How suitable is structure tensor analysis for real-time color image compression in context of high quality display devices

Fritz Lebowsky, Mariano Bona; STMicroelectronics; Grenoble, France

Abstract

High-end PC monitors and TVs continue to increase their native display resolution to 4k by 2k and beyond. At the same time, high dynamic range formats demand higher bit depth for the underlying color component signals. Subsequently, uncompressed pixel amplitude processing becomes costly not only when transmitting over cable or wireless communication channels, but also across on-chip image processing pipelines that access external memory units. We recently presented a block-based non-linear memory compression architecture for text, graphics, and video enabling multi-dimensional error minimization with context sensitive control of visually noticeable artifacts. The underlying architecture was constrained to a small block size of 4x4 pixels. To increase compression ratio as well as image quality, we propose a novel approach that converts image amplitudes into a pair of discrete structure and magnitude quantities on a pixel by pixel basis which has been inspired by structure tensor analysis. Graceful degradation of image information is controlled by a single parameter which aims at optimally defining sparsity as a function of image context. Furthermore, we apply error diffusion via a threshold matrix to optimally diffuse the residual coding error. A detailed error distribution analysis and comparison with our previous algorithms highlights the effectiveness of our new approach, identifies its current limitations with regard to high quality color rendering, and illustrates algorithm specific visual artifacts.

Introduction

Describing challenging engineering problems in tensor space has most often led to superior overall system performance most often due to modeling the underlying natural, physical, or biological phenomena in multidimensional differential parameter space. Therefore, we hypothesize a beneficial outcome of using the power of tensor analysis in image processing algorithms that target optimal adaptation to human vision and human visual perception, especially for video streams composed of natural scene content. Moreover, since color imaging is very often based on concurrent but independent multidimensional color component processing, we may also more easily observe critical differential errors between color components. As a result, highly suitable control and correction methods can be preciously elaborated in local and global context of image data.

Unfortunately, mastering tensor algebra demands understanding higher levels of functional abstraction than linear algebra. Investing in such additional mental effort can often be circumvented, especially in image processing, where simple amplitude quantities across a regular 2-dimensional (spatial) or 3-dimensional (spatio-temporal) pixel matrix constitute the input domain as well as the final output domain. Moreover, reaching a higher level of functional abstraction is very often linked to having to apply an increased number of operations per pixel which becomes prohibitive in most consumer devices that need to well minimize processing bandwidth and power consumption. For example, when looking at structure tensors, we already have to transform a single pixel amplitude quantity into a structure quantity and a magnitude quantity that operate concurrently on a pixel by pixel basis.

In a more general context, we are interested in specifically developing block based image compression architectures that enable context sensitive control of visually noticeable artifacts as a function of compression factor. Compared to well-known image compression methods such as JPEG, MPEG, or HEVC standards, which can achieve high compression factors, we aim at low compression factors in the range between 1.5 and 4. However, the fundamental architectural challenge arises from an implementation at a fraction of the cost of well-known compression methods.

So, shouldn't we stop dead right there when it comes to competitive performance of real-time consumer applications? Maybe not, especially when having grasped the unique opportunity of combining long term engineering experience of an applied mathematician (who had never worked on image processing tasks before) and a system architect dedicated to exploring advanced nonlinear functionalities in computational image processing. There are several intuitively predominant enigmas waiting to be lifted: (1) how useful are structure quantities towards avoiding some of the most common visually noticeable artifacts; (2) how well can a nonlinear structure & magnitude descriptor generate an optimally sparse representation of visual image information; (3) how well can the underlying computational method be optimized for highly efficient visually adaptive processing in local context; (4) how gracefully image degradation behaves visually as a function of image data compression factor; (5) how complex the emerging computational method becomes in terms of number of operations per pixel; (6) in view of all of the above, what are the most interesting publications in research and development addressing tensor based image processing tasks.

The breadth of structure tensor based image compression proposals address scenarios of high compression effort with compression factors in the range of 10x to 100x. In such scenarios, structure tensor processing on the encoder side is preceded by a Gaussian kernel that filters noise. Structure tensor processing on the decoder side is followed by an interpolation method mostly based on iterative total variation. The achievable image quality across low compression factors remains unknown and the overall computational complexity appears out of reasonable range.

Highly relevant foundations of our current problem statement have been described in [1]. Structurally adaptive low pass and high pass filtering methods use nonlinear structure tensors in context of local pixel neighborhood. Another interesting use case in temporal domain has been presented in [2]. A structure tensor framework serves as a detector of motion structure components in video sequences facilitating blind video quality evaluation in wavelet domain.

In more general terms and challenges, image texture and noise regions are considered as highly difficult structures to be compressed. Subsequently, more specific solutions have been proposed [3] that offer excellent visual image quality at high compression ratios, but with elevated processing cost. Unfortunately, in view of low compression factors, such solutions appear hard to be scaled down to obtain highly reduced computational complexity.

Continuing from the point of view of computational complexity, a first proof of concept and successful implementation has been described in [4]. The architectural simplicity was based on parametric exploration of a nonlinear domain using sorting of pixel amplitude values within a block of 4x4 pixels. The nonlinear system behavior enabled randomized spreading of residual error amplitudes. Such spatially and temporally randomized error amplitudes are less visually noticeable as long as medium and high error amplitudes remain sparse. Without significant increase of algorithmic complexity, a new error minimization strategy improved PSNR rating up to 12 dB [5]. Encouraged by the achieved high performance in image quality with compression factors in the range of 2x, we also became highly interested in exploring the potential of nonlinear structure tensor analysis.

In the remaining sections we first discuss fundamentals of discrete structure & magnitude processing inspired by analytical structure tensor processing, followed by a demonstration of the novel nonlinear method's efficiency towards image data compression which includes analysis of significant distribution functions as well as key ideas of adaptation to various types of image content. Furthermore, we introduce block based advanced scan path processing as well as block based advanced color component processing. Finally, we draw a conclusion and envision related work still ahead.

Fundamentals of discrete structure & magnitude processing inspired by analytical structure tensor processing

Before we discuss details of our novel approach of representing image information in discrete quantities that demonstrate surprisingly high precision and highly reduced computational effort, we review the underlying aspects of analytical structure tensor processing, since it provided us with a significant level of inspiration that could not have been leveraged otherwise.

Analytical structure tensor processing

In search of a simple and cost efficient computational method targeting image data compression, we were interested in better understanding some of the key properties of structure tensors. Thus we simply formulated the structure tensor in context of a spatial neighborhood spanning just across 2x2 pixels.

Eq. 1 and Eq. 2 summarize the fundamental mathematical relationship between image amplitudes $f(x,y)$ at positions relative to the actual center position (x_i, y_j) and their structure tensor (Eq. 2) to calculate the associated eigenvalue and eigenvector angle which can be derived from Eq. 2, whereas $\delta f(x_i, y_j)/\delta y$ represents the local image gradient amplitude in vertical direction and $\delta f(x_i, y_j)/\delta x$ represents the local image gradient amplitude in horizontal direction.

$$\begin{aligned} \delta f(x_i, y_j)/\delta y &= f(x_i, y_j) - f(x_i, y_{j-1}) \\ \delta f(x_i, y_j)/\delta x &= f(x_i, y_j) - f(x_{i-1}, y_j) \end{aligned} \quad (1)$$

The eigenvalue of the matrix equation (Eq. 2) represents the gradient magnitude of its vector norm and the eigenvector angle θ is defined as the angle between the x axis and the vector norm as illustrated in **Figure 1**, where for example the contour of a disk segment (red) is being tracked.

$$\begin{vmatrix} \left(\delta f(x_i, y_j)/\delta x\right)^2 & \delta f(x_i, y_j)/\delta y * \delta f(x_i, y_j)/\delta x \\ \delta f(x_i, y_j)/\delta y * \delta f(x_i, y_j)/\delta x & \left(\delta f(x_i, y_j)/\delta y\right)^2 \end{vmatrix} \quad (2)$$

At each pixel position we are primarily interested in the orientation of the contour's isophote, ideally showing a value equal to zero. We are also interested in estimating the maximum gradient magnitude in that location which doesn't necessarily always coincide with the vector norm (showing perpendicular to the isophote orientation). For example, some types of image objects like corners may generate ambiguity. We will discuss such scenario in more detail in the section where we focus on specifics of discrete processing of structure components.

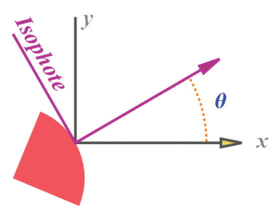

Figure 1. Illustration of relationships between Cartesian coordinates (x,y), the vector norm, its vector angle θ, and its isophote in reference to a contour of a disk segment (red).

Figure 2 shows an example test image of the Kodak image database together with the angle θ of the eigenvector of the structure tensor as described in Eq. 2. The angular quantities θ in the range of 0 to 2π have been converted to hue values in the range of Modulo[θ/2π,1] on a pixel by pixel basis. It may already become evident that this structural information appears surprisingly sufficient for the human visual system to decode image structure to the finest level, without any accompanying magnitude quantities. Consequently, we hypothesize that image structure quantities take precedence over gradient magnitude quantities which themselves take most likely precedence over image amplitude quantities. In other words, the HVS hardly cares about absolute luminance or chrominance values as long as the image content – presented by structure information – remains comprehensible.

Because image gradient magnitude quantities seem less important than image structure quantities, we now pursue the idea of compromising precision of image gradient magnitudes while primarily preserving image structure quantities. Although the structure quantity cannot be considered independent of the precision of the gradient magnitude quantity, we can accept higher level of error tolerance on gradient magnitude quantities than image structure quantities, especially in context of non-stationary local neighborhood. So we want to firstly extract and preserve image structure information on a pixel by pixel basis while secondly minimizing the number of gradient magnitude quantities as a function of ***acceptable*** image quality.

Besides, in the context of a predictor of image structure quantities, the structure tensor outperforms the well-known Median-Edge-Detector. The structure tensor method remains consistent across the entire frequency spectrum in terms of phase response while, for example, the median of 3 adjacent pixel amplitudes imposes a severe phase change around 1/3 of its sampling frequency.

Figure 2. A representative high quality test image (left) and its image structure quantities visualized as hue values representing the eigenvector angles at full spatial resolution (right).

Before we proceed with the description of our novel discrete representation of image structure quantities and gradient magnitude quantities, we would like to address the dilemma of extrapolation in image amplitude domain followed by an elegant solution that emerges from within the gradient magnitude domain.

Dilemma of extrapolation

Image amplitude prediction by extrapolation remains suboptimal in context of linear filtering when using linear convolution kernels in temporal domain,. Figure 3 illustrates the extrapolation scenario. A common solution to this dilemma is based on convolutional high pass filtering by considering several image amplitude samples within the neighborhood of the center pixel position (x_i, y_j). However, if we would like to predict the amplitude **APR** of the center pixel with just the image amplitude **APA** of the closest pixel position to the left and the image amplitude **APB** of the closest pixel position to the right of the center position, the solution appears impossible to be formulated in amplitude domain with only relative quantities that should be constraint to the normalized range of $[\mathbf{0} \dots \mathbf{1}]$, where this range represents the distance $|\boldsymbol{APA} - \boldsymbol{APB}|$. We would like to recall that the normalized range always holds in the interpolation scenario, where the amplitude **APR** is less or equal to **APA** and greater or equal to **APB**.

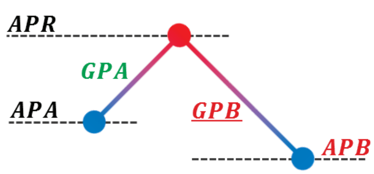

*Figure 3. Illustration of extrapolation dilemma in amplitude domain when considering normalized range of $[\mathbf{0} \dots \mathbf{1}]$, equivalent to the absolute amplitude difference between **APA** and **APB**.*

Solution in gradient magnitude domain

We overcome the dilemma of extrapolation by implementing the following solution. At first, we calculate the gradient magnitudes $\boldsymbol{GPA} = |\boldsymbol{APR} - \boldsymbol{APA}|$ and $\boldsymbol{GPB} = |\boldsymbol{APR} - \boldsymbol{APB}|$. Secondly, we create a polarity bit **POL** that indicates whether we process an interpolation scenario or an extrapolation scenario. Thirdly, we pick the minimum of **GPA** and **GPB** and normalize it using the sum of the distances **GPA** and **GPB**, therefore obtaining the normalized minimum gradient magnitude quantity **GPAN**, as summarized in Eq. 3. We also introduce an error quantity **ERR** which represents the degree of error tolerance (maximum error) to the quantization step that will be discussed in the section dedicated to efficiency towards image data compression. Please also note that, by our own convention, **GPA** will always represent the minimum gradient magnitude in local context. **Figure 4** illustrates the solution to the extrapolation scenario shown in Figure 3. Here we convert the amplitude of **APB** to **APR+GPB**. This step is seamlessly reversible with the help of '*memorizing*' the associated polarity information.

$$\boldsymbol{GPAN} = \frac{\boldsymbol{GPA}}{\boldsymbol{GPA+GPB}} = \mathbf{1} - \frac{\boldsymbol{GPB}}{\boldsymbol{GPA+GPB}} \pm \boldsymbol{ERR} \qquad (3)$$

The smaller the **GPAN** quantity becomes, the better we are approximating the 'ideal' isophote in relative quantities of pixel based local context, or, if looking at **GPA** and **GPB** concurrently, the larger **GPB** manifests itself as well. Therefore, if we search for the minimum gradient magnitude and the maximum gradient around the center pixel value to be predicted, we also seem to be using a powerful descriptor of local contrast that may easily enable minimizing the visibility of a voluntarily acceptable maximum prediction error **ERR** when remaining proportional to the just-noticeable-difference (JND) quantity for example. Subsequently, perceptual quantization tone curves (mean local luminance) together with perceptual contrast sensitivity (local variance) (see ref [6]) may enable excellent guidance for minimizing visibility of local errors resulting from a digitally encoded (quantized) representation of **GPAN**, which we define as the **GPAC** quantity.

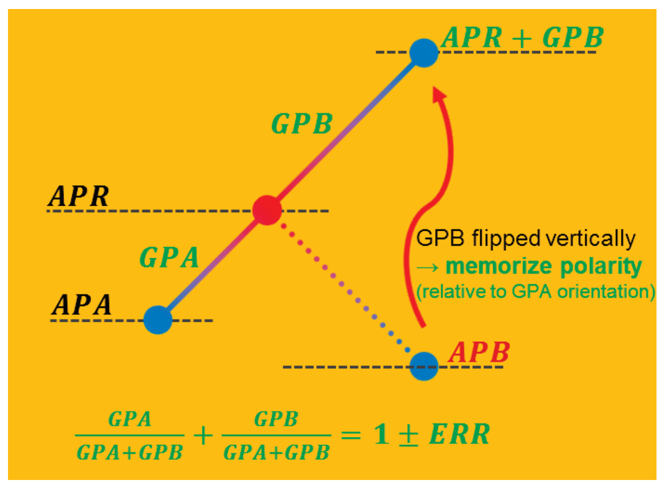

*Figure_4. Illustration of solving the extrapolation dilemma in gradient magnitude domain when considering normalized data range of $[\mathbf{0} \dots \mathbf{1}]$, equivalent to the sum of distances **GPA** and **GPB**.*

Discrete processing of structure components

Instead of estimating the eigenvector angle representing the isophote at high angular resolution, we would like to elaborate the idea of simply defining several structure elements which represent basis vectors within a local 3 by 3 neighborhood. To this end, **Figure 5** illustrates two fundamental basis vectors. The one on the left constitutes the well-known Cartesian grid (coordinate system) while the one on the right constitutes the Cartesian grid rotated by

45 degrees with a new basis of $\sqrt{2}$ in reference to the unit spacing of the underlying pixel grid. To encode the selectable basis vectors pixel by pixel, a single bit is sufficient. We select the basis vectors that contain the minimum gradient magnitude and encode on which 'branch' of the basis vectors it is located by designating a second bit. Finally, we add a third bit that describes the polarity **POL** of the gradient magnitude only when the encoded gradient magnitude **GPAC** is greater zero. Summarizing this simple but surprisingly efficient structure scenario, we will spend at most 3 bits per pixel to fully encode a local structure component composed of basis vectors, the relative location of the minimum gradient magnitude, and the minimum gradient magnitude's polarity.

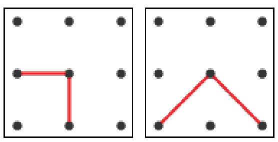

Figure 5. Illustration of a set of two basis vectors enabling discrete orientation identification of most suitable minimum and maximum gradient magnitude pairs in local neighborhood.

Taking the idea a step further for estimating the very best pair of minimum and maximum gradient magnitude in a local neighborhood of 3 by 3 pixels we could theoretically choose from a set of 28 basis vectors as shown in **Figure 6**. Each row shows a specific curvature with all possible orientations within each row. The set of curvatures also highlights that we are not at all restricted to orthonormal basis vectors, whether operating in 2-dimensional spatial domain or in vector spaces of higher dimensionality. We recently provided a theoretical performance analysis in the context of vectorized linear interpolation [7]. However, due to sequential processing constraints, such as raster scan mode, only the use of a 'single sided' subset of these 28 basis vectors remains practical. With all this in mind, one could also add more basis vectors by extending the local neighborhood and using any suitable location of already encoded pixels. In other words, we propose using the discrete pixel grid to optimally define discrete structure quantities in local context. In more general terms, we recommend reusing the already existing discrete grid spacing to ideally define most suitable basis vectors on a pixel by pixel basis.

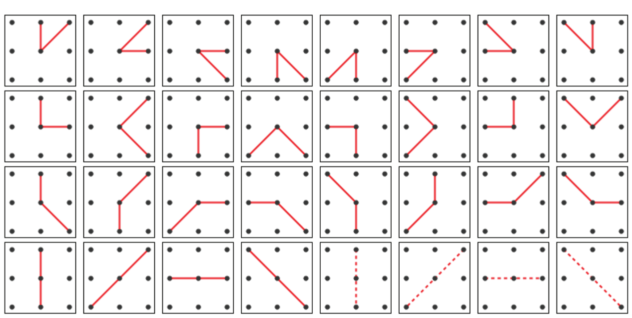

Figure 6. Illustration of a set of 28 basis vectors enabling discrete orientation & curvature identification of most suitable minimum and maximum gradient magnitude pairs in local 3 by3 pixel neighborhood.

Efficiency towards image data compression

At first we would like to point out that the image data compression is currently defined by a single parameter which we baptized target threshold parameter **TTP**. **TTP** controls the compression effort - to be increased or decreased - as a function of number of normalized minimum gradient values below threshold, for example calculated on a frame by frame basis of a video stream to be compressed. In other words, the more normalized minimum gradient quantities reach below TTP, the higher the overall compression factor becomes.

After having carried out careful simulation analysis across a significant number of natural images, we present our simulation results obtained from a most representative high quality test image of the Kodak image data base.

Simulation results – obtained from a representative set of example images

In search of demonstrating reasonable robustness in the presence of noise, we added 40dB of white Gaussian noise to the selected test image (fig. 2). Although the PSNR metric is strongly debatable as an absolute image quality metric, it appears well suitable for algorithm performance analysis in relative quantities. **Figure 7** illustrates the compression performance in PSNR quantities over a significant range of compression factors. The lower bound curve represents a simple version of the novel compression method, where each color component has been encoded independently. The upper bound curve represents an advanced compression scheme in which all three color components shall be processed concurrently with only one most suitable structure quantity per pixel to be transmitted.

The graph reveals three highly important performance features: (1) asymptotically approaching the maximum PSNR value towards a compression factor equal to one, (2) highly linear PSNR variation at high compression factors, and (3) less than 5 dB reduction in PSNR rating over more than 5x range of compression factors. Please also note that this range of compression factors is based on images having 8 bit pixel amplitude resolution. Interestingly, the compression factor will scale nearly linearly with pixel amplitude resolution. For example, images with 10 bit pixel amplitude resolution would achieve 25% higher compression factors when compared with 8 bit images.

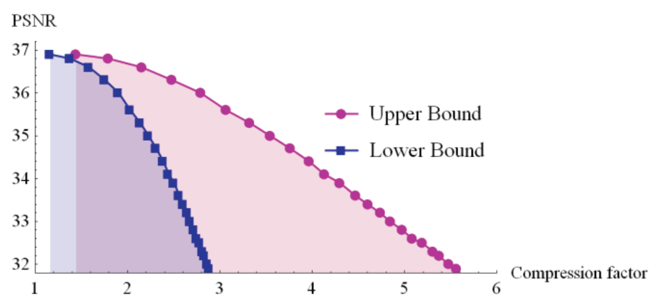

Figure 7. Compression performance in PSNR quantities of test image shown in fig. 2 with additive WGN of 40dB. The lower bound curve depicts a suitable range of compression factors when assigning structure quantities to each color channel separately while the upper bound curves Illustrates a suitable range of compression factors when assigning a single structure quantity to all three color channels.

For the analysis that follows, we chose a moderate compression factor having a lower bound of 2.3x, respectively upper bound of 3.8x with a PSNR rating of 34.7dB which shall demonstrate visual and statistical performance in view of compression factors targeted for real time compression.

Figure 8 compares the original image with the compressed version. The compressed image was obtained by calculating structure quantities for each color component separately, leading to a compression factor of 2.3x. The image pair on the left shows the reference image together with the red channel's hue coded eigenvector angle (leftmost position). The image pair on the right shows the compressed image and the red channel's hue coded eigenvector angle with a compression factor of 2.3x (rightmost position). The reduced variability of eigenvector angle in local context manifests itself in dominant hue values across the entire image.

We would like to also emphasize that the novel nonlinear method did not yet reveal any annoying visual artifacts even at high compression factors. We hypothesize that this is due to graceful pruning of prioritized local contrast quantities in context of visual masking. The graceful performance seems to be specifically revealed since we eliminated (1) visibility of ***blurriness*** (by escaping convolutional linear filtering), eliminated (2) visibility of ***blockiness***, eliminated (3) visibility of ***mosquito noise,*** and (4) minimized visibility of ***contouring***.

Figure 8. Compression performance of a detail region from test image shown in fig. 2. The image pair on the left shows the original and its hue coded eigenvector angle quantities (red channel); the image pair on the right shows the compressed version and its hue coded eigenvector angle quantities with a compression factor of 2.3 (lower bound) or 3.8 (upper bound - estimated).

Figure 9 visualizes the red color component's discrete structure quantities of the minimum gradient magnitude as well as the maximum gradient magnitude converted to color hue quantities. The image pair on the left shows the discrete structure components of minimum (leftmost position) and maximum gradient magnitude quantities without compression; the image pair on the right shows the compressed version of the discrete structure components of minimum and maximum gradient magnitude quantities (rightmost position) with a compression factor of 2.3(lower bound) or 3.8 (upper bound - estimated).

By comparing the compressed version with the uncompressed version we notice that preferred structure orientations appear in compressed mode which manifest itself in dominant global hue, both for the structure quantities of the minimum gradient magnitude as well and the maximum gradient magnitude. As expected, the structure quantities of maximum gradient magnitude show significantly more visually meaningful information.

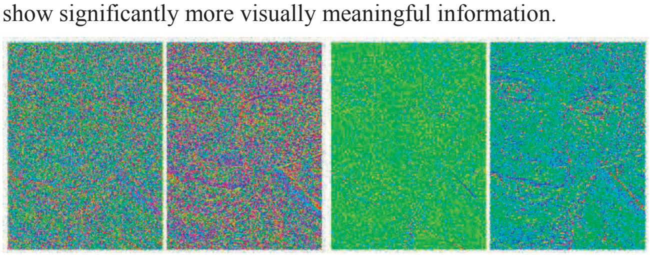

Figure 9. Compression performance illustrated in hue coded discrete structure quantities of a detail region from test image shown in fig. 2. The image pair on the left shows the discrete structure components of minimum and maximum gradient magnitude quantities without compression; the image pair on the right shows the compressed version of the discrete structure components of minimum and maximum gradient magnitude quantities with a compression factor of 2.3 (lower bound) or 3.8 (upper bound - estimated).

Analysis of significant distribution functions

In context of analyzing overall achievable coding efficiency, we created several different distribution functions that enable improved understanding of system functionality from a statistical point of view. Moreover, possible issues can be more easily revealed.

For best comparison results, in **Figure 10** we first recall the cumulative pixel amplitude distribution derived from the red color component of the test image already shown in fig. 8. The revealed S-shaped curve – very typical for natural images – illustrates its challenges with regard to efficient quantization, due to varying slope in the center of the CDF. Most importantly, we need to consider that simple amplitude representation does not sufficiently reflect dominant functionality of human visual perception. The CDFs of image amplitude are practically indistinguishable between the reference image data and the compressed image data. In other words, no significant amplitude errors have been introduced by the underlying nonlinear compression method.

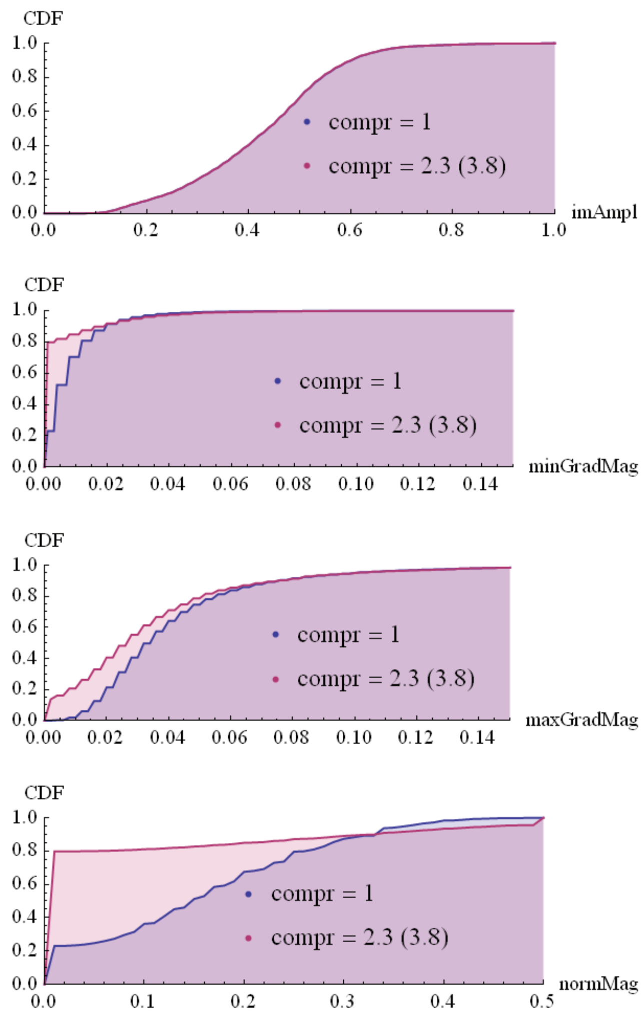

Figure 10. Cumulative distribution functions (CDF) demonstrating compression performance, derived from the red color component of the test image set shown in fig. 8. The blue curve refers to the uncompressed image and the red curve refers to the 2.3x compressed image: CDF of image amplitude (top), followed by CDF of minimum gradient magnitude and CDF of maximum gradient magnitude, and finally CDF of normalized minimum gradient magnitude (bottom).

Next, we compare the CDFs of minimum gradient magnitude. Approximately 20% of minimum gradient magnitude values are already zero in the reference image. However, the compressed image now contains about 80% of minimum gradient magnitude values at zero level. Moreover, CDF progression remains smooth.

The CDFs of maximum gradient magnitude values demonstrate significant spreading of values indicating the presence of important structural image information.

The CDFs of the normalized minimum gradient values reveal the most significant difference between reference and compressed image data which we consider as relevant evidence for proof of concept of our novel nonlinear compression method. As expected, the number of minimum gradient values – obtained via **TTP** thresholding – increase as a function of compression factor. Furthermore, the higher the compression factor the simpler becomes the nonlinear function that approximates the distribution of normalized minimum gradient magnitude values for optimal error minimization & binary encoding efficiency.

Adaptation to various types of image content

Flexible local adaptation to constantly varying image content is key to minimizing visually noticeable artifacts. Therefore, we consider three major categories of critical image content which are (1) homogeneous regions, (2) contour regions, and (3) texture/noise regions. To achieve improved visual image quality, we combined our novel differential error minimization method with a spatial error diffusion technique. However, instead of using advanced locally adaptive error diffusion concepts [8], we simply apply a threshold dither matrix enabling spatial error diffusion with significantly reduced visibility of visually noticeable artifacts versus ordered dither patterns that are more easily susceptible to objectionable moiré phenomena.

Applying adaptive error diffusion

The effectiveness of the dither matrix functionality is currently threefold: (1) we modulate the normalized minimum gradient magnitude quantity; (2) we spatially *seed* homogeneous regions with adequate gradient magnitude values; and (3) we spatially *prune* gradient magnitude values across high density texture and noise regions.

Progressive quantization of normalized minimum gradient magnitude quantities

At first we would like to recall that the image data compression is currently defined by a single parameter which we baptized target threshold parameter **TTP**. **TTP** controls the compression effort - to be increased or decreased - as a function of number of normalized minimum gradient values below threshold, for example calculated on a frame by frame basis of a video stream to be compressed. We advantageously add the normalized minimum gradient magnitude quantity and the dither matrix amplitude that has already been multiplied by the target threshold parameter **TTP** of the normalized minimum gradient magnitude.

Processing homogeneous regions

A dither matrix with 5 bit resolution provides spatial density increments in the order of 3% which appears highly suitable for spatially randomized seeding of normalized gradient magnitude quantities. Seeding in this context is equivalent of retaining normalized gradient magnitude quantities at specific locations defined by the threshold matrix values and the compression effort parameter.

Processing texture & noise regions

The same dither matrix also serves for spatially randomized pruning of gradient magnitude quantities in presence of high density texture or noise regions. Pruning in this context is equivalent of suppressing gradient magnitude quantities at specific locations as defined by the threshold matrix values and the compression effort parameter.

Advanced scan path processing

Until now, we have assumed a traditional raster scan path that progresses line by line over the entire image. However, a scan path which progresses macro block by macro block, using for example a macro block size of 8 lines by 32 pixels per line, appears highly suitable as well. Moreover, within each macro block we envision more elaborate scan paths in meander shape or helical shape. Particularly interesting seems a helical (spiral) scan path which starts in the center of the macro block and progresses to the boundary of the macro block. A small number of initial pixels in the center will be coded in amplitude quantities (providing stability to the nonlinear compression method) followed by structure & normalized gradient magnitude pairs all the way to the outmost pixel location within the macro block. In addition, scan path orientation could also be rotated by 90 degrees from one macro block to the next. **Figure 11** shows two examples of helical scan paths starting from the center position (p1) of the macro block.

p29	p21	p10	p9	p9	p7	p20	p28	p36
p30	p22	p11	p2	**p1**	p6	p19	p27	p35
p31	p23	p12	p3	p4	p5	p18	p26	p34
p32	p24	p13	p14	p15	p16	p17	p25	p33

p33	p25	p17	p7	p6	p5	p16	p24	p32
p34	p26	p18	p8	**p1**	p4	p15	p23	p31
p35	p27	p19	p9	p2	p3	p14	p22	p30
p36	p28	p20	p10	p11	p12	p13	p21	p29

Figure 11. A helical scan path (top) and an alternative helical scan path that has been rotated by 90 degree around the center position p1 (bottom).

Advanced color component processing

Since we do not need to apply spatial filtering before carrying out any compression steps, we can effectively implement separable interleaved sampling. For example, a quincunx raster appears well suitable for the YCC 420 format which is most commonly used in video streaming. Therefore we have one YCC sub-grid (A) combining all three color components (YCC) and two additional Y sub-grids (B) & (C), which simplify managing single structure component quantities across variable number of color components.

Separate processing of (B) and (C) also regularizes selection of relevant pixels in their neighborhood – defined by the available basis vector sets – to calculate local gradient values and their respective normalized minimum gradient magnitude. **Figure 12** illustrates a possible layout of the 3 interleaved grids.

Isn't all the above engineering ingenuity most playful artistic color processing at its best?

(A)	(B)	(A)	(B)	(A)	(B)	(A)	(B)	(A)
(B)	(C)	(B)	(C)	(B)	(C)	(B)	(C)	(B)
(A)	(B)	(A)	(B)	(A)	(B)	(A)	(B)	(A)
(B)	(C)	(B)	(C)	(B)	(C)	(B)	(C)	(B)

Figure 12. Advanced color component processing of YCC420 video format: For optimal compression performance, the 420 sampling grid has been separated into 3 different sub-grids (A), (B),and (C) – see text for details.

Conclusion

The novel nonlinear method enables significantly better local adaptation to non-stationary image information in context of human visual perception while accomplishing low computational complexity and preserving finest levels of structure and magnitude information. Error quantities are being processed in several complementary differential domains enabling advanced visual masking strategies. We also maintain full spectral bandwidth fidelity and optimal phase coherency by having eliminated any linear filtering or median filtering which was traditionally used to improve prediction and coding performance. Needless to say that less filtering also favorably translates into fewer number of operations per pixel. In addition, the new possibility of processing structure quantities and gradient magnitude quantities concurrently but independently enabled adjusting the underlying sparse representation of image information as a function of compression factor. Surprisingly, the associated image quality shows highly stable and relatively proportional to a varying compression factor. First estimates, which however still need to be confirmed, predict that compression factors of about 4x appear very reasonable for 10bit video streams - without creating any visually noticeable artifacts. We presume that the substantially high computational efficiency has been achieved by focusing on minimum and maximum gradient magnitude pairs that are easily derived from discrete pixel amplitudes already present on the underlying grid of the pixel matrix. Such a discrete pair of minimum and maximum gradient magnitude seems to imitate astonishingly well relative local contrast perception of the human visual system. The resulting normalized minimum gradient magnitude quantity therefore offers an excellent solution for representing a powerfully adaptive coefficient in local context.

We also imagine that the presented concept can be advantageously applied to multi-dimensional and multi-scale data sets of many other challenging engineering tasks achieving efficient computational performance and dedicated precision, especially where structural quantities represent important information in local context.

Acknowledgment

The intuitive desire of steadily pursuing the underlying key ideas enabling the discovery of most select discrete multi-dimensional representation of image information in visually meaningful quantities, in quest of beneficially reaching beyond linear system theory, had been a highly challenging endeavor, not only from the engineering point of view, but also across project management and project financing matters. A huge thank you goes to the Electronic Imaging conference community members for their continuous mutual support spanning over more than 25 years. The extremely valuable interdisciplinary exchange of unparalleled scientific and technical knowledge has proven to be the very best platform to create and follow up on exaggerated ideas. May profound gratitude especially reach Bernice Rogowitz, John McCann, Reiner Eschbach, Gabriel Marcu, Alessandro Rizzi, Jan Allebach, Sheila Hemami, Damon Chandler, Mylene Farias, Michael Kriss, Peter Burns, Scott Daly, Al Ahumada, Beau Watson, and Sergio Goma for their fascinating personal interactive engagement as well as wonderfully stimulating and encouraging scientific support over so many years.

Already several years ago, Randolph Fox, a 'retired' colleague at STMicroelectronics, had attached to the new topic in providing magnificent handwritten tutorials on understanding tensor algebra in engineering applications, followed by many highly fundamental technical discussions having enabled fascinating inspiration from structure tensor analysis. Moreover, Mariano Bona's excellent experience in discrete mathematics and control system theory and his limitless curiosity in applying his wonderful knowledge to the domain of digital image processing for the first time - just a year ago - let the discrete solution crystallize. Once again, like the year before, this novel development activity carrying significantly higher pre-estimated risk, could not have been sufficiently explored without Marina Nicola's strong dedication to much more urgent project tasks, in which she succeeded brilliantly. For those reasons, may all three colleagues receive deepest appreciation.

Last but not least, the project would not have been feasible without substantial funding from within the NANO2017 framework. Therefore we would like to also express ultimate gratefulness to Mario Diaz Nava for having offered his extraordinary experience in project funding as well as having early on gained great confidence in the potential project outcome.

References

[1] Thomas Brox, J. Weickert, B. Burgeth, P. Mrázek, "Nonlinear structure tensors," Image and Vision Computing, vol. 24, no. 1, pp. 41-55, 2006.

[2] Michele A. Saad, Alan C. Bovik, "Breaking down the problem of blind video quality evaluation," in Proc. SPIE 9016, Image Quality and System Performance XI, pp. 90160N, 2014.

[3] G. Jin, Y. Zhai, T. N. Pappas, and D. L. Neuhoff, "Matched-texture coding for structurally lossless compression," in Proc. Int. Conf. Image Processing (ICIP), pp. 1065 – 1068, 2012.

[4] Fritz Lebowsky, "Optimizing color fidelity for display devices using contour phase predictive coding for text, graphics, and video content," in Proc. SPIE 8652, Color Imaging XVIII: Displaying, Processing, Hardcopy, and Applications, pp. 86520X, 2013.

[5] Fritz Lebowsky, Marina Nicolas, "Preserving color fidelity for display devices using scalable memory compression architecture for text, graphics, and video," in Proc. SPIE 9015, Color Imaging XIX: Displaying, Processing, Hardcopy, and Applications, pp. 90150M, 2014.

[6] Scott Daly, S. A. Golestaneh, "Use of a local cone model to predict essential CSF light adaptation behavior used in the design of luminance quantization nonlinearities," in Proc. SPIE 9394, Human Vision and Electronic Imaging XX, pp. 939405, 2015.

[7] Marina Nicolas, Fritz Lebowsky, "Optimizing color fidelity for display devices using vectorized interpolation steered locally by perceptual error quantities," in Proc. SPIE 9395, Color Imaging XX: Displaying, Processing, Hardcopy, and Applications, pp. 939502, 2015.

[8] Altyngul Jumabayeva, Yi-Ting Chen, Tal Frank, Robert Ulichney, Jan Allebach, "Design of irregular screen sets that generate maximally smooth halftone patterns," in Proc. SPIE 9395, Color Imaging XX: Displaying, Processing, Hardcopy, and Applications, pp. 93950K, 2015.

Author Biography

Fritz Lebowsky received his MS (1985) and PhD (1993) in electrical engineering from the Technical University of Braunschweig, Germany. He began his professional career as a research and teaching assistant at the Institute of Telecommunications of the Technical University of Braunschweig in 1985. From 1991 he worked as a research and development engineer in the field of digital video processing at Micronas in Freiburg, Germany. In 1995 he joined Thomson Consumer Electronics Components in Meylan, France, as a development engineer modeling video processor networks as well as digital acquisition sub-systems for DVD ROM drives. In 2000 he joined the Imaging and Display Division of STMicroelectronics Inc. in San Jose, CA, developing advanced display engines for the PC flat panel monitor market. Since 2004 he is with STMicroelectronics in Grenoble, France, working on image quality improvement for consumer TV products.

Mariano Bona received his PhD (1983) in applied mathematics from the Scientific University of Grenoble, France. Since 1985 he is with STMicroelectronics, having primarily worked on front-end signal processing and speech compression. In 2015 he joined the research and development team focusing on innovative nonlinear image compression algorithms.