1-2002

On the Optimal Choice of Quality Metric in Image Compression

Olga Kosheleva
The University of Texas at El Paso, olgak@utep.edu

Vladik Kreinovich
The University of Texas at El Paso, vladik@utep.edu

Yeung Yam

Comments:
UTEP-CS-00-25c.
Published in the a*Proceedings of the IEEE Southwest Symposium on Image Analysis and Interpretation*, Santa Fe, New Mexico, USA, April 7-9, 2002, pp. 116-120.

On the Optimal Choice of Quality Metric in Image Compression

Olga Kosheleva and Vladik Kreinovich
Departments of Electrical & Computer Engineering and of Computer Science
University of Texas at El Paso, El Paso, TX 79968, USA, vladik@cs.utep.edu

Hung T. Nguyen
Math. Dept., New Mexico State University, Las Cruces, New Mexico 88003, USA

Abstract

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.

1. Formulation of the problem

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.

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 coexist 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?

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*.

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 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?

The larger the difference $\Delta I(x) = \widetilde{I}(x) - I(x)$ between the two images, the larger the "distance" $d(I, \widetilde{I})$ should be. Thus, it is natural to define the desired distance in terms of the difference $\Delta I(x)$. The question is: how exactly?

In this paper, we propose a three-step solution to this question:

- First, we use some reasonable arguments 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.

2. First step: using reasonable arguments to select a general class of quality metrics

The necessity to describe *preferences*, i.e., to describe the *utility* of different alternatives for different people, is extremely important in decision making, including decision making under conflict (also known under a somewhat misleading name of *game theory*). To describe these preferences (utilities), a special *utility theory* has been developed; see, e.g., [2].

The mathematical formalism of utility theory comes from the observation that sometimes, when a person faces several alternatives $A_1, \ldots, A_n$, instead of choosing one of these alternatives, this person may choose a *probabilistic* combination of them, i.e., A_1 with probability p_1, A_2 with a probability p_2, etc. For example, if two alternatives are of equal value to a person, that person will probably choose the first one with probability 0.5 and the second one with the same probability 0.5. Such probabilistic combinations are called (somewhat misleadingly) *lotteries*. In view of this realistic possibility, it is desirable to consider the preference relation not only for the original alternatives, but also for arbitrary lotteries combining these alternatives. Each original alternative A_i can be viewed as a *degenerate* lottery, in which this alternative A_i appears with probability 1, and every other alternative $A_j \neq A_i$ appear with probability 0.

The main theorem of utility theory states that if we have an ordering relation $L \succ L'$ between such lotteries (with the meaning "L is preferable to L'"), and if this relation satisfies natural consistency conditions such as transitivity, etc., then there exists a function u from the set $\mathcal{L}$ of all possible lotteries into the set R of real numbers for which:

- $L \succ L'$ if and only if $u(L) > u(L')$, and
- for every lottery L, in which each alternative A_i appears with probability p_i, we have

$$u(L) = p_1 \cdot u(A_1) + \ldots + p_n \cdot u(A_n).$$

This function u is called a *utility function*. Each consistent preference relation can thus be described by its utility function.

The correspondence between preference relations and utility functions is not 1-1: different utility functions may correspond to the same preference. For example, if the preference relation is consistent with the utility function $u(L)$, then, as one can easily see, it is also consistent with the function $u'(L) = a \cdot u(L) + b$, in which a is an arbitrary positive real number, and b is an arbitrary real number.

It is known that such linear transformations form the only non-uniqueness of utility function: namely, if two utility functions $u(L)$ and $u'(L)$ describe the same preference relation, then $u'(L) = a \cdot u(L) + b$ for some real numbers $a > 0$ and b.

In general, preference relations can be described by utility functions. Therefore, to describe preferences between images, we need a utility function that is defined on the set of all possible images.

In particular, if we consider realistic images I that consist of finitely many pixels and are thus described by a matrix with components I_{ij}, $1 \leq i, j \leq N$, then we need a utility function that depends on N^2 parameters ΔI_{ij}:

$$u(\Delta I) = u(\Delta I_{11}, \ldots, \Delta I_{1N}, \ldots, \Delta I_{N1}, \ldots, \Delta I_{NN}).$$

What function should we choose? An important feature of many image processing problems, including the two above problems, is their *localness*: different parts of the image are pretty much independent on each other. Let us explain what we mean on the above examples:

- In *astronomy*, the relative quality of two possible reconstructions of a part of this image (a part that contains, e.g., a galaxy), does not depend on the remaining part of the image.
- In *medicine*, the relative quality of reconstructing a part of the mammogram does not depend on whether there is anything in the rest of this mammogram.

In mathematical terms, if

- an image I_{ij} is preferable to an image I'_{ij} that differs from I_{ij} only in pixels (i, j) from some set P, and
- images $\widetilde{I}_{ij}$ and $\widetilde{I}'_{ij}$ coincide with each other for $(i, j) \notin P$ and with, correspondingly, I_{ij} and I'_{ij} for $(i, j) \in P$,

then $\widetilde{I} \succ \widetilde{I}'$.

This "localness" ("independence") is a frequent feature in practical problems, and utility theory has developed a precise description of utility functions that satisfy this property. Namely, it has been shown that when alternatives are characterized by n parameters $p_1, \ldots, p_n$, then the localness of the preference is equivalent to the utility function $u(p_1, \ldots, p_n)$ being of one of the two types [2]:

- *additive* $u(p_1, \ldots, p_n) = u_1(p_1) + \ldots + u_n(p_n)$ for some functions $u_i(p_i)$; or
- *multiplicative* $u(p_1, \ldots, p_n) = u_1(p_1) \cdot \ldots \cdot u_n(p_n)$ for some functions $u_i(p_i)$.

In utility theory, the values $u_i(p_i)$ are called *marginal utilities*.

For images, $n = N^2$, parameters $p_1, \ldots, p_n$ are the values of the difference $\Delta I_{ij} = \Delta I(x_{ij})$ at different points x_{ij} of the grid, and the resulting forms of the utility function are $u(\Delta I) = \sum\limits_{ij} u_{ij}(\Delta I_{ij})$ and $u(\Delta I) = \prod\limits_{ij} u_{ij}(\Delta I_{ij})$. So, to describe the utility function, we must describe how the value of the marginal utility depends on the point x_{ij} and on the brightness $\Delta I_{ij} = \Delta I(x_{ij})$ at this point. We can describe this dependence explicitly if, instead of the abbreviated notation ΔI_{ij}, we use $\Delta I(x_{ij})$, and if we describe the dependency $u_{ij}(\Delta I_{ij})$ as $U(x_{ij}, \Delta I(x_{ij})$. In this case, the formula for the utility function takes one of the forms $u(\Delta I) = \sum U(x_{ij}, \Delta I(x_{ij}))$ or $u(\Delta I) = \prod U(x_{ij}, \Delta I(x_{ij}))$.

So far, we were analyzing the problem of how to compare different pixel-by-pixel images. In real-life, the object whose image we want to describe is continuous, pixels are simply a useful approximation. It is, therefore, desirable to reconstruct not just the values on a grid, but also the entire brightness distribution, i.e., the values of $\Delta I(x)$ for every point x. To achieve this goal, we must be able to compare the quality of different functions $\Delta I(x)$, i.e., we must be able to describe the value of utility $u(\Delta I)$ for different functions ΔI.

The denser the pixels (i.e., the smaller the distances h_x and h_y between the neighboring pixels), the closer the pixel-by-pixel image to the continuous one. Therefore, as a utility $u(\Delta I)$ of a function ΔI, we can take the *limit* of the utilities of its pixel-by-pixel representation as $h_x \to 0$ and $h_y \to 0$. How can we describe such a limit?

This limit is easy to describe for the case when utility is a *sum* of marginal utilities: in this case, the sums are, in effect, integral sums, and therefore, as the pixels get denser, the sums tend to the integral $u(\Delta I) = \int U(x, \Delta I(x))\,\mathrm{d}x$. For the case when utility is a *product* of marginal utilities, the limit can be obtained indirectly: indeed, since utility is a product of marginal utilities, its *logarithm* is the *sum* of logarithms of marginal utilities:

$$v(\Delta I) = \log(u(\Delta I)) = \sum V(x_{ij}, \Delta I(x_{ij})),$$

where $V = \log(U)$. For these logarithms, we also get integral sums and therefore, a reasonable limit expression $v(\Delta I) = \int V(x, \Delta I(x))\,\mathrm{d}x$, and $u(\Delta I) = \exp(v(\Delta I))$. Our goal is to find a compression method for which $u(\Delta I) \to \min$. Since logarithm is monotonic, the condition $u(\Delta I) \to \min$ is equivalent to $v(\Delta I) = \log(u(\Delta I)) \to \min$. Therefore, in multiplicative case, we get the same problem $\int V(x, \Delta I(x))\,\mathrm{d}x \to \min$ as in the additive case.

The "quality" of a compression scheme should not change if we simply shift the image. Thus, the function U should not explicitly depend on x, i.e., and we should have $d(I, \widetilde{I}) = \int U(\Delta I(x))\,\mathrm{d}x$.

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 $U(0) = 0$. It is also reasonable to require that the function $U(z)$ be everywhere differentiable (i.e., smooth).

3. 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 to $\widetilde{I}_1$ 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: $\Delta I_{\mathrm{new}}(x) = \lambda \cdot \Delta I_{\mathrm{old}}(x)$. 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 $B(\Delta I) = \int U(\Delta I(x))\,\mathrm{d}x$ for some differentiable function $U(z)$ for which $U(0) = 0$. 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 $B(\Delta I) = \mathrm{const} \cdot \int |\Delta I(x)|^p\,\mathrm{d}x$ for $p \geq 1$.*

Proof. The proof of this theorem is similar to the proofs from [3]; for a general description of how symmetry requirements can help, see [5].

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 $U(z)$ is a differentiable function, when $\varepsilon \to 0$, the change in $B(\Delta I) = \int U(\Delta I(x))\, \mathrm{d}x$ is asymptotically equal to

$$\varepsilon \cdot \int U'(\Delta I(x)) \cdot \delta I(x)\, \mathrm{d}x,$$

where $U'(z)$ is the derivative of $U(z)$, and the corresponding change in $B(\lambda \cdot \Delta I)$ is asymptotically equal to

$$\varepsilon \cdot \lambda \cdot \int U'(\lambda \cdot \Delta I(x)) \cdot \delta I(x)\, \mathrm{d}x.$$

Thus, if $\int U'(\Delta I(x)) \cdot \delta I(x)\, \mathrm{d}x = 0$, then we have

$$\int U'(\lambda \cdot \Delta I(x)) \cdot \delta I(x)\, \mathrm{d}x = 0.$$

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

$$U'(\lambda \cdot \Delta I(x)) = c(\lambda) \cdot U'(\Delta I(x)).$$

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., [5]). Since the function $U(z)$ is everywhere differentiable, the value $U'(0)$ must be finite, i.e., $\alpha \geq 0$. Hence, $U(z) = c_2 \cdot z^p + c_2$, where $p = \alpha + 1 \geq 1$. From $U(0) = 0$, 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?

4. Final step: selecting the parameter p of the quality metric

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 $\Delta c = \int \Delta I(x) \cdot a(x)\, \mathrm{d}x$ 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 $d(I, \widetilde{I}) = \int |\Delta I(x)|^p\, \mathrm{d}x$. 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 $p, q \geq 1$ for which $1/p + 1/q = 1$, we have

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

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

Moreover, it is known that for any function $g(x)$, the largest possible value of $|\int f(x) \cdot g(x)\, \mathrm{d}x|$ over all functions $f(x)$ with a given L^p-norm $\|f\|_p = \left(\int |f(x)|^{1/p}\, \mathrm{d}x\right)^{1/p}$ is equal to $\|f\|_p \cdot \|g\|_q$, where $\|g\|_q = \left(\int |g(x)|^{1/q}\, \mathrm{d}x\right)^{1/q}$. 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 $A = \int |a(x)|^q\, \mathrm{d}x$. 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.

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.

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, $|\Delta I(x)| \leq \Delta_0$ and $|\Delta I(x)|^p \leq \Delta_0^p$.

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

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

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; differentiating it relative to p and equating the derivative to 0, we get the following equation:

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

or, equivalently, $r = \dfrac{p-1}{2p-1} \cdot e^{p/(2p-1)}$. 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 $\dfrac{p}{(p-1)\cdot(2p-1)}$, 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.

5. Image compression: 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., $\Delta I(x) = \widetilde{I}(x) - I(x) \ne 0$. 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 micro-calcification 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 grants NCC5-209 and NCC 2-1232, by the Air Force Office of Scientific Research grant F49620-00-1-0365, and by NSF grants CDA-9522207, ERA-0112968 and 9710940 Mexico/Conacyt.

References

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

[2] P. C. Fishburn, *Nonlinear preference and utility theory*, The John Hopkins Press, Baltimore, MD, 1988.

[3] 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.

[4] 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.

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