

REDES NEURONALES ARTIFICIALES APLICADAS AL PROCESAMIENTO DE VIDEO

Juan Sebastián Valencia Villa ¹, Mónica Ayde Vallejo Velásquez ²

¹Ingeniero de Control, Escuela de Mecatrónica, Universidad Nacional de Colombia Sede Medellín, jsvalenc@unal.edu.co. Carrera 80 Número 65 – 223.

²Magister en Ingeniería, Profesor Auxiliar, Escuela de Mecatrónica, Universidad Nacional de Colombia Sede Medellín, mavallejov@unal.edu.co. Carrera 80 Número 65 – 223.

RESUMEN

La estimación de movimiento es un bloque esencial en la etapa de predicción del estándar de codificación de video H.264/AVC, el cual es determinante para obtener una tasa de compresión efectiva. Diferentes algoritmos han sido propuestos en la literatura con el fin de optimizar la arquitectura de este bloque, ya que éste demanda un prolongado tiempo de ejecución y un significativo número de recursos computacionales. Las estrategias tradicionales basadas en la búsqueda bloque a bloque aunque garantizan alta calidad no son apropiadas en cuanto a la eficiencia computacional; otros métodos basados en búsqueda por regiones aligeran la carga de procesamiento, pero no garantizan una buena calidad. En este trabajo se implementa una red neuronal tipo perceptrón multicapa que utiliza un algoritmo de aprendizaje tipo *backpropagation* supervisado que permite realizar la estimación de movimiento.

Palabras clave: Estándar de codificación, Estimación de Movimiento, Redes Neuronales Artificiales.

Recibido: 31 de enero de 2012. Aceptado: 18 de mayo de 2012

Received: January 31th, 2012. Accepted: May 18th, 2012

ARTIFICIAL NEURAL NETWORKS APPLIED TO VIDEO PROCESSING

ABSTRACT

The motion estimation is an essential block in the prediction stage of the H.264/AVC video compression standard, which is crucial for obtaining an effective coding rate. Different algorithms have been proposed in the literature to optimize the architecture of this block, since it requires a long execution time and a significant number of computational resources. Traditional strategies based on the search on each block allow ensuring high quality but are inappropriate in terms of computational efficiency; other methods based on search by regions reduce the processing load, but do not guarantee a good quality. In this paper it is implemented a multilayer perceptron neural network with supervised backpropagation learning algorithm which allows making the estimation.

Keywords: *Encoding Standard, Motion Estimation, Artificial Neural Networks.*

1. INTRODUCCIÓN

La estimación de movimiento es un algoritmo esencial en la etapa de predicción de un codificador de video ya que éste afecta directamente la velocidad de compresión y la calidad de la imagen [1]. Sin embargo, éste es uno de los bloques más intensivos a nivel computacional, representando en el estándar H.264/AVC más del 60% de todo el conjunto de cálculos [2] [3].

Los algoritmos de estimación de movimiento tienen como función principal predecir el movimiento de una trama de video a través de *frames* previamente codificados con el fin de reducir la cantidad de datos del video codificado [4]. La coincidencia se evalúa según un criterio de decisión, el cual es definido por un método de estimación de error.

Diferentes aproximaciones han sido propuestas en la literatura para el algoritmo de estimación de movimiento siendo hasta ahora los algoritmos de coincidencia de bloque (BMA) la técnica más exitosa. Los métodos de ajuste de bloques basados en el algoritmo BMA, tales como *full-search (FS)* y *fast full-search (FFS)*, proporciona una solución óptima mediante la búsqueda exhaustiva de todos los candidatos dentro de la ventana de búsqueda. Sin embargo, su alta complejidad computacional hace difícil la ejecución en tiempo real [5]. Otros métodos basados en búsquedas rápidas por regiones tales como: *Three Step Search (TSS)* [6], *Four Step Search (FSS)* [7], *diamond search (DS)* [8], *Two Dimensional Logarithmic Search (TDL)* [9], permiten reducir la complejidad computacional mientras brindan una calidad de video cercana a la original y una velocidad de procesamiento más alta, sin embargo pueden caer fácilmente en óptimos locales y en consecuencia perder desempeño para secuencias de video que presenten grandes movimientos [10].

Otros estudios revelan el uso de técnicas basadas en inteligencia artificial para la implementación de algoritmos de estimación de movimiento [11] [12] [13]. Por su potencial para reconocer patrones de información, las redes neuronales artificiales, son una herramienta apropiada para resolver problemas relacionados con el procesamiento de imágenes.

En este artículo se presenta el uso de una red neuronal tipo perceptrón multicapa que utiliza un algoritmo de aprendizaje tipo *backpropagation* supervisado, como método de predicción para la

implementación del bloque de estimación de movimiento. Dichas redes han sido elegidas gracias a su habilidad natural para el procesamiento paralelo de la información basado en células nerviosas y a su capacidad de deducir las propiedades emergentes para resolver problemas complejos [14].

El artículo está organizado de la siguiente forma: en la sección 2 se describe en forma general el algoritmo de estimación de movimiento. En la sección 3 se presenta el algoritmo propuesto basado en la definición de una Red Neuronal. En la sección 4 se presentan los resultados de simulación. Finalmente las conclusiones y trabajos futuros son presentadas en la sección 5.

2. MÓDULO DE ESTIMACIÓN DE MOVIMIENTO

Los algoritmos de estimación se basan en realizar una división de cada una de las imágenes de una secuencia en bloques para detectar el movimiento en éstas con respecto a los bloques que las constituyen. Para esto, se realizan una serie de comparaciones entre cada uno de los bloques de la imagen previa con el mismo bloque en la imagen actual y se calcula la función de coincidencia entre ellos. Estas comparaciones se realizan determinando una región de búsqueda y encontrando el bloque para el cual el error es mínimo. El vector de movimiento (MV) estimado sería el desplazamiento entre ambos bloques [1] [15].

DIAGRAMA DE BLOQUES DE ESTIMACIÓN DE MOVIMIENTO

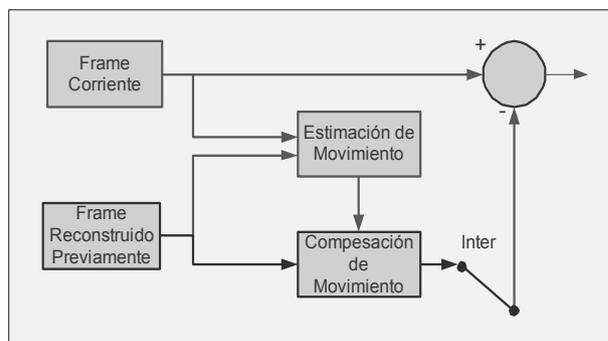


Fig.1. Diagrama de bloques de la etapa de estimación de movimiento [16]

La coincidencia entre bloques de imágenes secuenciales se evalúa según un criterio de decisión, el cual es definido por un método de

estimación de error como por ejemplo: el método de Error Cuadrático Medio (MSE), el método del Error Absoluto Medio (MAE), Suma de Diferencias Absolutas (SAD), siendo éste último método el más usado en este tipo de algoritmos [1].

3. IMPLEMENTACIÓN

Una red neuronal está compuesta por unidades básicas de procesamiento operando en paralelo denominados neuronas. La función de la red es determinada por las conexiones entre las neuronas. En consecuencia, una red puede ser entrenada para realizar una función específica ajustando los valores de las conexiones (pesos) entre las neuronas [17].

En este trabajo es usada una red neuronal tipo perceptrón multicapa con algoritmo de aprendizaje tipo *backpropagation* supervisado, la cual es ampliamente estudiada en la literatura. En la implementación más sencilla de aprendizaje *backpropagation*, los pesos y los umbrales de la red son actualizados con base en el gradiente descendente. La modificación de la función de error es de acuerdo a la pendiente de pesos de la red neuronal, y los pesos iniciales se generan al azar [14].

En particular, la red neuronal usada tiene tres capas distribuidas así:

- La capa de entrada consistente de 176 neuronas que representan el número de píxeles del vector columna de cada *frame* de la trama de video.
- Una capa oculta dividida en cuatro subcapas distribuidas en 528 neuronas, 264 neuronas, 132 neuronas y 66 neuronas respectivamente y con funciones de activación tangente inversa sigmoidea y logarítmica sigmoidea.
- Una capa de salida consistente de 144 neuronas con función de activación lineal pura que representan el residual de predicción.

Para el entrenamiento de la red, se utilizó el algoritmo *trainrp* del *toolbox* de redes neuronales de Matlab y los siguientes parámetros de entrenamiento:

- Iteraciones: 1000
- Meta: 0.0001
- Tasa de Aprendizaje: 0.001

- Mínimo Valor de Gradiente Descendiente: 1×10^{-20}

En las redes neuronales multicapas, no se pueden dar reglas concretas para determinar el número de neuronas o el número de capas de una red para resolver un problema concreto [18]. La arquitectura de la red propuesta se obtuvo a través de pruebas con diferentes tipos y fue seleccionada ya que presentó el menor error en la etapa de entrenamiento.

La red neuronal propuesta efectúa el cálculo de la función de error, modifica la región de búsqueda y activa el motor de estimación en una unidad integrada. El funcionamiento del algoritmo, se describe a continuación y se muestra gráficamente en la figura 2.

a) Se almacena una imagen residual de la trama como objeto de comparación u objetivo de la red neuronal, para que a partir de esta se modifiquen los pesos de dicha red y se pueda predecir el residual del siguiente *frame* de la trama de video.

b) Cada *frame* de la trama de video, se separa por vectores columna, los cuales almacenaran la información de tiras verticales de la imagen que conforma el *frame*.

c) Cada uno de los vectores del *frame* se compara con cada uno de los vectores columna que conforman la imagen objetivo, es decir, el vector columna 1 que conforma el *frame* actual que se está prediciendo, se compara con el vector columna 1 de la imagen objetivo, con lo cual para cada iteración de la red a través de este vector, se modificarán los pesos y se logrará una aproximación de los residuales del siguiente *frame* de la trama. Este procedimiento ocurre con todas las columnas que conforman la imagen.

d) Cuando se tiene la red completamente entrenada para cada uno de los vectores que conforman la imagen, se implementa esta red utilizando el *frame* siguiente de la trama de video, dando como resultado el residual de predicción que va a la etapa de codificación del estándar.

e) La imagen objetivo se actualiza para el *frame* siguiente y comienza de nuevo el entrenamiento de la red, esto ocurre hasta finalizar con la *cantidad* de *frame* que hacen parte de la trama de video.

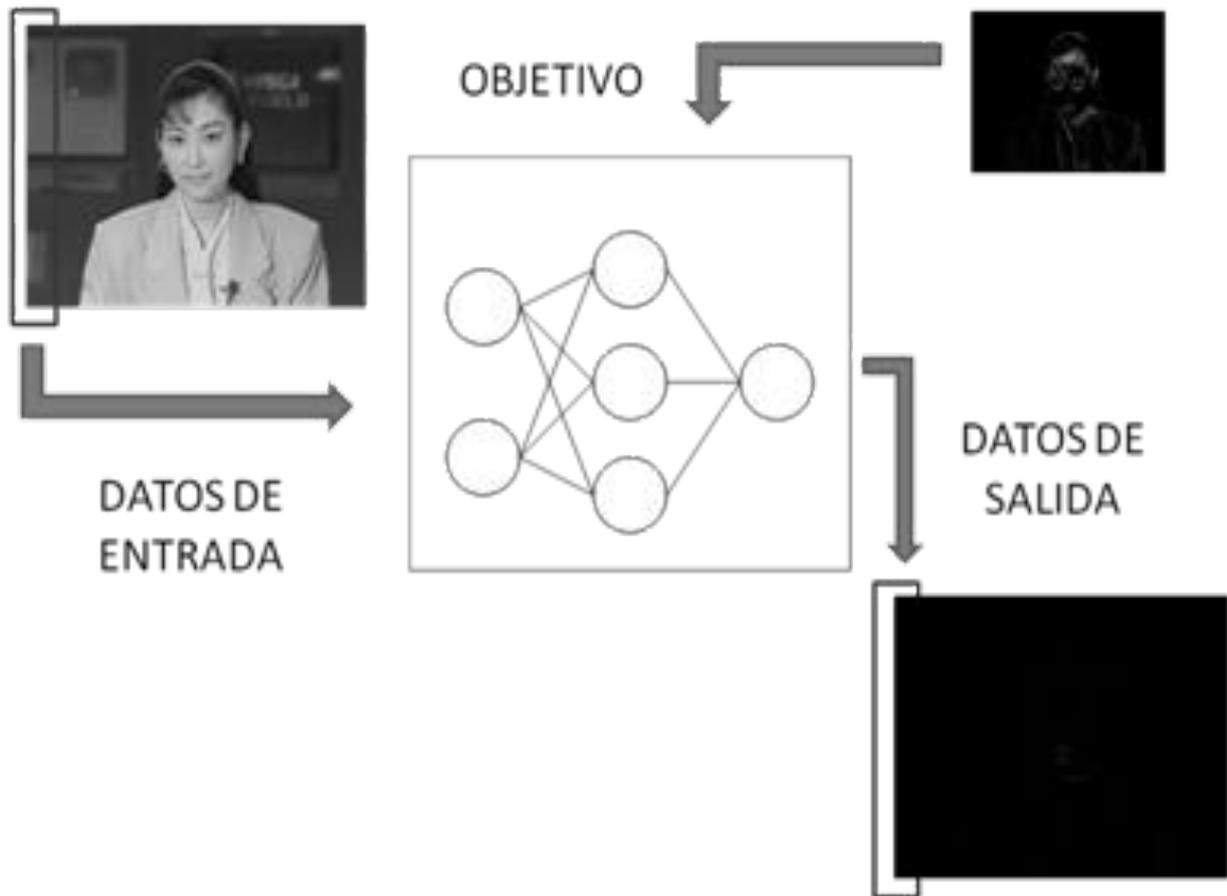


Fig.2. Funcionamiento de la red neuronal

4. RESULTADOS

Para evaluar la funcionalidad de la red neuronal para generar el *frame* residual y de esta forma poder determinar su efectividad como motor de búsqueda dentro del bloque de estimación de movimiento, se utilizaron dos secuencias de video en formato QCIF (144 X 176). La secuencia *akiyo* muestra una presentadora de noticias sentada frente de un escenario estático con lo cual hay poco movimiento. La secuencia *foreman*, es tomada con una cámara de mano muy inestable y muestra a un hombre hablando con expresiones faciales muy animadas y detalladas. De ambas secuencias se extrajeron cinco *frames* consecutivos, cuatro *frames* para entrenamiento y un *frame* de prueba, esto debido a la complejidad que conlleva manejar la gran cantidad de datos que involucra las aplicaciones de procesamiento de imágenes y video [19] [20].

La secuencia de video *akiyo*, mostrada en la figura 3 es de poco movimiento, por lo cual, se evidencian pocas líneas en los *frame* que constituyen los residuales de la trama de video que se quiere codificar.

La secuencia de video *foreman*, que se muestra en la figura 4, tiene mayor movimiento, por lo cual se pueden evidenciar una mayor cantidad de líneas en los residuales.

Para ambos casos, solo es comprobado que el objetivo deseado (*frame* residual) refleje los cambios de movimiento en relación con las características particulares de cada secuencia de video analizada. Los *frames* residuales son comparados en forma subjetiva, a través de observaciones visuales con *frames* residuales obtenidos mediante otras técnicas para las mismas secuencias de video en la literatura [21 – 23].



Fig.3. Secuencia 1 "AKIYO"



Fig.4. Secuencia 2 "FOREMAN"

5. CONCLUSIONES Y TRABAJO FUTUROS

Aunque existen modelos convencionales que tienen menor consumo computacional que el modelo de redes neuronales, la aproximación presentada en este artículo es funcional y puede ser una opción a la hora de implementar la etapa de estimación de movimiento de una aplicación especial.

Como trabajo futuro se planea evaluar nuestra propuesta de forma más rigurosa, usando más muestras de entrenamiento y pruebas, y mediante la aplicación de diferentes métricas que permitan comparar en forma objetiva los resultados frente a los obtenidos con otros métodos con el propósito de garantizar que esta técnica es apropiada para usar como motor de búsqueda dentro del bloque de estimación de movimiento. Por otra parte también se evaluará la factibilidad de implementar esta red neuronal en hardware reconfigurable como las FPGAs que admiten procesamiento paralelo.

6. AGRADECIMIENTOS

Los autores agradecen a la Dirección de Investigación DIME y la Vicerrectoría Nacional de Investigación de la Universidad Nacional de Colombia por el apoyo brindado para llevar a cabo este trabajo mediante la financiación del proyecto con código 20201007832.

7. REFERENCIAS

[1] I.E.G. Richardson, H.264 and MPEG-4 Video Compression: Video Coding for Next-generation, Wiley, 2003.

[2] Junwoo Lee; Minsu Choi; Youngmin Cho; Jinsang Kim; Won-Kyung Cho; , "Fast H.264/AVC motion estimation algorithm using adaptive search range," Integrated Circuits, ISIC '09. Proceedings of the 2009 12th International Symposium on, vol., no., pp.336-339, 14-16, Dec. 2009

[3] J.S. Valencia, "Descripción e implementación del bloque de estimación de movimiento para el estándar de codificación de video H.264 usando redes neuronales artificiales". Proyecto Fin de Carrera, Universidad Nacional de Colombia, sede Medellín, 2010.

[4] Kalaycioglu, C.; Ulusel, O.C.; Hamzaoglu, I., "Low power techniques for motion estimation hardware," Field Programmable Logic and Applications, 2009. FPL 2009. International Conference on, vol., no., pp.180-185, Aug. 31 2009-Sept. 2 2009.

[5] Shenolikar, P.C.; Narote, S.P., "Different approaches for motion estimation," Control, Automation, Communication and Energy Conservation, 2009. INCACEC 2009. 2009 International Conference on, vol., no., pp.1-4, 4-6 June 2009.

[6] L.-P. Chau and X. Jing, "Efficient Three-Step Search Algorithm for Block Motion Estimation in Video Coding," Proc. IEEE Int'l Conf. Acoustics, Speech, and Signal Processing 2003 (ICASSP '03), pp. 421- 424, Apr. 2003.

[7] Duanmu, C.J., "Fast Scheme for the Four-Step Search Algorithm in Video Coding," Systems, Man and Cybernetics, 2006. SMC '06. IEEE International Conference on, vol.4, no., pp.3181-3185, 8-11 Oct. 2006.

[8] Jo Yew Tham; Ranganath, S.; Ranganath, M.; Kassim, A.A.; "A novel unrestricted center-biased diamond search algorithm for block motion estimation," Circuits and Systems for Video Technology, IEEE Transactions on , vol.8, no.4, pp.369-377, Aug 1998.

[9] Jain, J.; Jain, A.; , "Displacement Measurement and Its Application in Interframe Image Coding," Communications, IEEE Transactions on , vol.29, no.12, pp. 1799- 1808, Dec 1981.

[10] Zhenyu Liu; Junwei Zhou; Goto, S.; Ikenaga, T., "Motion Estimation Optimization for H.264/AVC Using Source Image Edge Features," Circuits and Systems for Video Technology, IEEE Transactions on , vol.19, no.8, pp.1095-1107, Aug. 2009.

[11] Michaelis. B., Schnelting. O., Seiffert. U., Mecke. R.: Motion Estimation Using a Compounded Self Organizing Map-Multi Layer Perceptron. WCNN '95. World Congress on Neural Networks. Washington USA, July 1995.

[12] Dae-Hyun Ryu; , "Block matching algorithm using neural network," TENCON '97. IEEE Region 10 Annual Conference. Speech and Image

Technologies for Computing and Telecommunications, Proceedings of IEEE, vol.1, no., pp.379-382 vol.1, 4-4 Dec. 1997.

[13] Ryu D.H., Kim C.R., Kim S.W., Choi T.W., Kim J.C., Motion Search Region Prediction Using Neural Network Vector Quantization. Signal Processing, 3rd International Conference on, 14-18 Oct 1996. 1473 - 1476 vol.2. 1996.

[14] Berkane, M.; Clarysse, P.; Magnin, I.E.; "A Neural Network-Based Approach to Motion Estimation with Discontinuities," Hybrid Intelligent Systems, 2008. HIS '08. Eighth International Conference on, vol., no., pp.356-361, 10-12 Sept. 2008.

[15] P.C. Benito. Integración de información de movimiento en la segmentación de secuencias de vídeo basada en el modelado de fondo. Proyecto Fin de Carrera. Universidad Autónoma de Madrid. 2010.

[16] Wiegand T., Sullivan G.J., Bjontegaard G., Luthra A. Overview of the H.264/AVC Video Coding Standard. IEEE, Transaction on Circuits and System for Video Technology, Vol.13, no7, pp 560-573. July 2003.

[17] Flórez R., Fernández J. Las Redes Neuronales Artificiales, Fundamentos Teóricos y Aplicaciones Prácticas. Netbiblo S.L., ISBN: 9788497452465. 2008.

[18] L.K Toscano. Reconocimiento dinámico y estático de trazos. Tesis de doctorado. Escuela Superior de Ingeniería Mecánica y Eléctrica Unidad Culhuacan, México 2005.

[19] C.V. Gámez Jiménez. Diseño y Desarrollo de un sistema de Reconocimiento de Caras. Proyecto Fin de Carrera de Ingeniera de Telecomunicaciones. Universidad Carlo III, Madrid. 2009.

[20] Adnan Khashman. Automatic Detection, Extraction and Recognition of Moving Objects. International Journal Of Systems Applications, Engineering & Development Issue 1, Volume 2, pp 43-51, 2008.

[21] Eckert, M. Compensación de movimiento avanzada para codificación de video. Tesis de doctorado. Universidad Politécnica de Madrid. 2003

[22] Fang Zhijun, Xu Shenghua, Wu Shiqian, Wang Zhengyou, Zeng Weiming. SVD Based Multi-Watermark Method in Multiwavelet Domain, Proceedings of the International Computer Conference 2006 on Wavelet Active Media Technology and Information Processing. Chongqing, China, pp.807-812.

[23] Y. Wang, J. Ostermann y YQ. Zhang, Video Processing and Communications. Prentice Hall, 2002.