



HAL
open science

Neuro-inspired Compression of RGB Images

Effrosyni Doutsis, George Tzagkarakis, Panagiotis Tsakalides

► **To cite this version:**

Effrosyni Doutsis, George Tzagkarakis, Panagiotis Tsakalides. Neuro-inspired Compression of RGB Images. EUSIPCO (27th European Signal Processing Conference), Sep 2019, Coruña, Spain. hal-02146564

HAL Id: hal-02146564

<https://hal.science/hal-02146564>

Submitted on 8 Jan 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Neuro-Inspired Compression of RGB Images

Effrosyni Doutsis
Institute of Computer Science
FORTH
Heraklion, Greece
edoutsis@ics.forth.gr

George Tzagkarakis
Institute of Computer Science
FORTH
Heraklion, Greece
gtzag@ics.forth.gr

Panagiotis Tsakalides
Institute of Computer Science
FORTH
and
Computer Science Department
University of Crete
Heraklion, Greece
tsakalid@ics.forth.gr

Abstract—During the last decade, there is an ever increasing interest about the decryption and analysis of the human visual system, which offers an intelligent mechanism for capturing and transforming the visual stimulus into a very dense and informative code of spikes. The compression capacity of the visual system is beyond the latest image and video compression standards, motivating the image processing community to investigate whether a neuro-inspired system, that performs according to the visual system, could outperform the state-of-the-art image compression methods. Inspired by neuroscience models, this paper proposes for a first time a neuro-inspired compression method for RGB images. Specifically, each color channel is processed by a retina-inspired filter combined with a compression scheme based on spikes. We demonstrate that, even for a very small number of bits per pixel (bpp), our proposed compression system is capable of extracting faithful and exact knowledge from the input scene, compared against the JPEG that generates strong artifacts. To evaluate the performance of the proposed algorithm we use Full-Reference (FR) and No-Reference (NR) Image Quality Assessments (IQA). We further validate the performance improvements by applying an edge detector on the decompressed images, illustrating that contour extraction is much more precise for the images compressed via our neuro-inspired algorithm.

Index Terms—Retina-inspired filter, Leaky Integrate-and-Fire model, spikes, FR-IQA, NR-IQA, edge detection.

I. INTRODUCTION

We live in an era when the amount of digital information is expanding at astonishing rates due to the Internet of Things (IoT). According to recent studies, 2.5 quintillion (10^{18}) bytes of data are produced every day. In this data deluge, images and video play a substantial role as they occur in different resolutions depending on the device that produces them. However, the resolution of an image or a video should be supported by efficient processing tools, such that the quality of the retrieved information after storage and/or transmission will remain high and free of visual artifacts. This necessitates the design of new, profitable and energy-efficient mechanisms, which enable the detection of the most informative part of the visual scene, while also extracting knowledge about the content semantics without any prior information.

Image and video compression algorithms are among the key processing tools for balancing between the assigned bits per

pixel (bpp) and the achieved reconstruction quality. During the last decades, significant efforts have been made for the development of efficient image and video compression standards. JPEG and JPEG2000 are the most widely used image compression algorithms, while H.264 and HEVC achieve state-of-the-art performance in video compression applications. The common drawback of all these systems is that the input signal is processed without any knowledge about the content of the visual scene. In addition, their embedded mechanisms, such as macroblock splitting, cause displeasing and distracting artifacts to the human perception, especially for low bpp values.

The human visual system (HVS) perception has been gaining an increasing interest as a powerful mechanism for improving the performance of the compression systems. The HVS enables the intelligent capturing and transformation of the visual information into a sequence of biphasic events, the so-called *spikes*, which constitute a highly compact, yet very informative and efficient, neural code able to be transmitted undamaged to the visual cortex, the part of the brain that processes the visual information. Recently, the structure and functionalities of the HVS have motivated the signal processing community to build *neuro-inspired systems* that mimic the brain. Among others, neuromorphic chips [1], bioinspired cameras [2], convolutional neural networks, virtual retina [3] and neuromimetic image and video compression mechanisms [4]–[8] are the most well-known brain-inspired systems.

This paper proposes a novel algorithm for neuro-inspired compression of RGB images, based on a recently introduced retina-inspired coding/decoding algorithm for grayscale images. In particular, a lossy compression algorithm is designed consisting of a retina-inspired filter [9] and a spike-based coding/decoding scheme [8]. The algorithm mimics the way the visual system captures and transforms the visual information along time into a sequence of spikes, which are then used to reconstruct the image. We compare the performance of the proposed system against the JPEG standard and we demonstrate that, for a given bpp value, our system is better capable of describing the details of the visual scene, while avoiding the block-effect artifacts, especially for a small bpp value. We use Full-Reference Image Quality Assessments (FR-IQA) like the Peak Signal to Noise Ratio (PSNR) and the Structural SIMilarity index (SSIM) [10], which are highly used

when the reference undistorted image is available. However, this is an ideal scenario since, in practice, neither the reference image is available nor any of its features. This is the case of the No-Reference Image Quality Assessments (NR-IQA) like BRISQUE [11]. Additionally, we employ the Sobel edge detection algorithm to illustrate the performance improvements of our method by showing that, for the same bpp, our compression system is much more efficient in detecting the objects' contours in a given image, in contrast to JPEG.

The rest of the paper is organized as follows: Section II overviews the retina-inspired filter and the neuro-inspired quantization scheme, which constitute the main processing tools of our proposed compression method, hereafter called NICE. Section III analyzes the architecture of NICE applied on RGB images, and describes the quality metrics to be used for the evaluation of its performance. Section IV presents a comparison between our proposed algorithm and the JPEG applied to RGB images, while Section V summarizes the main outcomes of this study and proposes directions for future extensions.

II. BASICS OF NEURO-INSPIRED COMPRESSION

A. Retina-inspired Filter

The retina-inspired filter (RIF) [9], which is motivated by retina-related tools [12], has been proposed to approximate precisely the dynamic properties of the outer plexiform layer (OPL) of retina. OPL is responsible for capturing and transforming the visual stimulus into electrical current. The RIF is a bunch of time-varying difference of Gaussian filters, the so-called weighted difference of Gaussians (WDoGs), defined by

$$\phi(x, t) = a(t)G_c(x) - b(t)G_s(x), \quad (1)$$

where $G_c(x)$, $G_s(x)$ are the center and surround Gaussian filter, respectively, and $a(t)$, $b(t)$ are two time-varying weights. Assuming a uniform temporal sampling, t_1, \dots, t_m , a different DoG filter $\phi_j(x) = \phi(x, t_j)$ is defined for each time instance t_j , $j = 1, \dots, m$. Let a temporally constant input defined by

$$f(x, t_j) = f(x) \mathbb{1}_{[0, T]}(t_j), \quad j = 1, \dots, m, \quad (2)$$

where $f(x) = \{f(x_1), \dots, f(x_n)\}$ is the input signal that consists of n pixels, $\mathbb{1}$ is the indicator function that equals 1, if $0 \leq t \leq T$, and 0 otherwise, and T is the time for which the signal is flashed. Then, as it is proven in [9], for each time instance the RIF yields

$$A_j(x) = A(x, t_j) = \phi_j(x) \overset{x}{*} f(x), \quad j = 1, \dots, m, \quad (3)$$

where $A_j(x)$ is the transformed signal at time j and $\overset{x}{*}$ denotes a spatial convolution. In addition, the RIF has been proven to be an invertible transform according to the frame theory [13], meaning that it perfectly reconstructs the input signal.

B. Spike-based Coding

The Leaky Integrate-and-Fire (LIF) model approximates a neuron by an electrical circuit [14]. Then, the pixels' intensity is determined by the current I that runs the circuit. Under the assumption that the input current is constant during a period of time, $I(t) = I \mathbb{1}_{[0 \leq t \leq T]}(t)$, each I can be described by the arrival delay, d , of the first spike as follows,

$$d = \begin{cases} +\infty, & RI < \theta, \\ h(RI; \theta) = -\tau_m \ln \left[1 - \frac{\theta}{RI} \right], & RI > \theta, \end{cases} \quad (4)$$

where R is the resistance, $\tau_m = RC$ is the leaky integrator term, C is the capacitor of the electrical circuit and θ is the membrane threshold of the neuron. Due to the high memory cost of the perfect LIF [15], it was proven in [8] that by encoding the number of spikes, N , during the observation window T , instead of the delays, provides enough information for recovering the input signal,

$$N = \begin{cases} 0, & d > T, \\ \left\lfloor \frac{T}{d} \right\rfloor, & d \leq T. \end{cases} \quad (5)$$

Doing so, it is possible to approximate the spike arrival delays $\tilde{d} = T/N$ and reconstruct the best possible values,

$$\tilde{I} = \begin{cases} 0, & N = 0, \\ h^{-1}(\tilde{d}; \theta) = h^{-1} \left(\frac{T}{N}; \theta \right), & N \neq 0. \end{cases} \quad (6)$$

III. PROPOSED NEURO-INSPIRED COMPRESSION METHOD FOR RGB IMAGES

The visual system is more sensitive to the luminance of light than the color. This is confirmed by the fact that the retina consists of two different kinds of photoreceptors; the rods, which are responsible for capturing light, and the cones, which are used to detect color, where the population of rod cells is much higher compared to the one of cone cells. As a consequence, the limited number of cones keeps within bounds their physiological and mathematical analysis, whereas the vast number of rods facilitated their study. The RIF filter is based on neuroscience models that approximate the shape and behavior of rod cells. Thus, when we apply the RIF filter to an RGB image, f , that consists of three different color channels (f_R , f_G and f_B , that correspond to the red, green and blue colors, respectively) we proceed by applying the RIF and the spike-based coding to each individual channel, since there are no available neuro-inspired models for approximating the color perception motivated by the cones in the retina of the eye.

Fig. 1 illustrates the proposed architecture, where each color channel is transformed by the RIF filter, whose output is a bunch of layers $A_j(x)$, $j = 1, \dots, m$. Let x_1, \dots, x_n be some spatial samples and $A_j(x_1), \dots, A_j(x_n)$ the associated pixels of the j -th output layer of the RIF, which is then fed to the *spike-based encoder*. For each pixel intensity of the j -th layer, the spike-based encoder calculates the corresponding number of emitted spikes via (5). By encoding and transmitting the

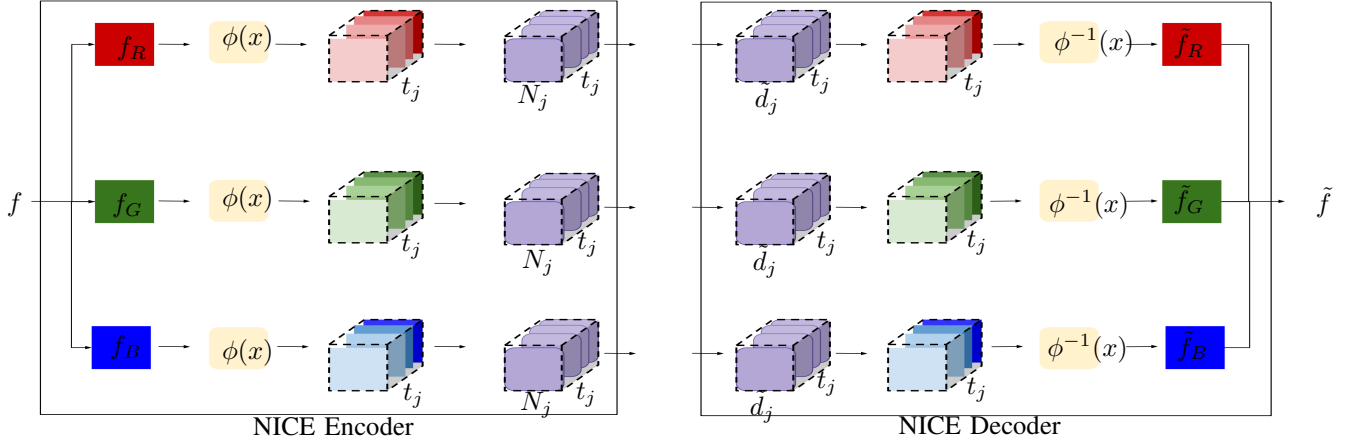


Fig. 1. Our proposed NICE compression system tailored to RGB images. Each channel is filtered by the RIF and then sampled by the spike-based encoder.

number of spikes to the *spike-based decoder*, it is possible to approximate the time delay each spike was emitted via the inverse function, h^{-1} . Then, a reconstruction is obtained for the intensity of each pixel, \tilde{A}_j , using (6). According to [13], the RIF constitutes a frame, thus we can apply the inverse RIF transform to the group of the quantized RIF layers in order to reconstruct an approximation of each color channel, \tilde{f}_R , \tilde{f}_G and \tilde{f}_B , and to this extend an approximation of the input RGB image, \tilde{f} .

A. Rate Calculation

To evaluate the performance of our proposed NICE compression system we employ the Shannon entropy, H , for calculating the number of bits per pixel,

$$H = \frac{1}{m} \sum_{j=1}^m H_j, \quad (7)$$

$$H_j = - \sum_{k=1}^{N_j} P(k) \log_2 P(k), \quad j = 1, \dots, m, \quad (8)$$

where P is the probability mass function of the number of spikes k that correspond to each layer. Throughout this paper, the entropy is given in bits per pixel.

B. Quality Metrics

Two famous FR-IQA metrics are used to evaluate the quality of the reconstructed image; the peak signal-to-noise ratio (PSNR) (10) and the structural similarity index (SSIM). For an RGB image, the PSNR is given by

$$\text{PSNR}(f, \tilde{f}) = \frac{1}{3} (\text{PSNR}_R + \text{PSNR}_G + \text{PSNR}_B), \quad (9)$$

where the PSNR of each color channel $c \in \{R, G, B\}$, is defined by

$$\text{PSNR}_c = 10 \log_{10} \frac{255^2}{\text{MSE}(f_c, \tilde{f}_c)}, \quad (10)$$

$$\text{MSE}(f_c, \tilde{f}_c) = \frac{1}{n} \sum_{i=1}^n \|f_c(x_i) - \tilde{f}_c(x_i)\|^2. \quad (11)$$

Since the proposed architecture is based on neuroscience models, we also employ the SSIM (12), which is considered to be a more accurate visual perception quality metric, defined by

$$\text{SSIM}(f, \tilde{f}) = \left(\frac{2\mu_f \mu_{\tilde{f}} + c_1}{\mu_f^2 + \mu_{\tilde{f}}^2 + c_1} \right) \left(\frac{2\sigma_f \sigma_{\tilde{f}} + c_2}{\sigma_f^2 + \sigma_{\tilde{f}}^2 + c_2} \right) \left(\frac{\sigma_{f, \tilde{f}} + c_3}{\sigma_f \sigma_{\tilde{f}} + c_3} \right), \quad (12)$$

where μ_f and $\mu_{\tilde{f}}$ denote the average of f and \tilde{f} , respectively, σ_f^2 and $\sigma_{\tilde{f}}^2$ are the variances of f and \tilde{f} , and $\sigma_{f, \tilde{f}}$ is the covariance of f and \tilde{f} . The constants $c_1 = k_1 L^2$, $c_2 = k_2 L^2$ and $c_3 = c_2/2$ are positive numbers used to stabilize the division with a weak denominator, where L is the dynamic range of the pixel values and $k_1 = 0.01$, $k_2 = 0.03$.

However, the performance of the FR-IQA metrics is questionable for several reasons. First, in real-life applications, it is impossible to provide the receiver with the ground truth signal. The only available information is the distorted image to be assessed. Furthermore, although widely used, the FR-IQA are inconsistent with the human visual perception. Thus, the blind reference-less image spatial quality evaluator (BRISQUE) [11] has been also used in this work to evaluate the capacity of the neuro-inspired compression architecture according to the natural scene statistics that appear in the spatial domain.

IV. EXPERIMENTAL EVALUATION

In this section we compare the performance of the proposed NICE system with the state-of-the-art JPEG standard, in terms of the rate-distortion tradeoff, for a set of images obtained from the ImageNet database. In addition, an edge detection algorithm is employed to demonstrate the fact that, when the entropy is very low (i.e., the distortion of the reconstructed image is high), the NICE system is capable of extracting the meaningful information content from an RGB image, which is reflected in the precise description of the contours in the input scene. MATLAB was used to generate the distinct JPEG

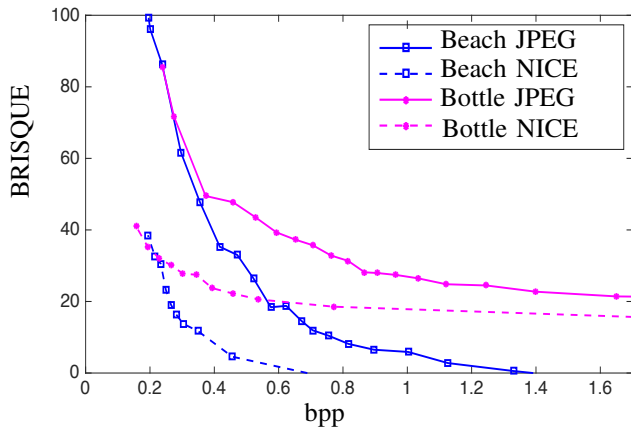


Fig. 2. Performance comparison of the proposed NICE system against JPEG. BRISQUE is used as a NR-IQA. The NICE scores are lower than the JPEG scores meaning that the perceptual quality is better in the proposed neuro-inspired compression architecture.

qualities with the built-in function `imwrite` which allows us to create and write `.jpeg` files of 100 different qualities. Subsequently, the calculation of the number of bits per pixel for each different quality is carried out using the built-in function `iminfo`.

Fig. 3 shows the reconstruction results for the proposed NICE system and the JPEG standard for low and high dynamic images. Although for similar bpp values the two quality FR-IQA metrics (PSNR and SSIM) are higher for JPEG (currently, the reconstructed images using NICE present a decrease in the dynamic range yielding darker images), however, it is generally admissible that such quality metrics are typically inconsistent with the human visual perception. For this reason, we compare in Fig. 2 the performance between the proposed NICE and the JPEG using the BRISQUE metric. BRISQUE quality scores are typically between $[0,100]$. The lower the score is, the better the perceptual quality of the image. In addition, special attention has been paid recently towards investigating new perceptual metrics capable of evaluating the visual quality of an image in accordance to the human visual system. Motivated by this, we employ an edge detection algorithm based on the Sobel filter as an indirect way to evaluate the performance of our proposed compression system.

Nevertheless, edge detection is a critical step in several tasks, such as segmentation, classification, etc. The edge detection results illustrate the superiority of NICE in better preserving the meaningful information content of an RGB image yielding an enhanced detection and description of the contours in the image. This is in contrast to JPEG, whose blocking artefacts effect, especially for very low bpp values, deteriorates the performance of the edge detector.

V. CONCLUSIONS

In this work, we proposed a neuro-inspired compression system tailored to RGB images. Specifically, a retina-inspired filter was employed, combined with spike-based compression mechanisms to design an integrated encoding/decoding sys-

tem. The performance of the proposed NICE compression method was compared against the JPEG standard, in terms of the rate-distortion tradeoff using both FR-IQA (PSNR and SSIM) and NR-IQA (BRISQUE) quality metrics. Furthermore, we suggested the use of edge detection (based on the Sobel filter in this study) as an indirect way to demonstrate the superiority of NICE in better describing the meaningful information content of RGB images, yielding increased edge detection accuracy, as opposed to JPEG, especially for low bpp values.

As a future work, we are interested in demonstrating the increased efficiency of the NICE system against the state-of-the-art compression standards in more complex tasks, such as image classification. According to the experimental results presented herein, we envision that our neuro-inspired compression scheme will yield an improved performance compared with the existing image compression standards, when coupled with pre-trained convolutional neural network architectures.

REFERENCES

- [1] C. S. Thakur, J. L. Molin, G. Cauwenberghs, G. Indiveri, K. Kumar, N. Qiao, J. Schemmel, R. Wang, E. Chicca, J. Olson H., J. Seo, S. Yu, Y. Cao, A. van Schaik, and R. Etienne-Cummings, "Large-scale neuromorphic spiking array processors: A quest to mimic the brain," *Frontiers in Neuroscience*, vol. 12, pp. 891, 2018.
- [2] C. Posch, T. Serrano-Gotarredona, B. Linares-Barranco, and T. Delbruck, "Retinomorph Event-Based Vision Sensors: Bioinspired Cameras With Spiking Output," *Proceedings of the IEEE*, vol. 102, no. 10, pp. 1470–1484, 2014.
- [3] A. Wohrer, P. Kornprobst, and M. Antonini, "Retinal filtering and image reconstruction," Tech. Rep., Inria Research ReportRR- 6960, 2009.
- [4] S. J. Thorpe, "Spike arrival times: A highly efficient coding scheme for neural networks," R. Eckmiller, G. Hartmann & G. Hauske (Eds.), *Parallel processing in neural systems and computers (pp. 91-94): North-Holland Elsevier*, 1990.
- [5] K. Masmoudi, M. Antonini, and P. Kornprobst, "Frames for exact inversion of the rank order coder," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 2, pp. 353–359, 2012.
- [6] A. A. Lazar, "Population Encoding With Hodgkin-Huxley Neurons," *IEEE Transactions on Information Theory*, vol. 56, no. 2, pp. 821–837, 2010.
- [7] E. Doutsis, L. Fillatre, M. Antonini, and J. Gaulmin, "Retina-inspired Video Codec," *Picture Coding Symposium (PCS), IEEE Press*, pp. 1–5, 2016.
- [8] E. Doutsis, L. Fillatre, M. Antonini, and J. Gaulmin, "Neuro-inspired Quantization," in *IEEE International Conference on Image Processing*, Athens, 2018.
- [9] E. Doutsis, L. Fillatre, M. Antonini, and J. Gaulmin, "Retina-inspired Filter," *IEEE Transactions on Image Processing*, vol. 27, no. 7, pp. 3484–3499, 2018.
- [10] A. Horé and D. Ziou, "Image quality metrics: PSNR vs. SSIM," *20th International Conference on Pattern Recognition, ICPR*, pp. 2366–2369, 2010.
- [11] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, Dec 2012.
- [12] Bruno Cessac, Pierre Kornprobst, Selim Kraria, Hassan Nasser, Daniela Pamplona, Geoffrey Portelli, and Thierry Viville, "Pranas: A new platform for retinal analysis and simulation," *Frontiers in Neuroinformatics*, vol. 11, pp. 49, 09 2017.
- [13] E. Doutsis, L. Fillatre, M. Antonini, and J. Gaulmin, "Retina-inspired filtering for dynamic image coding," *IEEE International Conference in Image Processing (ICIP)*, pp. 3505–3509, 2015.
- [14] W. Gerstner and W. Kistler, *Spiking neuron models: Single Neurons, Populations, Plasticity*, Cambridge University Press, 2002.
- [15] E. Doutsis, L. Fillatre, M. Antonini, and J. Gaulmin, "Bio-inspired Sparse Representation of Images," in *Groupe d'Etudes du Traitement du Signal et des Images (Gretsi)*, 2017.



(a) NICE
 $H = 0.21$ bpp
 PSNR = 7.92 dB
 SSIM = 0.50



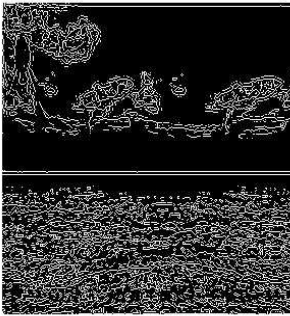
(b) NICE
 $H = 0.45$ bpp
 PSNR = 10.91 dB
 SSIM = 0.69



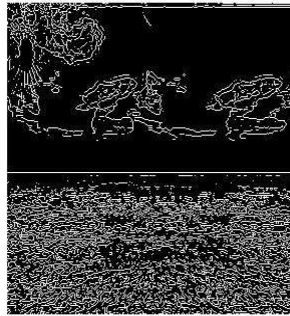
(c) NICE
 $H = 0.22$ bpp
 PSNR = 8.77 dB
 SSIM = 0.3



(d) NICE
 $H = 0.3$ bpp
 PSNR = 10.09 dB
 SSIM = 0.43



(e)



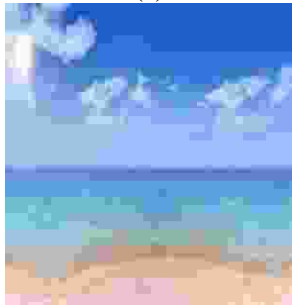
(f)



(g)



(h)



(i) JPEG
 $H = 0.21$ bpp
 PSNR = 26.44 dB
 SSIM = 0.95



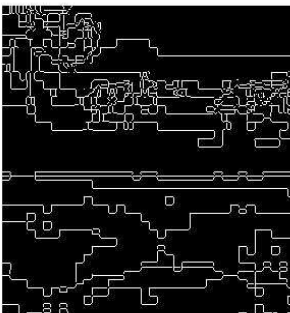
(j) JPEG
 $H = 0.45$ bpp
 PSNR = 32.23 dB
 SSIM = 0.97



(k) JPEG
 $H = 0.23$ bpp
 PSNR = 21.37 dB
 SSIM = 0.81



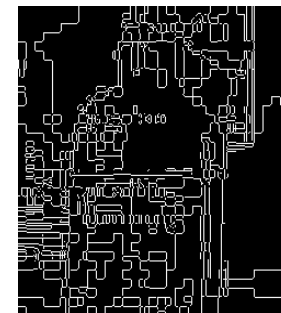
(l) JPEG
 $H = 0.31$ bpp
 PSNR = 24.82 dB
 SSIM = 0.92



(m)



(n)



(o)



(p)

Fig. 3. Visual inspection of the reconstruction performance using NICE and JPEG applied to RGB images. (a)-(d) proposed NICE system; (i)-(k) JPEG standard. (e)-(h) and (m)-(p) correspond to the edge detection output from the NICE-based reconstructions and the JPEG-based reconstructions, respectively (set of parameters: $R = 1000 \Omega$, $C = 1$ F, $T = 100$ msec, $\theta \in [100000, 400000]$, Sobel factor = 0.4).