On the Optimal Choice of Quality Metric In Image Compression: A Soft Computing Approach

Hung T. Nguyen[1], Olga Kosheleva[2], and Vladik Kreinovich[3]

***Abstract*— Images take lot of computer space; in many practical situations, we cannot store all original images, we have to use compression. Moreover, in many such situations, compression ratio provided by even the best lossless compression is not sufficient, so we have to use lossy compression. In a lossy compression, the reconstructed image $\widetilde{I}$ is, in general, different from the original image I. There exist many different lossy compression methods, and most of these methods have several tunable parameters. In different situations, different methods lead to different quality reconstruction, so it is important to select, in each situation, the best compression method. A natural idea is to select the compression method for which the average value of some metric $d(I, \widetilde{I})$ is the smallest possible. The question is then: which quality metric should we choose? In this paper, we show that under certain reasonable symmetry conditions, L^p metrics $d(I, \widetilde{I}) = \int |I(x) - \widetilde{I}(x)|^p \, \mathrm{d}x$ are the best, and that the optimal value of p can be selected depending on the expected relative size r of the informative part of the image.**

I. Formulation of the Problem

A. Image Compression Is Necessary

Images tend to take up a lot of computer space, so in many applications, where we cannot store the original images, we must use image compression. Ideally, we would like to use a lossless compression, but unfortunately, there are serious limitations on how much we can compress without losing information. For a more radical compression, we must therefore use lossy compression schemes. In these schemes, some information about the image is lost; as a result, for every point x, the intensity $\widetilde{I}(x)$ of reconstructed image $\widetilde{I}$ at this point may be slightly different from the intensity $I(x)$ of the original image I at this point.

B. It Is Important to Select Optimal (Or At Least Good Enough) Compression Scheme

There exist many different compression schemes, from standard ones like gif, jpg, jpg2000, etc., to specially designed ones. Most of these schemes comes with one or several parameters which we can select.

One of the reasons why so many different schemes co-exist is that in different applications, different schemes (with different values of parameters) work better. It is vitally important to select an appropriate compression scheme, i.e., a scheme which provides the best compression ratio within the same accuracy. How can we do that?

C. The Notion of a Quality Metric

Intuitively, the quality of a compression scheme can be characterized by how close the decompressed image is to the original one. In order words, the quality of a compression scheme can be described by using an appropriate *metric* $d(I, \widetilde{I})$ on the set of all images. Such metrics describing the "distance" $d(I, \widetilde{I})$ between the two images I and $\widetilde{I}$ are called *quality metrics*.

D. It Is Important to Select Optimal (Or At Least Good Enough) Quality Metric

If we select a quality metric, then we can choose the optimal compression scheme as the one for which the average value of the selected metric is the smallest possible. So, within this approach, in order to select the optimal compression scheme, we must first select the appropriate quality metric.

In some cases, it is clear how to select the quality metric. For example, in some practical applications, we are interested in only one characteristic $c(I)$ of the observed image I: e.g., we may only need to know the total intensity $c(I)$ within a certain zone which characterizes the tumor size. In such cases, our goal is to reconstruct the value $c(I)$ as closely as possible, so we can take the absolute value $|c(I) - c(\widetilde{I})|$ of the difference $c(I) - c(\widetilde{I})$ as

[1]Department of Mathematical Sciences, New Mexico State University, Las Cruces, NM 88003, USA, email hunguyen@nmsu.edu
[2,3]Departments of [2]Electrical and Computer Engineering and [3]Computer Science, University of Texas at El Paso, El Paso, TX 79968, USA, emails olga@ece.utep.edu, vladik@cs.utep.edu

the desired metric $d(I,\widetilde{I}) = |c(I) - c(\widetilde{I})|$. In such applications, the choice of the best compression is straightforward: there is no need to store the entire image I, it is sufficient to store only the single value $c(I)$ as the compressed image. This compression is, in general, extremely lossy, but from the viewpoint of the problem of reconstructing the value $c(I)$, this compression is lossless.

Similarly, if we intend to use only a few characteristics $c_i(I)$ $(1 \le i \le m)$ of an image I, it is natural to compress an image I by storing only the values of these characteristics $c_1(I), \ldots, c_m(I)$. Thus, we get a drastic compression ratio and a perfect reconstruction of all desired values $c_i(I)$.

In many practical situations, however, we do not know *a priori* which characteristics we will be interested in; depending on the situation, we may use the stored image to evaluate many different characteristics. How can we determine the metric in this case?

E. What We Are Planning to Do

Since the problem is not precisely formulated, it is reasonable to use method of soft computing (specifically, fuzzy logic) which were specifically designed to formalize imprecise statements to formalize and solve this problem. In this paper, we propose a three-step solution to this problem:

- First, we use soft computing techniques to describe a general class of quality metrics.
- Second, we use natural symmetry requirements to select a subclass of quality metric characterized by a single parameter p.
- Finally, we show how to select the best value of the parameter p depending on the image.

As a result, we get a data-driven technique for selecting the optimal quality metric and thus, of the optimal compression scheme.

II. First Step: Using Soft Computing Techniques to Select a General Class Of Quality Metrics

Intuitively, the two images are close if for every pixel x, their intensities are close $\widetilde{I}(x) \approx I(x)$, i.e., if for every pixel x, the difference $\Delta I(x) = \widetilde{I}(x) - I(x)$ is small. To formalize this requirement, we will follow the path described in [8, 12].

Let us denote the value of the difference $\Delta I(x)$ at a pixel with coordinates i and j by ΔI_{ij}. In terms of this notation, require that *all* the difference values are small, i.e., that ΔI_{11} is small, and ΔI_{12} is small, and ...

A natural way to formalize this requirement is to use *fuzzy logic*. Let $\mu(x)$ be a membership function that describes the natural-language term "small". Then,

- our degree of belief that ΔI_{11} is small is equal to $\mu(I_{11})$;
- our degree of belief that ΔI_{12} is small is equal to $\mu(I_{12})$,
- etc.

To get the degree of belief b that all conditions are satisfied, we must use a t-norm (a fuzzy analogue of "and"), i.e., use a formula

$$b = \mu(\Delta I_{11}) \,\&\, \mu(\Delta I_{12}) \,\&\, \ldots,$$

where & is this t-norm.

In [13] (see also [2, 3, 4]), we have shown that within an arbitrary accuracy, an arbitrary t-norm can be approximated by a strictly Archimedean t-norm. Therefore, for all practical purposes, we can assume that the t-norm that describes the experts' reasoning, is strictly Archimedean and therefore, has the form $a\&b = \varphi^{-1}(\varphi(a) + \varphi(b))$ for some strictly decreasing function φ [5, 14]. Thus,

$$b(\Delta I) = \varphi^{-1}(\varphi(\mu(\Delta I_{11})) + \varphi(\mu(\Delta I_{12})) + \ldots.$$

We want to find the compression scheme for which, on average, the difference ΔI is as small as possible, i.e., for which our degree of belief $b(\Delta I)$ (that the difference is small) is the largest possible. Since the function φ is strictly decreasing, $b(\Delta I)$ attains its maximum if and only if the auxiliary characteristic $B(\Delta I) = \varphi(b(\Delta I))$ attains its minimum. From the formula that describe $b(\Delta I)$, we can conclude that

$$B(\Delta I) = \varphi(\mu(\Delta I_{11})) + \varphi(\mu(\Delta I_{12})) + \ldots$$

This sum is, in effect, an integral sum, and therefore, as the pixels get denser, this sum tends to the integral

$$B(\Delta I) = d(I,\widetilde{I}) = \int F(\Delta I(x))\,\mathrm{d}x,$$

where $F(z)$ denotes $\varphi(\mu(z))$. Thus, this integral is a reasonable metric.

Since we interpret a metric as a distance, we want the metric to be equal to 0 when the compression is lossless, i.e., when $\Delta I(x) = 0$ for all x. Thus, we want $F(0) = 0$. It is also reasonable to require that the function $F(z)$ be everywhere differentiable (i.e., smooth).

III. Second Step: Using Natural Symmetry Requirements to Select a 1-Parametric Subclass of Quality Metrics

Once a metric $d(I,\widetilde{I})$ is fixed, we can determine which compression is better: if $d(I,\widetilde{I}_1) < d(I,\widetilde{I}_2)$,

then the compression which leads $\widetilde{I}_1$ is clearly better.

In principle, we can use different units to measure the image's intensity. When we select a new unit which is λ times smaller than the old one, then the numerical values of intensity $I(x)$, $\widetilde{I}(x)$, and $\Delta I(x)$ gets multiplied by λ: $\Delta I_{\rm new}(x) = \lambda \cdot \Delta I_{\rm old}(x)$. As a result, the numerical value of the metric may change. It is, however, reasonable to expect the mere change of the measuring unit should not affect our conclusion on which compression was better. Thus, we arrive at the following definition:

Definition. *By a quality metric, we mean the expression of the type $B(\Delta I) = \int F(\Delta I(x))\,{\rm d}x$ for some differentiable function $F(z)$ for which $F(0) = 0$. We say that a quality metric is unit-invariant if for every $\lambda > 0$, the inequality $B(\Delta I_1) < B(\Delta I_2)$ implies that $B(\lambda \cdot \Delta I_1) < B(\lambda \cdot \Delta I_2)$.*

Theorem. *The only unit-invariant quality metrics are the L^p-metrics $B(\Delta I) = {\rm const} \cdot \int |\Delta I(x)|^p\,{\rm d}x$ for $p \ge 1$.*

Main idea of the proof. (Details are similar to [6]; for a general description of how symmetry requirements can help, see [11].) From unit-invariance, one can conclude that if the change $\Delta I(x) \to \Delta I(x) + \varepsilon \cdot \delta I(x)$ does not affect $B(\Delta I)$, then this change should not affect $B(\lambda \cdot \Delta I)$ either. Since $F(z)$ is a differentiable function, when $\varepsilon \to 0$, the change in $B(\Delta I) = \int F(\Delta I(x))\,{\rm d}x$ is asymptotically equal to $\varepsilon \cdot \int F'(\Delta I(x)) \cdot \delta I(x)\,{\rm d}x$, where $F'(z)$ is the derivative of $F(z)$, and the corresponding change in $B(\lambda \cdot \Delta I)$ is asymptotically equal to $\varepsilon \cdot \lambda \cdot \int F'(\lambda \cdot \Delta I(x)) \cdot \delta I(x)\,{\rm d}x$. Thus, if $\int F'(\Delta I(x)) \cdot \delta I(x)\,{\rm d}x = 0$, then we have $\int F'(\lambda \cdot \Delta I(x)) \cdot \delta I(x)\,{\rm d}x = 0$.

In terms of L^2-metric $\int (f(x) - g(x))^2\,{\rm d}x$ on the space of all functions, the condition of this implication means that the function δI is orthogonal to the function $F'(\Delta I)$. Thus, the implication says that every function δI which is orthogonal to $F'(\Delta I)$ is also orthogonal to $F'(\lambda \cdot I)$. From the geometric viewpoint, this can happen only if the functions $F'(\Delta I)$ and $F'(\lambda \cdot I)$ are collinear, i.e., when $F'(\lambda \cdot \Delta I(x)) = c \cdot F'(\Delta I(x))$ for all x.

The coefficient c does not depend on x, but it may depend on λ and also on the function $x \to \Delta I(x)$. From the above condition, however, we can conclude that the coefficient c depends only on the value $\Delta I(x)$ at a given point x. So, if two different functions have the same value somewhere ($\Delta I_1(x) = \Delta I_2(y)$), the corresponding values of c are the same. Hence, c can only depend on λ: $F'(\lambda \cdot \Delta I(x)) = c(\lambda) \cdot F'(\Delta I(x))$.

This is a known functional equation, whose only differentiable solutions are $F'(z) = c_1 \cdot z^\alpha$ for some real numbers c_1 and α (see, e.g., [11]). Since the function $F(z)$ is everywhere differentiable, the value $F'(0)$ must be finite, i.e., $\alpha \ge 0$. Hence, $F(z) = c_2 \cdot z^p + c_2$, where $p = \alpha + 1 \ge 1$. From $F(0) = 0$, we can conclude that $c_2 = 0$. The theorem is proven.

Comment. The L^p-quality metrics are indeed widely used. The value $p = 2$ (corresponding to the mean square decompression error) is most widely used, because for $p = 2$, the optimality criterion is quadratic, and thus, when we minimize it by equating the derivatives to 0, we get an easy-to-solve linear system of equations. However, in many cases, different values of p lead to better compressions.

So, the question is: how to select the value p which is the best for a given practical problem?

IV. Final Step: Selecting the Parameter p of the Quality Metric

A. Our Main Idea

We are interested in the values of several characteristics $c(I)$ of the image I. Instead of using the original image I, we use a degraded image $\widetilde{I} = I - \Delta I$. Since the corruption is small, we can neglect the terms quadratic in ΔI in the expression for the resulting error $\Delta c = c(\widetilde{I}) - c(I)$, and get a linear integral expression $\Delta c = \int \Delta I(x) \cdot a(x)\,{\rm d}x$ for some function $a(x)$.

For each choice of the parameter p, the only information that we have about the difference ΔI is the upper bound D on the corresponding distance $d(I, \widetilde{I}) = \int |\Delta I(x)|^p\,{\rm d}x$. According to the Hölder-Minkowski inequality (see, e.g., [1], Section 4.11.2), for every two integrable functions $f(x)$ and $g(x)$, and for every two real numbers $p, q \ge 1$ for which $1/p + 1/q = 1$, we have

$$\left| \int f(x) \cdot g(x)\,{\rm d}x \right| \le$$

$$\left(\int |f(x)|^{1/p}\,{\rm d}x \right)^{1/p} \cdot \left(\int |g(x)|^{1/q}\,{\rm d}x \right)^{1/q}.$$

Moreover, it is known that for any function $g(x)$, the largest possible value of $|\int f(x) \cdot g(x)\,{\rm d}x|$ over all functions $f(x)$ with a given L^p-norm $\|f\|_p = \left(\int |f(x)|^{1/p}\,{\rm d}x\right)^{1/p}$ is equal to $\|f\|_p \cdot \|g\|_q$, where $\|g\|_q = \left(\int |g(x)|^{1/q}\,{\rm d}x\right)^{1/q}$. Thus, we can conclude that the best possible upper bound for $|\Delta c|$ is the product $D^{1/p} \cdot A^{1/q}$, where $1/p + 1/q = 1$ and $A = \int |a(x)|^q\,{\rm d}x$. It is therefore reasonable to choose p for which the maximum of this product (over the functions $a(x)$ which correspond to all desired characteristics $c(I)$) is the smallest possible.

B. Case Study: Description

As a case study, we take the imaging problems in which the goal is to find the center of gravity of the bright zone, e.g., the center of a tank or the center of a tumor in a medical image. Let us show how, in these problems, we can estimate the values D and A and how we can find the optimal value of p.

C. Case Study: Formalization

Let is first estimate $d(I,\widetilde{I}) = \int |\Delta I(x)|^p \, \mathrm{d}x$. To find the upper bound D for this distance, we need to estimate the difference $\Delta I(x) = \widetilde{I}(x) - I(x)$ between the reconstructed and the original images, and we also need to estimate the area over which we integrate this difference.

Let Δ_0 denote a "typical" error of reconstructing an image from its lossy compression. Then, we can expect that on average, $|\Delta I(x)| \le \Delta_0$ and $|\Delta I(x)|^p \le \Delta_0^p$.

Let us now estimate the area. In the above-described problems, we are only interested in the points x which are reasonably bright, i.e., for which the brightness $I(x)$ exceeds a certain threshold I_0. In such problems, after we reconstruct the image, we can eliminate all the values for which $\widetilde{I}(x) < I_0$. Thus, when the reconstruction is good enough, both the original image $I(x)$ and the reconstructed image $\widetilde{I}(x)$ are concentrated approximately at the same zone. So, their difference $\Delta I(x)$ is also concentrated on this same zone.

Let L denote the linear size of the entire image $[0,L]\times[0,L]$ (including pixels with 0 intensity). Let r denote the portion of the image which we expect to be informative (i.e., filled with non-zeros). Then, the total area of the informative part of the image is approximately equal to $r \cdot L^2$.

Since $|\Delta I(x)|^p \le \Delta_0^p$, and the integration area is equal to $r \cdot L^2$, we have $d(I,\widetilde{I}) = \int |\Delta I(x)|^p \, \mathrm{d}x \le \int \Delta_0^p \, \mathrm{d}x = \Delta_0^p \cdot r \cdot L^2$. Thus, we can take $D = \Delta_0^p \cdot r \cdot L^2$.

For each component c_i of the center of gravity, e.g., for the 1st component, we have

$$c_1(I) = \frac{\int x_1 \cdot I(x)\,\mathrm{d}x}{\int I(x)\,\mathrm{d}x}.$$

In our problems, the value d of the denominator stays approximately the same, namely,

$$d \approx I_{\text{av}} \cdot r \cdot L^2.$$

Therefore, when we substitute $\widetilde{I} = I - \Delta I$ into the above formula and ignore quadratic and other terms, we conclude that $a(x) = x_1/d$. Therefore,

$$A = \int |a(x)|^q \, \mathrm{d}x = \frac{1}{d^q} \cdot \int_0^L \mathrm{d}x_2 \int_0^L |x_1|^q \mathrm{d}x_1 = \frac{1}{d^q} \cdot L \cdot \frac{L^{q+1}}{q+1}.$$

We want to find the value p which minimizes the product

$$D^{1/p} \cdot A^{1/q} = \Delta_0 \cdot r^{1/p} \cdot L^{2/p} \cdot \frac{1}{d} \cdot L^{1/q} \cdot L^{1+1/q} \cdot (q+1)^{-1/q}.$$

D. Case Study: Optimization

Some factors in the minimized expression, such as Δ_0, do not depend on p at all. So, when we find the minimum of the product, we can as well divide by these factors and only minimize what remains without changing the value of p for which this product takes the smallest possible value. Since $1/p + 1/q = 1$, the product of all powers of L in this expression is equal to $L^{2/p+2/q+1} = L^3$, so these terms can also be eliminated without changing the desired value of p. As a result, we must find p as a solution to the following problem: $r^{1/p} \cdot (q+1)^{-1/q} \to \min$. Since $1/q = 1 - 1/p = (p-1)/p$, we have $q = p/(p-1)$ and $q+1 = (2p-1)/(p-1)$, so the minimization problem takes the form:

$$r^{1/p} \cdot (2p-1)^{1/p-1} \cdot (p-1)^{1-1/p} \to \min.$$

Minimizing this expression is equivalent to minimizing its logarithm

$$\frac{1}{p} \cdot \ln(r) - \left(1 - \frac{1}{p}\right) \cdot \ln\left(\frac{2p-1}{p-1}\right) \to \min.$$

Differentiating the minimized function relative to p and equating the derivative to 0, we get the following equation:

$$-\frac{1}{p^2} \cdot \ln(r) - \frac{1}{p^2} \cdot \ln\left(\frac{2p-1}{p-1}\right) - \left(1 - \frac{1}{p}\right) \cdot \left(\frac{2}{2p-1} - \frac{1}{p-1}\right) = 0.$$

If we subtract the fractions, multiply both sides by p^2, and move $\ln(r)$ to the other side, then we get the following equation

$$\ln(r) = -\ln\left(\frac{2p-1}{p-1}\right) + \frac{p}{2p-1}, \qquad (1)$$

or, equivalently, $r = \frac{p-1}{2p-1} \cdot e^{p/(2p-1)}$. In particular, the L^2-method ($p=2$) is optimal for $r = e^{2/3}/3 \approx 0.65$.

If we differentiate the right-hand side of the equation (1) and add the resulting fractions, we conclude that this derivative is equal to $\frac{p}{(p-1)\cdot(2p-1)}$, i.e., it is always positive. Thus, $\ln(r)$ is an increasing

function of p and hence, p is a increasing function of $\ln(r)$ and hence, of r.

When $p \to 1$, we get $\ln(r) \to -\infty$ and $r \to 0$. When $p \to \infty$, we get $\ln(r) \to -\ln(2) + 0.5$, i.e., $r \to \sqrt{e}/2 \approx 0.82$. Thus, we arrive at the following conclusions:

Conclusions

In many image-processing situations, we must select the optimal lossy compression scheme. Due to the lossiness of such compression schemes, the reconstructed image $\widetilde{I}$ may differ from the original image I, i.e., $\Delta I(x) = \widetilde{I}(x) - I(x) \neq 0$. We show that a natural way to select an optimal compression scheme is to select a scheme for which the average value of the quality metric $d(I, \widetilde{I}) = \int |\Delta I(x)|^p \, \mathrm{d}x$ if the smallest possible. The value p should be selected depending on what images we expect:

- If we expect a small image (e.g., a microcalcification in mammography), then the optimal value of p is close to 1, corresponding to $d(I, \widetilde{I}) = \int |\Delta I(x)| \, \mathrm{d}x$.
- When r increases, the value of p increases, and it reaches $p = \infty$, which corresponds to $d(I, \widetilde{I}) = \max |\Delta I(x)|$, for $r = \sqrt{e}/2 \approx 0.82$.

In general, to find the optimal value of p, the must solve the equation (1).

Acknowledgments

This work was supported in part by NASA under cooperative agreement NCC5-209, by NSF grants No. DUE-9750858 and CDA-9522207, by the United Space Alliance, grant No. NAS 9-20000 (PWO C0C67713A6), by the Future Aerospace Science and Technology Program (FAST) Center for Structural Integrity of Aerospace Systems, effort sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant number F49620-95-1-0518, and by the National Security Agency under Grants No. MDA904-98-1-0561.

References

[1] R. E. Edwards, *Functional analysis: theory and applications*, Dover, N.Y., 1995.

[2] S. Jenei, "Limit theorems and the family of nilpotent ordinal sums", In: S. Gottwald and E. P. Klement (eds.), *Abstracts of the Conference on Enriched Lattice Structures for Many-Valued and Fuzzy Logics LINZ'97, Linz, February 25–March 1, 1997*, pp. 110–114.

[3] S. Jenei and E. Pap, "Smoothly generated Archimedean approximation of continuous triangular norms", *Fuzzy Sets Syst.*, 1999, Vol. 104, No. 1, pp. 19–25.

[4] E. P. Klement, R. Mesiar, and E. Pap, *Triangular norm*, 2000 (to appear).

[5] G. Klir and B. Yuan, *Fuzzy sets and fuzzy logic: theory and applications*. Prentice Hall, Upper Saddle River, NJ, 1995.

[6] O. Kosheleva, "Symmetry-group justification of maximum entropy method and generalized maximum entropy methods in image processing", In: G. J. Erickson, J. T. Rychert, and C. R. Smith (eds.), *Maximum Entropy and Bayesian Methods*, Kluwer, Dordrecht, 1998, pp. 101–113.

[7] O. Kosheleva et al., "Compression degradation metrics for analysis of consistency in microcalcification detection", *Proc. 1998 IEEE Symposium on Image Analysis and Interpretation*, April 5–7, Tucson, Arizona, USA, pp. 35–40.

[8] O. Kosheleva et al., "Operations with Fuzzy Numbers Explain Heuristic Methods in Image Processing", *Proceedings of the International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'98)*, Paris, France, July 6–10, 1998, pp. 265–272.

[9] O. Kosheleva, C. Mendoza, and S. D. Cabrera, "Task-specific image quality metrics for lossy compression", In: I. Kadar (ed.), *Signal Processing, Sensor Fusion, and Target Recognition VIII*, Proceedings of the SPIE/International Society for Optical Engineering, Vol. 3720, Orlando, FL, 1999, pp. 397–408.

[10] O. M. Kosheleva et al., "Metrics for evaluation and control of lossy compression effects on the processing of SMD images", *Proceedings of the 1999 IEEE Midwest Symposium on Circuits and Systems*, Las Cruces, New Mexico, August 8–11, 1999 (to appear).

[11] H. T. Nguyen and V. Kreinovich, *Applications of continuous mathematics to computer science*, Kluwer, Dordrecht, 1997.

[12] H. T. Nguyen, V. Kreinovich, and B. Bouchon-Meunier, "Soft Computing Explains Heuristic Numerical Methods in Data Processing and in Logic Programming", L. Medsker (ed.), *Frontiers in Soft Computing and Decision Systems*, AAAI Press (Publication No. FS-97-04), 1997, pp. 30–35.

[13] H. T. Nguyen, V. Kreinovich, and P. Wojciechowski, "Strict Archimedean t-Norms and t-Conorms as Universal Approximators", *International Journal of Approximate Reasoning*, 1998, Vol. 18, Nos. 3–4, pp. 239–249.

[14] H. T. Nguyen and E. A. Walker, *A first course in fuzzy logic,* CRC Press, Boca Raton, Florida, 1999.