7/21/'95

Optimal data compression of visual images

version 1

Thomas I. Seidman
Department of Mathematics and Statistics
University of Maryland Baltimore County
Baltimore, MD 21228, USA
e-mail: ⟨seidman@math.umbc.edu⟩

ABSTRACT: We consider a general framework for consideration of optimality in the compression of visual images and then, noting that human perception plays a special role in this analysis, discuss some currently available models of the 'eye metric' for measuring the degradation of image quality associated with loss of information and propose a novel compression scheme based on this. The setting for this discussion is that of a single[1] gray-scale image, initially expressed in terms of the 'intensity function' $f : \Omega \to [0, 1]$ with f in some (rather special) class $\mathcal{F}$ of potential 'natural images'.

This is a somewhat condensed version of [27] so the discussion and justification of the principal points to be made are intended for presentation at somewhat greater length there.

[1]At this point we simplify our discussion by omitting consideration of the additional issues raised by stereoscopic data fusion (depth perception, etc.) and by the context of evolution in time (tracking motion in a sequence of images, etc.); similarly, we omit consideration of issues related to color, although those should involve rather less modification of the ideas presented here.

1. Framework for general analysis

In general, the cost $\mathcal{J}$ associated with use of any compression scheme is a sum of three components:

$$\begin{aligned} \mathcal{J} = \quad & \mathcal{J}_1 && \textit{[cost of compressed size]} \\ & +\mathcal{J}_2 && \textit{[cost of image degradation]} \qquad (1.1) \\ & +\mathcal{J}_3 && \textit{[computational cost].} \end{aligned}$$

Note that $\mathcal{J}_2 = 0$ when one restricts attention to lossless compression but this will here be precisely the term of dominant interest. On the other hand, we will simplify our analysis by assuming that computation is cheap so we may neglect $\mathcal{J}_3$ and focus on the relevant trade-offs[2] between $\mathcal{J}_1$ and $\mathcal{J}_2$. Note that $\mathcal{J}_1$ may typically be taken in the form[3]

$$\mathcal{J}_1 := \alpha K \qquad K := \#\text{bits (compressed image)} \tag{1.2}$$

[2]Note that this viewpoint ignores such visual processing as deblurring (apart from filtering) and enhancement, to the extent that these refer only to modification of the appearance of the reconstructed image without alteration of the intrinsic information content. We are now implicitly assuming that, in each case, we exhaust the possibilities of lossless compression and are primarily concerned with the *additional* compressive savings to be obtained by discarding some information to the extent that this can be done without significant loss of the information which is genuinely *useful*. Thus, this analysis ignores the actual form of the compression coding as a data representation. Note that, in this context, *any* reconstruction theory is also a compression theory, since data which could be reconstructed involves only a representation change without information loss while any information actually lost in a successful reconstruction is necessarily information which could be also discarded in compression without significant cost. The primary distinction is that the distortion occurring in a situation for which reconstruction is needed is imposed on us whereas we may choose what is to be discarded in compression.

We insert, at this point, mention of the possible desirability of a segmented (multistage) compression scheme in which, e.g., the 'first segment' serves dual functions: being both part of the complete 'compressed code' for reconstruction and also providing an automated *labeling*: acting as a set of keys to permit/facilitate intelligent retrieval from an archive of visual images.

[3]One must also include in K the specification of (the parameters of) the compression scheme, if this is not fixed *a priori*.

but that the form of $\mathcal{J}_2$ is much less clear: we only know that $\mathcal{J}_2$ should represent a sensitivity, measuring the gradient of an 'information-value functional' with respect to perturbation of the reconstructed image.

What we would like, for compression, would be a localization — in some (conceptual) representation space $\mathcal{R}$ for image data — specifying a region $\mathcal{U}$ likely to contain the 'useful information' in the sense of $\mathcal{J}_2$, whatever that might be. In focusing our attention on the informational and optimality aspects of $\mathcal{J}_1$, $\mathcal{J}_2$, we neglect the algorithmic mechanics of representation conversion and reconstruction. Essentially, a compression scheme then just consists of the excision of the complement $\mathcal{R}\backslash\mathcal{U}$ from this representation as presumably containing only negligible information. In some sense, $\mathcal{J}_1$ is the size[4] of the defining region $\mathcal{U}$ in this decomposition. Now think of the density of 'value of information' as a function ϕ which is small on $\mathcal{R}\backslash\mathcal{U}$ but may never quite vanish. Since we cannot expect any perfect decomposition, we accept some degradation in picture quality for the reconstructed image and $\mathcal{J}_2$ measures the discarded (useful) information[5] $\int_{\mathcal{R}\backslash\mathcal{U}} \phi$. A major part of our goal is to obtain some insight into the nature of this (conceptual) function ϕ to permit an optimal choice of $\mathcal{U}$.

As in [25], [26], we first argue that $\mathcal{J}_2$ is sharply application-dependent, so no general form for $\mathcal{J}_2$ is to be expected — although we will later argue that a special role will be played by the so-called 'eye metric', corresponding to the sensitivities of human perception.

Assuming that there is some (specific) purpose[6] associated with the trans-

---

[4]I.e., we may take $\mathcal{J}_1 \approx \alpha \int 1_{\mathcal{U}}$ to the extent that the implied measure on $\mathcal{R}$ corresponds well to 'information content'.

[5]Suppose, for the same representation, we denote the noise level as distributed over $\mathcal{R}$ by ν. Then, similarly, 'noise filtering' essentially consists of excision of a region $\mathcal{V} \subset \mathcal{R}$, selected so the remaining noise $\int_{\mathcal{R}\backslash\mathcal{V}} \nu$ is small while the lost useful information $\int_{\mathcal{V}} \phi$ is again negligible. Comparing $\mathcal{V}$ to $\mathcal{R}\backslash\mathcal{U}$, we see that these problems share the same $\mathcal{J}_2$ so insight into either can be applied to obtain understanding of the other.

[6]We insert a brief annotated list of some possible types of applications for subsequent more detailed investigation.

medical imaging These applications are highly contextual: detection of fractures or of arterial blockages, location of possible tumors, diagnosis of tuberculosis CAT or MRI or sonogram interpretation, etc.

aerial or satellite photography Note the need to correct for distortion (due to height and angle of view as well as nonlinear distortion due to lens aberration or surface curvature); note also the 'registration problem' (cf. e.g., [13]) in aligning adjacent mission/storage of the image, we may expect an 'application map'

$$A : \boxed{f = \text{image}} \mapsto \boxed{a = \text{action}} : \mathcal{F} \to \mathcal{A} \tag{1.3}$$

where the space $\mathcal{A} = \mathcal{A}_A$ of possible 'actions' may be either discrete (e.g., a 'recognition' problem) or continuous (e.g., robotic control). The appropriate cost $\mathcal{J}_2$ should then be taken as the economic cost of error in the action taken. For the discrete case $\mathcal{J}_2$ might be an expected cost for mis-identification in the presence of noise whereas for the continuous case it would be related to the closeness of the action $\tilde{a}$ computed from the decompressed image $\tilde{f}$ from the 'correct' action $\bar{a}$ so

$$\begin{aligned} \mathcal{J}_2(\tilde{f}) &:= \hat{\mathcal{J}}_A(\tilde{a}|\bar{a}) \qquad \tilde{a} = A(\tilde{f}) \\ &:= [\text{cost of using } \tilde{a} \text{ when } \bar{a} \text{ is correct}]. \end{aligned} \tag{1.4}$$

What then makes compression possible is the observation that (especially in the presence of *a priori* contextual function) the image f is special: $f \in \mathcal{F}$ with $\mathcal{F}$ compact in the application-defined topology[7] so we can optimize $\mathcal{J}$

images for concatenation (allowing for variable cloud cover, etc.) or relating image sequences at different scales. These applications frequently also involve difficult questions of interpretation for data fusion, e.g., in updating weather data.

tracking and real-time control E.g., robotic vision, automated vehicle navigation, etc. We might include here also some applications of quality control, as detection of flaws (mismatches from a 'standard' image) in a VLSI circuit chip (also, e.g., cf., [31]).

recognition and matching This ranges from IFF in real time to reading handwriting (recognition of letters or of letter clusters, using local context) and includes archival compressed storage of finger prints or ID photos for later matching — again note the problems of distortion correction and registration — as well as recognition of 'interesting' astronomical or cloud chamber events. Perhaps counting problems (e.g., whitw cells in a blood sample) could be included here. One might also include a potential for intelligent automated TV channel surfing...

etc. ...

[7]Note that we expect that $\mathcal{U}(\varepsilon;\bar{a})$ will be open in the action space $\mathcal{A}_A$ for $\varepsilon > 0$ so, assuming continuity of the map A in (1.3), each $\mathcal{V}(\varepsilon;g)$ will be open and it is precisely the compactness of $\mathcal{F}$ which ensures the finiteness of $N(\varepsilon)$. This may be compared with the notion of 'relevant information' implicit in Kolmogorov's definition of 'ε-entropy'.

through the minimization

$$
(1.5)\qquad
\begin{aligned}
\mathcal{J}_{\text{opt}} = &\ \min_{\varepsilon} \{\varepsilon + \alpha\lceil \ln_2 N(\varepsilon)\rceil\} \\
N(\varepsilon) := &\ \#\{ \text{ a finite subcover of } \mathcal{F} \text{ from} \{\mathcal{V}(\varepsilon; g) : g \in \mathcal{F}\}\} \\
\text{where} &\quad \begin{aligned} \mathcal{V}(\varepsilon; g) &:= A^{-1}\left[\mathcal{U}(\varepsilon; A(g))\right] \\ \mathcal{U}(\varepsilon; \bar{a}) &:= \left\{\tilde{a} \in \mathcal{A}_A : \hat{\mathcal{J}}_A(\tilde{a}|\bar{a}) < \varepsilon\right\} \end{aligned}
\end{aligned}
$$

Once ε is determined in (1.5), specification of the compression scheme consists of specification of the subcover to be used — e.g., in terms of a set of 'centers' or, more effectively, in terms of the corresponding set of actions $\{A\,(g_k) =: a_k : k = 1, \ldots, N(\varepsilon)\}$: one is really precomputing and compressing action specifications rather than images. A somewhat similar comment may be made for the discrete case: the compression need only label the 'index' $a \in \mathcal{A}_A$ and, given suitable *a priori* information, quite spectacular compression ratios may be possible.[8]

In many settings, however, the specific application may not (yet) be known at the time of compression or may involve some later data fusion. For such settings A, hence (1.5), is unavailable. We might be able to replace (1.5) by

$$
(1.6)\qquad \mathcal{J}_{\text{opt}} = \min_{\varepsilon} \{\varepsilon + \alpha \boldsymbol{\mathcal{E}}_A \lceil \ln_2 N(\varepsilon)\rceil\}\,, \quad \text{etc.}
$$

where $\boldsymbol{\mathcal{E}}_A$ denotes the expected value corresponding to some probability distribution over 'potential applications' $A \in \boldsymbol{\mathcal{A}}$ (alternatively, one might use a 'worst case' analysis for $A \in \boldsymbol{\mathcal{A}}$), although a corresponding construction of the specific compression scheme may not then be possible. It is in this setting that we argue for the special role of human perception and the 'eye metric'.

2. Human perception

The argument for the special role of human perception is based on three considerations:

[8] The ISBN code, widely available in practice, is an extreme case: e.g., one has lossless compression of the complete 435-page book [22] to 10 digits: 0–12–597301–2, with exact reconstruction available through a bookstore.

- This is the 'default standard' — universally available and widely used: most images are judged/interpreted by humans.[9]

- Historically, many applications exist precisely because humans have been able to do them adequately and find them consequently useful.[10]

- The human eye has evolved to handle a variety of (moderately universal) basic tasks, so it is not surprising that the processing modes of human perception accomplish these rather well.[11] Thus, techniques of compression–reconstruction which facilitate human perception must necessarily preserve the information needed for all applications for which success can be based on these tasks.

Our first observation is that the precise nature of human vision is quite complicated and is rather imperfectly understood. We are forced to infer relevant characteristics from optical and neurological theory,[12] from experimen-

---

[9]We quote from the chapter in [3] on 'Detection, recognition, and subjective image quality' —

> "The history of image evaluation in aerial photography exhibits an intriguing counterpoint on the theme of objective and subjective image quality. Variations appear, disappear and resurface according to the shifting emphasis in the appreciation that there can be no progress in accuracy without taking the human being out of image evaluation, and no confidence in its significance without putting him in again."

[10]Consider, e.g., reading handwritten material. This is important because we have structured our activities to make it so, which was possible only because this is something we *can* do. These applications are typically the automation of human tasks without restructuring the activities which depend on them.

[11]In appealing to human perception we may explicitly include consideration of *enhancement*: we are not seeking to make the reconstructed image prettiest or most like the original — indeed, e.g., in the case of deblurring, the reconstructed image may be considerably preferable to the original — but to maximize the facility of a human participant in interpreting this image for a range of applications so it is entirely plausible to compare images *after* the deliberate introduction of useful exaggeration/distortion, etc.

[12]The influences of [11] and [10], in particular, were noted in [18] and we quote,

> "A striking confirmation of the importance of edge detection in vision (which even leads the perceptual system to 'see' illusory contours) was found by R. von der Heydt and his laboratory [9]. The responses of single cells in visual areas are codified by describing the visual field: the area within which

tation, and from interpretations[13] of 'what works' in algorithms for various aspects of computer vision. Much of the 'conventional wisdom' was applied to these questions in a spate of books [2], [16], [22] in 1982 but we shall find particularly useful such more recent work as [7] (in the 1992 special issue on wavelets of the IEEE Trans. Inf. Theory) and the discussion of segmentation (feature detection) still more recently in [18].

There seems to be a fairly clear distinction between preconscious[14] low-level vision processing and higher level cortical processing for humans. This is somewhat like the distinction between a 'cipher' and a 'code': if it would be feasible to segment and identify semantically meaningful visual elements within an image, then we might achieve significantly greater compression by labeling these (perhaps with annotation for relevant parameters) for reference to some library of picture fragments.[15] At this point we do not pursue this line of analysis and will not attempt to analyze or to imitate (as an AI problem) the more interpretive higher level functions of human image processing — which, in any case, are typically relevant to fairly specific contexts, rather than to generic processing tasks — but will focus on the considerations associated with 'early vision'. As 'visual semioticians' we will focus on the syntaxis, rather than the semantics or pragmatics, of image analysis.

Low-level vision seems particularly effective in detecting *change*.[16] We also note that the number of receptors in the human eye, viewed as a sampling

> moving or stationary bars and edges produces activity. They found, however, that many cells in visual area $V2$ responded when no actual stimulus was present in their visual field, but when edges outside this field produced illusory contours that cross the field."

[13]We have already noted the close conceptual relation between compression and noise filtering in their use of degradation measures.

[14]The word 'subliminal' (='below threshhold') would be ambiguous here: the process is typically below some threshhold for attracting *conscious* attention — indeed, we are here regarding it as a precursor to conscious processing — but if the presented data would actually be below the perceptive threshhold then this processing could not occur at all.

[15]This seems to be the case for dreams and much human memory (although there *is* evidence for some genuinely eidetic memory) and is, e.g., implicit in the treatments of text and of textural data within 'painting' software.

[16]Note the effectiveness of the 'blink comparator', which immediately draws attention to a single point with significant change in position or brightness (planet, comet, nova) from the mass of data represented by a star field. Experiment shows that a frog *only* sees motion: EEG measurement shows no signal to the brain from a time-constant image.

grid, is regular enough and coarse enough that it should — by standard signal processing theory — produce noticeable aliasing effects and the non-appearance of these may be related to a time-averaging effect of rapid eye motion. This jiggling of registration, on the other hand, could be expected to convert (local) spatial discontinuities (edges) to temporal changes detectible by a single receptor: one has a coding of the spatial intensity gradient at a scale related to the jiggling amplitude.

Next, we note that the eye adapts to the illumination level[17] through a *local normalization*: interpretation of standard optical effects (cf., e.g., Figures 1,2,3 on pp. 54,55 of [22]) shows that apparent brightness is judged by comparison with the local illumination level. Thus one concludes that:

- The eye metric for degradation must be inherently nonlinear.
- One should expect edge effects for variation of brightness on some spatial scale when quite near an edge (associate with a coarser scale) — although, perhaps surprisingly, the eye does seem to compensate partially in attempting to normalize with respect to a neighborhood within a 'perceived region'.
- One should anticipate similar phenomena of local normalization for other components of vision. This has long been known for color contrast but we now observe that this also seems recently established [28] for orientation.[18]

While we are not considering the data fusion involved in stereoscopic vision, we note that a dominant determinant of depth perception is perceived occlusion, with the eye interpreting edges as projections of the boundaries of 2-D or 3-D objects.[19] Thus, edge detection (and the related segmentation of the image into significant subregions) appears as a major processing

[17]If this occurred only on a global basis (even allowing for the obvious saturation effects), there would be only a minor difficulty, repairable through multiplication of the norm by a suitably adjusted weight.

[18]The modeling of [28] was based on 'masking' experiments, involving the superposition of two patterns. A linear theory would suggest simple addition of the energies in orthogonal components but this was demonstrably false: one observed an interaction, attributed in [28] to a nonlinear local normalization.

[19]This shows up particularly in the observed significance of T-junctions — interpreted as the occlusion of an edge by a foreground object. One also, of course, has edges which are, e.g., interpretable as shadow boundaries.

task, reflecting — perhaps determining — much of the underlying structure of low-level vision and human perception. Apart from the relevance of edge detection to so-called 2.1-D depth perception, we observe that one of the major stages of syntaxic visual processing is the identification of 'salient features', attracting attention as carrying the greatest (density of) significant information.[20] Note that these perceived 'features' may be (small) regions, but are quite likely to be lower-dimensional: especially '1-sets' (1-dimensional curves, as bounding edges, etc.[21]) and 0-sets (both isolated points and as curve ends and intersections or corners, etc.). The induced 'significance measure' is unlikely to be absolutely continuous with respect to the Lebesgue measure on an image window, suggesting some of the considerations of geometric measure theory (cf., e.g., [8], as well as the middle third of [18]). It is also this geometric localization which emphasizes the advantages of wavelet expansion over Fourier analysis.

3. Diffusive filtering, etc.

We note here an influential conjecture by D. Marr [16] that the edge representation at a variety of scales should suffice for complete reconstruction of a visual image. The paper [15] describes a successful[22] concrete implementation (proposed and discussed earlier in [32], [14]) of image reconstruction from edge data (more precisely, from knowledge of the 0-crossings of a wavelet

[20]That much of this identification of 'significance' occurs at a preconscious level is indicated by the almost 'wired-in' recognition of a 'face' by infants and by dichotomous illusions of recognition (as, e.g., the well-known example of 'vase *vs.* faces').

This identification of significance strongly affects the subjective judgment of picture quality. Experimentation at Kodak indicates [5] that "In digital imaging manipulation of faces, for example, ... the few pixels which represent specular reflection of light sources in the eye image are critical to viewers judging the image as good or bad."

[21]Non-edges to be considered might be projections of space curves [= 1-D objects in the physical 3-D space — although these may effectively be treated as 3-D objects which, except for one dimension, have 'small scale'] but 1-sets might also arise as artifacts of finer-scale variation [= texture].

[22]Meyer has shown by a counterexample [17] that uniqueness fails in this context. Presumably the success noted for this technique must again reflect the fact that 'natural images are special'.

It is also worth noting at this point (a) that [15] utilizes the L^2-norm for $\mathcal{J}_2$ and (b) that one continues to get good results after omission of certain 'edges' which appear to be spurious.

transform) at several scales. We do not pursue this goal of computation-reconstruction exclusively through such edge representations, but are now led to seek a formulation of the eye metric consistent with a multiscale analysis and the dominant significance of edges. Although the book [18] is primarily intended to address issues of segmentation rather than compression, we have found the literature survey therein particularly valuable for these considerations, while noting that it has not always been easy to extract the precisely relevant information.

The focus of [18] is minimization of a version of the 'Mumford–Shah energy functional'

$$\boldsymbol{\mathcal{E}}(K) = \nu \mathcal{H}^1(K) + \inf_u \left\{ \int_{\Omega \setminus K} \left[\lambda |\nabla u|^2 + |u - f|^2 \right] \right\} \tag{3.1}$$

where $\mathcal{H}^1$ denotes the length (more precisely, the 1-dimensional outer Hausdorff measure) of K and $f : \Omega \to [0,1]$ is the given intensity function.[23] The parameter λ in (3.1) is related to scale and an approximate relation is noted between minimization of $\boldsymbol{\mathcal{E}}$ and the evolution of u from $u_0 = f$ through an associated diffusion equation up to a small 'time' $t = \lambda^2$ (with K then appearing as the set on which $\Delta u = 0$ and u itself giving a filtered version of the possibly noisy f). The principal results of the book are that the minimizer K exists and has certain regularity properties, making it a reasonable choice as the 'edge set' for f.

We next observe that, while most of the earlier work in 2-D signal processing — and also much contemporary investigation — uses the mathematically convenient $L^2(\Omega)$-norm as a degradation measure (the usual 'least squares' error criterion), there is now a fairly unanimous agreement that this is *not* consistent with the facts of human perception for visual images. A cogent argument, based on the slope of the log-log CST[24] is presented in [7] for use

[23]The original Mumford–Shah functional ([19] or pp. 36*ff* of [18]) essentially restricted the $\inf_u$ in (3.1) to functions piecewise constant on the connected components of $\Omega \setminus K$ (also allowing for occluded portions) and so ommitted the $|\nabla u|^2$ term; it also integrated $\varphi(\kappa)$ (a function of curvature along K, of linear growth for large $|\kappa|$) rather than using $\mathcal{H}^1(K)$. For this or for the version (3.1), the minimization balances some notion of regularity against the (L^2) faithfulness to f, much along the lines of the optimization for (1.1) or treatment of ill-posed problems via Tikhonov regularization (for which, compare [29].

[24]Contrast Sensitivity Threshhold, marking the level of contrast (amplitude of brightness oscillation) at which this oscillation, at a given spatial frequency, is just noticeable.

of the L^1-norm instead. One easily sees that this cannot quite be true and, indeed, that *no* L^p-norm can give satisfactory agreement with human perception. We have already noted the nonlinearity of local normalization and have also observed the concentration of significant information along 1-sets (edges, etc.). It is also easy to see that several characteristic length scales[25] occur as relevant to human perception — e.g., the distinction between central and peripheral vision and, implicitly, the scale of localization for normalization effects — which again is incompatible with a simple L^p theory.[26] It would not be too difficult to accommodate local normalization as a minor modification of the L^1 theory of [7], but the other objections noted require us to take a somewhat different approach, emphasizing geometric localization and the significance of edge effects.

An 'edge', for our purposes, is a 'curve of discontinuity' across which one has a *significant abrupt change*.[27] Looking, for simplicity, at comparable 1-D

Displaying a generated image with frequency increasing along one axis and contrast falling off along the other, one observes a division into a region of perceptible variation and a complementary gray region. One easily observes the separating CST curve and, with a log-log scaling, notes that (approximately) it is a straight line with slope -2. Since $\|c \sin 2^k \pi x\|_{L^p(I)}$ with contrast c and scale 2^k gives this behavior only for $p = 1$, [7] concludes that the L^1-norm is the correct measure (while noting that this remains somewhat arbitrary as one would obtain similar scaling behavior for the $W^{-1/2,2}$-norm, etc.).

[25]That this occurs temporally is, of course, what makes motion continuity possible in TV and movies; similarly, the success of image presentations with finite resolution cannot reflect a band limitation occurring in the physical world (as might be a quite naive interpretation in reading the statement of the Shannon–Whittaker Sampling Theorem), but rather a sensitivity cutoff in human perception. This actually takes the form of characteristic visual angles, partly adjustable by 'focus of attention', so the actual lengths are quite different for a movie projection screen as compared to a picture in a book. [We might also note the iconic correlation between importance and apparent size.]

[26]For $p = 2$, the L^2 isometry of the Fourier transform does just convert such scales to localization in the frequency domain — but we have already rejected the mathematically convenient L^2 metric as an incorrect description of human perception.

[27]I.e., in the intensity function f (or, as we shall see, in a smoothed version of f) — although, consistent with an interpretation of edges in terms of occlusion, one might also want to consider, e.g., abrupt changes in textures (changes in the distribution of the [locally normalized] energy with respect to refined scale and orientation: this (and other comparatively delicate approaches) may be needed to infer (part of) a separating boundary when there is little difference in the illumination levels of the adjacent regions: a good edge detection theory should reconstruct missing segments as connections or prolongations of curves already identified.

examples, we note that the graph of $f(x) = \tan^{-1} 100x$ looks like that of the *signum* function $\operatorname{sgn}(x) := \{-1 \,\text{for}\, x < 0;\, +1 \,\text{for}\, x > 0\}$, with a 'discontinuity' which we locate at the inflection point ($x = 0$, where $f'' = 0$) and an 'associated scale' (interface width — where the gradient f' remains large) roughly corresponding to the distance between the nearby maxima of $|f''|$; by comparison, we may think of $\tan^{-1} x$ as also having a 'discontinuity', but now on a scale 100 times as coarse. A consideration of the graph of $f(x) = \tan^{-1} 100x + \sin 100\pi x$ suggests that we may think of this as a 'texture' $\sin 100\pi x$ superposed on (approximately) $\operatorname{sgn} x$ so the underlying discontinuity (edge) shows up at $\{\hat{f}'' = 0\}$, where $\hat{f}$ is a smoothed[28] version of f, e.g., convolution with a Gaussian having standard deviation larger than .01; one could equally well treat a one-sided additive texture, e.g., replacing $\sin 100\pi x$ by $\{0 \,\text{for}\, x < 0; \sin 100\pi x \,\text{for}\, x > 0\}$. Similar considerations arise, of course, for 2-D visual images, although one then can make a distinction in treatment between filtering out patterned oscillations parallel to the edge — equivalent to those arising in the 1-D case above — and those orthogonal or oblique to the edge, with the possibility that the latter may also involve new edges as intersections or T-junctions. We see the relevant consideration, then, as involving the appropriate filtering for selective smoothing — closely relating this, as we have noted earlier, to the possibilities of compression.

We now expand on the previous mention, following (3.1), of filtering by diffusion. Most simply, the heat equation

$$u_t = D\Delta u \qquad \left(u\Big|_{t=0} = f\right) \tag{3.2}$$

has well-known smoothing effects: essentially it is a low-pass filter, convolving f with a Gaussian of variance Dt. Observing that

$$u(t, x) := \exp\left[-D|\omega|^2 t\right] \sin \omega \cdot x$$

One certainly also wishes to include free-standing curves, without change across them but corresponding to a narrow strip involving intensity differences much greater than oscillations in the adjacent regions.

[28]Once the effective point of discontinuity is found, it becomes possible to separate the scales (here the same .01) associated with the texture and with the underlying 'image' signal, since by definition the texture extends well outside the interface. Related considerations apply if one has superposition of noise along with texture; this is closely related to so-called 'change point' analysis in Statistics (cf., e.g., [12] for a statistical treatment of visual images from this point of view).

is a solution with spatial scale $\lambda := |\omega|/2\pi$, we see that we get exponential decay of features with scale λ at a rate of exponential decay of order $D\lambda^2$ or, conversely, that the 'disappearance' of spatial scales λ in the diffusion (3.2) has a rate $\lambda \sim \sqrt{Dt}$. We note that (3.2) is exactly the 'gradient flow' with speed D along[29] steepest descent in function space for the Dirichlet energy $\|\nabla u\|^2_{L^2(\Omega)}$. We will next wish to improve on this simple linear filtering.

A first interesting/useful modification of (3.2) would be through the replacement:

$$\Delta u \longleftarrow |\nabla u| \nabla \cdot \left(\frac{\nabla u}{|\nabla u|} \right) =: \Delta_\perp u. \tag{3.3}$$

Here $\Delta_\perp$ indicates the lower-dimensional Laplace operator associated with the subspace orthogonal to the gradient direction ∇u and the use of

$$u_t = D \Delta_\perp u \qquad \left(u\Big|_{t=0} = f \right) \tag{3.4}$$

provides anisotropic diffusive filtering — precisely in the directions parallel to level curves of u. One also gets some smoothing of the edges to be detected, related to the geometric interpretation which makes (3.4) so interesting (and extensively studied): this is known as the 'mean curvature equation' since level curves 'propagate' with normal motion proportional to mean curvature.

A second modification of (3.2) would be the use of a non-constant diffusion coefficient D. The version proposed by Perona and Malik [21] takes $D = D(|\nabla u|)$ with $D(\cdot)$ positive but decaying. This is intended to retain the sharpness of edges, where $|\nabla u|$ is comparatively large, while rapidly filtering out noise (and fine-scale variation) elsewhere. Interpreting this in divergence form, one has

$$u_t = \nabla \cdot D(|\nabla u|) \nabla u \qquad \left(u\Big|_{t=0} = f \right) \tag{3.5}$$

and we note, e.g., that there are known existence results[30] when $sD(s)$ is nondecreasing on $\mathbb{R}_+$. Unfortunately, use of such a gradient-dependent diffusion coefficient seems quite sensitive to noise (e.g., for the relevant small

[29] This assumes the standard duality map: $[f, g] \mapsto \int_\Omega fg$ and modification of this would lead to modification of (3.2). For example, as noted by P. Fife, use of a $W^{-1,2}$-duality would lead to the Cahn–Hilliard equation, arising [4] in the metallurgical theory of phase transition (solidification).

[30] The analysis in [23], [24] indicates that (3.5) should determine $D\nabla u$ stably with respect to an L^p metric if $s^2 D(s)$ grows like s^p $(p \geq 1)$, which here suggests the appropriate topology for edge detection, etc. Taking D decreasing as rapidly as this would make

t corresponding to scales of interest), [6] suggested the use of $D = D(|\nabla \hat{u}|)$ where $\hat{u}$ is a smoothed version of $u_0 = f$, such as the solution at time τ for (3.2).[31] While there seems to be no theory for such equations, anisotropic diffusive filtering in the form[32]

$$u_t = \nabla \cdot D(|\nabla \hat{u}|)\nabla u \qquad \left(u\Big|_{t=0} = f\right) \tag{3.8}$$

seems to work well in practice, even for quite rapidly decaying $D(\cdot)$ — e.g., [31] successfully uses $D(s) := \exp[-(1/5)(s/\lambda)^5]$ and obtains quite impressive-looking pictures for the examples considered there. It is pointed out in [18] that, as with (3.2), each of these variations is also related to a notion of energy minimization; for example, (3.6) corresponds to gradient flow for minimization of the $W^{1,1}$-seminorm $\|\nabla u\|_{L^1(\Omega)}$.

We observe at this point the close relation of these approaches to the mathematical modeling of phase transition in solidification. In each case one has a balance of minimization between terms penalizing rapid variation concentrated at an interface[33] and penalizing differences from the initial state or original image. As with (3.1), the Cahn–Hilliard and Allen–Cahn equations involve a small parameter, related to the interface thickness. Mathematical analysis has shown (This has been especially developed in the related 'phase field model') that this evolution tends to motion by mean curvature of a sharp

---

possible (i.e., $p = 1$ so $D(s) = 1/s$; note that the form (3.4) with variable D and the divergence form of (3.5) are then identical), one obtains the Osher-Rudin version [20]

$$u_t = \nabla \cdot \left(\frac{\nabla u}{|\nabla u|}\right) \tag{3.6}$$

of the mean curvature equation, related to the propagation of shock waves in fluid dynamics. Rather generally, the use of (3.5) with decreasing $D(\cdot)$ will provide some of the same anisotropic diffusive filtering as does (3.4).

[31] Presumably one could take account of local normalization in an analogous way, but this variation seems not to have been tried.

[32] Combining some of these ideas, [1] proposed the form

$$u_t = D(|\nabla \hat{u}|)\Delta_\perp u \tag{3.7}$$

for this diffusive filtering.

[33] For solidification, this is related to the surface component of the Gibbs free energy, typically expressed as 'surface tension'. The geometric fact that mean curvature is the derivative of surface area (curve length) with respect to normal motion seems to underly the significance of curvature for regularization.

front as this parameter goes to 0. The extensive study of the Cahn–Hilliard equation has also led to a rather detailed understanding of this evolution as a 'coarsening' of the segmentation. A particularly interesting effect is the development of metastable states (not genuine steady states, but persisting for extremely long times on the scale of the evolution); it would be interesting to know whether this can occur in some of these edge detection models — and, especially, whether this might correspond to real effects ('optical illusions') in human perception.

4. A proposed compression scheme

Based on the discussion above, we propose a novel compression scheme for visual images which seems to have several advantages. This scheme has two stages with adaptively determined cutoffs. We may label these stages, which we will then describe[34] and justify in a bit more detail, as *segmentation* and *residual compression*. The purpose of the first stage is to obtain a partition of the image — primarily by edge detection and anisotropic diffusive filtering at decreasing scales. Note that the first part of the compressed coded would (automatically?) be ordered by the decreasing scale of analysis so that an initial segment would be effectively a cartoon line drawing of the image, which could serve as a key for archival retrieval. As discussed, this should already carry a large part of the significant image information.

Then, in 'patches' aligned with the 'edge skeleton' created in the first stage, the 'residual' significant information is to be compressed by a wavelet expansion approach corresponding to that used globally in [7]. We note that the real importance, for [7], of the specification of the L^1-norm for measuring degradation of picture quality lies in the quantization of expansion coefficients — one retains fewer significant bits as one moves to finer scales, with a relation here tailored to the L^1-norm. What enables DeVore, Jawerth, and Lucier to obtain (good) estimates for the compression rate, is the compactness of the embedding of the Besov space $B^\alpha_q(L^q)$ into L^1. The choice of such a Besov space corresponds to a specification/assumption about the smoothness of the intensity function f, with better estimates available under stronger smoothness assumptions. In [7] there is a fundamental limitation on the smoothness which might reasonably be assumed for the image as a

[34] Fuller detail and some suggestions for implementation are necessarily deferred to [27].

whole, since one does expect discontinuities (edges) in the images under consideration. On the other hand, since we would only be using this compression technique in the segmented regions, we may note that f (as restricted to any such region) will be much smoother than could be globally true — simply by the definition/construction of the segmentation.

This is, of course, only an indication of the nature of the compression scheme and much work remains between this description and any implementation. Nevertheless, I would consider this as a promising approach to compression within the context of single gray-scale 'natural' visual images.

References

[1] L. Alvarez, P.-L. Lions, and J.M. Morel, *Image selective smoothing and edge detection by nonlinear diffusion,* pp. 845–866, SIAM J. Numer. Anal. **29** (1992).

[2] D.H. Ballard and C.M. Brown, *Computer Vision*, Prentice–Hall, Englewood Cliffs, 1982.

[3] G.C. Brock, *Image Evaluation for Aerial Photography,* Focal Press, London, 1970.

[4] J.W. Cahn and J.E. Hilliard, *Free energy of a nonuniform system, I: Interfacial free energy*, J. Chem. Phys. **28**, pp 258–267, (1958).

[5] P.E. Castro, personal communication (Nov., 1994).

[6] F. Catté, T. Coll, P.-L. Lions, and J.M. Morel, *Image selective smoothing and edge detection by nonlinear diffusion,* SIAM J. Numer. Anal. (1991).

[7] R. DeVore, B. Jawerth, and B. Lucier, *Image compression through wavelet transform coding,* pp. 719–746, IEEE Trans. Inf. Theory [Special issue on wavelets] **38** (1992).

[8] L.C. Evans and Gariepy, *Lecture Notes on Measure Theory and Fine Properties of Functions,* CRC Press, 1992.

[9] R. von der Heydt, E. Peterhans, and G. Baumgartner, *Illusory contours and cortical neuron responses,* pp. 1260–1262, Science **224** (1984).

[10] D.H. Hubel and T.N. Wiesel, *Receptive fields, binocular interaction and functional architecture in the cat's visual cortex,* pp. 106–154, J.Physiol. **160** (1962).

[11] G. Kanisza, *Organization in Vision,* Praeger, NY, 1979.

[12] A.P. Korostelev and A.B. Tsybakov, *Minimax Theory of Image Reconstruction,* (Lecture Notes in Statistics #82) Springer-Verlag, New York, 1993.

[13] J. LeMoigne, *Parallel registration of multi-sensor remotely sensed imagery using wavelet coefficients*,
CESDIS report, NASA Goddard, 1994.

[14] S.Mallat and S. Zhong, Complete signal representation with multiscale edges, Robotics report # 219, CIMS–NYU, 1989.

[15] S.Mallat and W.L. Hwang, *Singularity detection and processing with wavelets*, pp. 617–643. IEEE Trans. Inf. Theory [Special issue on wavelets] **38** (1992).

[16] D. Marr, *Vision,* Freeman, 1982.

[17] Y. Meyer, *Un contre-example a la conjecture de Marr et a celle de S. Mallat,* preprint, 1991.

[18] J.-M. Morel and S. Solimini, *Variational Methods in Image Segmentation,* Birkhäuser, Boston, 1995.

[19] D. Mumford and J. Shah, *Optimal approximations by piecewise smooth functions and associated variational problems,* Comm. Pure Appl. Math. **XLII**, (1989).

[20] S. Osher and L. Rudin, *Feature-oriented image enhancement using shock filters,* pp. 919–940, SIAM J. Numer. Anal. **27** (1990).

[21] P. Perona and J. Malik , *A scale space and edge detection using anisotropic diffusion*, in *Proc. IEEE Com. Soc. Workshop on Computer Vision*, IEEE, 1987.

[22] A. Rosenfeld and A.C. Kak, *Digital Picture Processing* (vol. 1), Academic Press, Orlando, 1982.

[23] T.I. Seidman, *A class of nonlinearly elliptic problems,* pp. 151–173, J. Diff. Eqns. **60** (1985).

[24] T.I. Seidman, *Continuous dependence for a degenerate parabolic equation,*
J. Diff. Eqns. **70**, pp. 293–308 (1987).

[25] T.I. Seidman, *Some 'complexity' issues for ill-posed problems,* pp. 17–35 in *Inverse Problems in Partial Differential Equations* (D. Colton, R. Ewing, W. Rundell, eds.), SIAM, Philadelphia, 1990.

[26] T.I. Seidman, *Interpretations and estimates for ill-posed problems*, pp. 108–119 in *Proc. St. Wolfgangsee Conf. on Parabolic Ill-Posed Problems* (H. Engl, W. Rundell, eds.), SIAM, Philadelphia, 1995.

[27] T.I. Seidman, *Optimal Data Compression of Visual Images,* final version in preparation.

[28] P.C. Teo and D.J. Heeger, *Perceptual image distortion*, pp. 127–141, Proc. SPIE **2179** (1994).

[29] A.N. Tikhonov and V.I. Arsenin, *Solutions of Ill-Posed Problems,* Winston, DC, 1977.

[30] F.M. Wahl, *Digitale Bildsignalverarbeitung,* Springer-Verlag, Heidelberg, 1984. (English trans., Artech House, Boston, 1987.)

[31] J.Weickert, *Anisotropic diffusion filters for image processing based quality control,* pp. 355–362 in *Proc.* 7^{th} *Conf. on Math. in Industry,* (A. Fasano and M. Primicerio, eds.) Teubner, Stuttgart, 1994.

[32] S. Zhong, *Edge representation from wavelet transform maxima,* Ph. D. dissertation, NYU, 1990.