distribución es utilizada para definir la forma del histograma que produce el resultado de mejor calidad.

Este algoritmo se encarga de mejorar el contraste de la imagen y la precisión de extracción, comparando el histograma local de la imagen, y redistribuyendo el contraste para cambiar el contraste de la imagen y obtener más detalles en el contraste. En comparación con la ecualización del histograma adaptativo (AHE), CLAHE puede suprimir eficazmente la amplificación de ruido y acelerar el funcionamiento con el algoritmo de interpolación.

El algoritmo CLAHE es el siguiente:

1. Adquisición de la imagen borrosa.
2. Obtener los valores de entrada como son, el número de regiones en fila y columna por separado, el rango dinámico (número de bandejas usadas en la función de transformación del histograma), límite del clip, tipo de parámetros de distribución.
3. La imagen original es dividida en varias regiones para las entradas preprocesadas.
4. Proceso aplicado sobre la región de interés (mosaico).
5. Genera mapeo de nivel de grises e histograma recortado. En la región de interés, los números de píxeles se dividen por igual en cada nivel de gris, por lo que el número promedio de píxeles es el nivel de gris que se describe a continuación:

$$N_{avg} = \frac{N_{CR-Xp} * N_{CR-Yp}}{N_{gray}} \quad (1)$$

Donde

N_{gray} = Número de nivel de gris en la región de interés

N_{CR-Xp} = Número de píxeles en la dirección X de la región de interés

N_{CR-Yp} = Número de píxeles en la dirección Y de la región de interés

Después de lo anterior, calcular el límite actual del clip.

$$N_{CL} = N_{CLIP} * N_{avg} \quad (2)$$

6. Se hace una interpolación mediante un mapa de escala de grises para generar una imagen mejorada.

De tal forma que este método modifica el contraste de la imagen original para realizar una mejor clasificación en la red neuronal convolucional [3].

2.1.3 Operaciones Morfológicas

Las operaciones morfológicas aplican elementos de estructuración a una imagen de entrada, creando una imagen de salida del mismo tamaño. Independientemente del tamaño del elemento estructurado, el origen está ubicado en su centro. Existen distintos operadores útiles definidos en morfología matemática, como son dilatación, erosión, apertura y cierre. Las transformaciones morfológicas básicas en la morfología matemática son dilatación morfológica y erosión mediante elementos estructurantes.

Estas transformaciones básicas están dadas por

$$\delta_{\lambda B}(f)(x) = (f \oplus \lambda B)(x) = \vee \{f(x - y) : y \in \lambda B\}$$

$$\varepsilon_{\lambda B}(f)(x) = (f \ominus \lambda B)(x) = \wedge \{f(x - y) : y \in \lambda B\} \quad (3)$$

donde B es el elemento estructurado elemental (3x3 píxeles en este trabajo) y λ es un parámetro de transformación [4].

2.1.4 Red Neuronal Convolucional (CNN)

Una de las técnicas de clasificación del aprendizaje automático es el aprendizaje profundo (*Deep Learning*), el cual utiliza la arquitectura de las redes neuronales, debido a esto se les denomina redes neuronales profundas, donde la profundidad hace referencia al número de capas ocultas en la red neuronal.

Uno de los tipos más populares de redes neuronales profundas son las conocidas como redes neuronales convolucionales (CNN o ConvNet). Una CNN convolucionada las características aprendidas con los datos de entrada y emplea capas convolucionales 2D, lo cual hace que esta arquitectura resulte adecuada para procesar datos 2D, tales como imágenes.

Las redes convolucionales combinan tres ideas de arquitecturas para asegurar algún grado de desplazamiento, escala y distorsión de invarianza.

1. Campos locales receptivos
2. Pesos compartidos (o replicación de pesos)
3. Submuestreo espacial o temporal.

2.1.4.1 LeNet-5

La arquitectura de la red neuronal convolucional LeNet-5 comprende 7 capas sin contar la entrada, pero en este trabajo se empleó una LeNet-5 modificada, la cual tiene 9 capas.

En LeNet-5 el conjunto de centros de los campos receptivos de la última capa convolucional forma un área 20x20 en el centro de la imagen de entrada de 32x32. Los valores de los píxeles de entrada son normalizados, ya que el nivel de fondo (blanco) corresponde a un valor de -0.1 y el primer plano (negro) corresponde a 1.175.

Esto hace que la media de la entrada rigurosamente sea 0 y la varianza 1, lo cual acelera el aprendizaje [5], [6], [7].

2.1.5 Base de datos MNIST

En la Figura 4 se muestra un ejemplo de imágenes de la base de datos MNIST. En nuestro trabajo, los datos se obtuvieron del conjunto de datos MNIST que consiste en diferentes imágenes de tamaño (28 x 28) píxeles. Entonces la capa de entrada contiene 784 neuronas de entrada. Los píxeles se guardan como caracteres sin signo (1 byte) y se consideran valores de 0 a 255. Los primeros 28x28 bytes de los datos coinciden con el primer entrenamiento ejemplo, los siguientes 28x28 bytes se comparan con el siguiente ejemplo y así. En la siguiente figura se muestran las imágenes utilizadas.

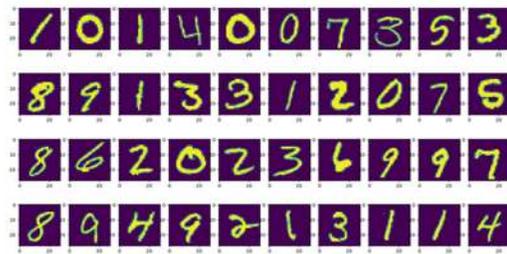


Figura 4. Ejemplo de imágenes muestra de la base de datos MNIST.

El entrenamiento se realizó previamente y el conjunto de prueba fue fácil, distinguible y diferente, que era necesario para propósito de ayudar a la red neuronal a aprender del conjunto de datos solamente. Por lo tanto, en lugar de seguir el camino natural de memorizar todo el conjunto de datos y luego reciprocando el mismo, en realidad se predice el conjunto de prueba [8].

3. Resultados

A continuación, en la Figura 5 se presenta la matriz de confusión del conjunto de imágenes para validar la respuesta de la red. Se pueden observar los valores esperados contra los valores obtenidos referente a la salida de la red neuronal convolucional.

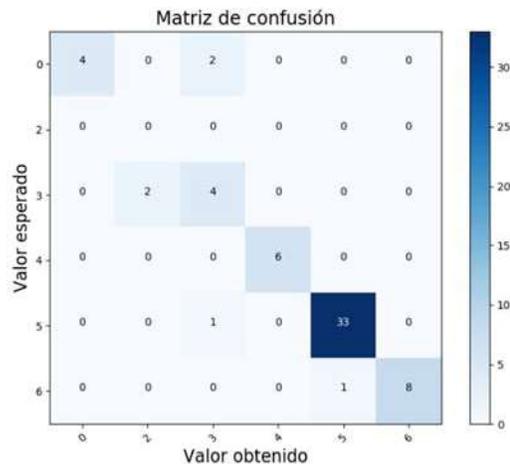


Figura 5. Matriz de confusión: Tabla del valor esperado con respecto al valor obtenido.

Como se aprecia en la matriz de confusión anterior, se obtuvo una mejor exactitud en la mayoría de los números clasificados, el número de muestras utilizado es de 64 imágenes, las cuales fueron proporcionadas de la base de datos Laboratory for Intelligent & Safe Automobiles (LISA) [9].

Se utilizaron las siguientes señales de límite de velocidad 35, 40, 45, 55 y 65 para la clasificación de cada número, los cuales fueron 0, 3, 4, 5 y 6 obteniendo los resultados mostrados en la tabla 1.

Tabla 1. Tabla de resultados

Número	Muestras	Aciertos	Exactitud
0	6	4	66.66%
3	6	4	66.66%
4	6	6	100%
5	34	33	97.05%
6	12	8	66.66%

Como se puede apreciar se logró un 85.93% de aciertos en promedio total, de acuerdo con las pruebas realizadas el algoritmo propuesto de preprocesamiento ayuda a mejorar la respuesta de la red neuronal convolucional.

4. Conclusiones

Se propuso un algoritmo de preprocesamiento para adecuar la imagen de entrada a la red neuronal convolucional. Los siguientes puntos se consideran como los más relevantes:

- La red neuronal convolucional fue entrenada con cierto patrón de características. Al realizar pruebas sin ajustar el contraste ni utilizar la dilatación, la red no respondía como se esperaba, y al momento de efectuar el preprocesado la respuesta de la red aumento considerablemente.
- Al momento de realizar la operación de escalamiento se puede reemplazar una de las capas de la red neuronal convolucional dedicada a esta.
- Una red neuronal convolucional está diseñada para la clasificación de imágenes y necesita un entrenamiento exhaustivo con diferentes tipos de patrones, de tal forma que se requiere de grandes cantidades de muestras y se requiere la utilización de GPU.

El algoritmo de preprocesamiento es de gran ayuda cuando se utiliza una técnica de clasificación de aprendizaje automático, al realizar una mejoría del contraste y amplificación de las características de la imagen se obtendrá una mejor respuesta en el procesamiento de la red neuronal convolucional.

5. Agradecimientos

Se agradece al Instituto Politécnico Nacional (IPN) y al Consejo Nacional de Ciencia y Tecnología por el apoyo de la beca de estudio para el programa de Maestría en Ciencias en Sistemas Digitales del CITEDI-IPN.

Referencias

1. Laguna, R., Barrientos, R., Blázquez, L. F., & Miguel, L. J. (2014). Traffic sign recognition application based on image processing techniques. *IFAC Proceedings Volumes*, 47(3), 104-109. doi:10.3182/20140824-6-za-1003.00693
2. Adatrao, S., & Mittal, M. (2016). An analysis of different image preprocessing techniques for determining the centroids of circular marks using hough transform. *2016 2nd International Conference on Frontiers of Signal Processing (ICFSP)*. doi:10.1109/icfsp.2016.7802966
3. Yadav, G., Maheshwari, S., & Agarwal, A. (2014). Contrast limited adaptive histogram equalization based enhancement for real time video system. *2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. doi:10.1109/icacci.2014.6968381
4. Terol-Villalobos, I. R. (2001). Morphological image enhancement and segmentation. *Advances in Imaging and Electron Physics*, 207-273. doi:10.1016/s1076-5670(01)80106-3
5. Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324. doi:10.1109/5.726791
6. Ide, H., & Kurita, T. (2017). Improvement of learning for CNN with ReLU activation by sparse regularization. *2017 International Joint Conference on Neural Networks (IJCNN)*. doi:10.1109/ijcnn.2017.7966185

7. Coskun, M., Ucar, A., Yildirim, O., & Demir, Y. (2017). Face recognition based on convolutional neural network. 2017 International Conference on Modern Electrical and Energy Systems (MEES). doi:10.1109/mees.2017.8248937
8. Makkar, T., Kumar, Y., Dubey, A. K., Rocha, A., & Goyal, A. (2017). Analogizing time complexity of KNN and CNN in recognizing handwritten digits. 2017 Fourth International Conference on Image Information Processing (ICIIP). doi:10.1109/iciip.2017.8313707
9. Mogelmoose, A., Trivedi, M. M., & Moeslund, T. B. (2012). Vision-Based Traffic Sign Detection and Analysis for Intelligent Driver Assistance Systems: Perspectives and Survey. *IEEE Transactions on Intelligent Transportation Systems*, 13(4), 1484-1497. doi:10.1109/tits.2012.2209421