Flexible Storage of Still Images with a Perceptual Quality Criterion

Vincent Ricordel, Patrick Le Callet, Mathieu Carnec, Benoît Parrein

▶ **To cite this version:**

Vincent Ricordel, Patrick Le Callet, Mathieu Carnec, Benoît Parrein. Flexible Storage of Still Images with a Perceptual Quality Criterion. Advanced Concepts for Intelligent Vision Systems (ACIVS), Sep 2005, Antwerp, Belgium. hal-00451334

HAL Id: hal-00451334
https://hal.science/hal-00451334

Submitted on 28 Jan 2010

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

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

Flexible Storage of Still Images with a Perceptual Quality Criterion

Vincent Ricordel, Patrick Le Callet, Mathieu Carnec and Benoit Parrein

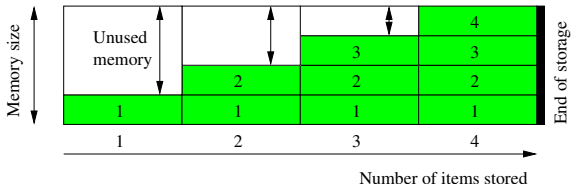
Image and Video Communication Team, IRCCyN UMR CNRS 6597
Ecole Polytechnique de l'Université de Nantes
La Chantrerie, BP 50609, 44306 Nantes Cedex 3
vincent.ricordel@polytech.univ-nantes.fr

Abstract. The purpose of the paper is to introduce a new method for flexible storage of still images. The complete design of the system is described with the scalable encoding, the distortion computation, the bits allocation strategy, and the method for the memory management. The main improvement is the full exploitation of a perceptual metric to assess precisely the introduced distortion when removing a layer in a scalable coding stream. Experimental results are given and compared with a system which uses the PSNR as distortion metric.

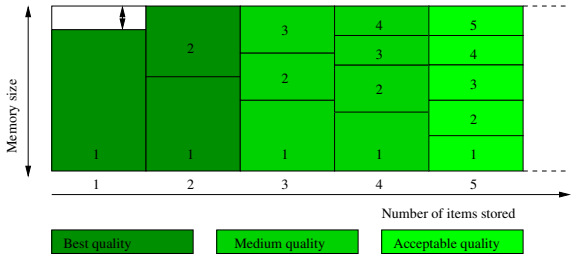
1 Introduction

In the paper we focus on the problem of digital still image storage. The problem occurs when the total memory size is limited, the amount of data to be stored is large and the quality of the decoded images must be high. Basic memory systems store each image, by granting each item a fixed share of the memory (see fig 1 [a]). The maximal number of stored images is thus limited. An improvement of these systems is usually based on the image encoding in order to minimize the compressed image size and to save a bigger memory area. The classical "flexible" storage implementation [1] aims at improving both these aspects : the image encoding and the storage process.
More precisely, the flexible method takes avantage of the scalability concept which provides a way to separate the image quality in several accessible layers. Then it is possible to keep the most significant image layers and to remove the least significant image layers already stored in the memory area. An optimal flexible allocation of storage capacity is performed with the use of flexible storage because : (*i*) the available storage capacity is fully exploited, (*ii*) the number of images to store is variable and can be changed at anytime, (*iii*) the highest possible quality is preserved when a new image is stored. While a basic system stores only necessary layers to get the minimal acceptable quality, a flexible system allows to store extra layers up to the best quality. Figure 1 [b] shows the advantages of the flexible storage principle with respect to a basic scheme. We can see that the allocated image size is distributed according to a quality criterion, which allows having little quality difference between the stored images.

The use of a memory whose size is fixed but with the properties associated to the flexible memory may have a lot of applications in image database storage or digital still cameras.



[a] Traditional memory usage



[b] Elastic memory usage

Fig. 1. Elastic memory principle.

2 System overview

Figure 2 illustrates the general structure of the system. To deliver a compressed version with several layers for each image, we use as scalable encoding scheme the standard JPEG2000 with quality scalability mode i.e. Signal to Noise Ratio (SNR) scalability. Storage a new image is achieved using a process of memory allocation. This implies a rate *vs.* distortion tradeoff to insert some new image layers by remaining the most relevant ones of the previous stored images. To measure the distortion, we have to use a quality criterion that must produce objective quality scores in good correlation with human judgment. Usually, PSNR is computed as a quality criterion. Unfortunatly, it is too poorly correlated with human judgement. We prefer to use a perceptual criterion. To present the entire process, the extraction and decoding steps are included in the figure 2 for the final image visualisation even if they are not specific to flexible memory. The following subsections detail more precisely the system overview.

2.1 The JPEG2000 quality layers

The SNR scalability included in JPEG2000 is based on a rate / distortion optimization, called Post Compression Rates Distortion (PCRD-opt), included in

the algorithm Embedded Coding with Optimal Truncation (EBCOT). This stage carried out after the contextual arithmetic coding is a rate allocation to various layers of quality which are separated between them by truncation points. Within an image, layers are hierarchically ordered which means that a layer might be exploited only if the first layers are available. From a practical point of view, the JPEG2000 header provides the set of truncation points between layers. So, it is easy to reach the associated rate of each layer and to remove unnecessary layers according to the hierarchical order. For this purpose, we have elected to take the JP2000 stream as it comes without modifying the rate / distortion process. This leads to define the optimal size of the layers for each image independently for our system.

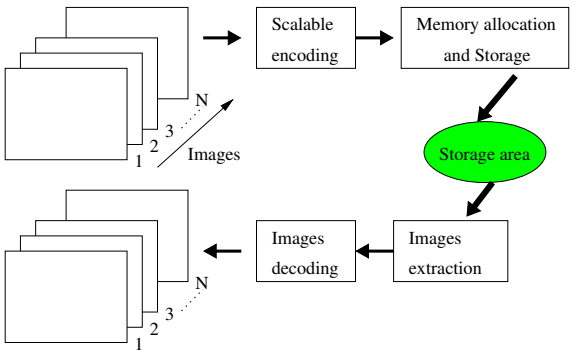


Fig. 2. Flexible storage implementation.

2.2 Perceptual Distortion

By the past we have developped several criteria based on human vision model for still color image quality assessment. Here, we use a reduced reference quality criterion that is well suited for distortions stemming from JPEG 2000 compression [2]. This criterion is divided into two steps. At first, a reduce description is built for both original and distorted image by extracting features in a psychovisual space. Then, the two reduced descritions are compared to produce the quality score.

To build the reduced description of an image, we project it in a psychovisual space. This is achieved through different functions including perceptual colorspace transformation (Krauskopf'one [3]), contrast sensitivity function, perceptual channels decomposition and masking effect functions. These different steps lead to a subband representation for the achromatic component (17 subbands) and to two images for the chromatic components. The next step, feature extraction is achieved only on the subbands of the achromatic component but taking into account the chromatic data. We select locations, called "characteristic points", in the subband representation in order to extract features on them.

These points are located on concentric ellipses, centered on the image center. We choose a fixed number of points per ellipse so these points are more concentrated in the image center (which generally gathers the objects of interest). At last, we extract several features on each characteristic point. First, a linear structure is extracted by a "stick growing algorithm". This structure is described by its orientation, its size (length and width), its maximum contrast amplitude. We also extract the mean values of achromatic and chromatic components computed on a circular neighboorhood of radius 5 pixels around the structure. All these features for all the characteristic points constitute the reduced description of the image.
The computation of an image quality from the reduced description of its reference and its distorted version, is also achieved in several steps. First, we produce for each characteristic point a correspondence coefficient per feature which indicates the similarity between the feature values of the distorted and the reference image. Then these coefficients are combined locally to produce a local similarity measure. The global similarity measure is defined as the mean of the local similarity measures. Finally, the objective quality score is produced using a linear transformation of the global similarity measure. Parameters of this linear transformation can be adapted to a particular distortion system. Here we use parameters adapted for JPEG2000. With this criterion [1], we get a value of 0.95 for the correlation coefficient between objective quality score and human judgment on the JPEG2000 images of the LIVE database [4].

2.3 Memory management

The bit allocation problem occurs in order to share the memory between the compressed images. It consists in minimizing the global distortion D subject to the constraint that the global rate R is under the memory size R_d. For each JPEG2000 stream corresponds a set of layers. Let be j the stream index, and i the layer number. $d_{i,j}$ and $r_{i,j}$ are respectively the distortion and the rate for the j-th stream with i layers (note that i can be different for each stream). If there are M streams, we assume that the problem is:

$$\min D = \min \sum_{j=0}^{M-1} d_{j,i} \text{ subject to } R = \sum_{j=0}^{M-1} r_{j,i} \leq R_d \tag{1}$$

A combination of M streams involves a point (R, D) in the distortion / rate space, all the combinations produce a cluster. The problem becomes the determination, on the cluster convex hull, of the point whose rate is just lower than R_d. In order to reduce the complexity the Lagrange-multiplier method is introduced to solve:

$$\min(D + \lambda \times R) \Longleftrightarrow \sum_{j=0}^{M-1} \min(d_{i,j} + \lambda \times r_{i,j})$$

[1] The implementation is available freely (http://www.dcapplications.t2u.com).

The complexity decreases because the reduction of distortion is now achieved separately from each stream. The general form of the algorithm is:
1. The convex hull $d(r)$ for each stream is directly computed, by storing the distortion / rate point corresponding to each layer;
2. The point on the global convex hull is determined, its rate is just below R_d.
For this second step, the algorithm proposed by Shoham [5] is particularly adapted. It is based on the computation of singular values of the Lagrange-multiplier λ. Precisely a singular value of λ is the slope of the line that pass through two consecutive points of the convex hull. So from a first point on the hull, by successive computations of singular values, we get the global convex hull.
In practice a first point is (R_{max}, D_{min}) where D_{min} equals to the sum of the distortions when considering each stream with all its layers, the corresponding rate R_{max} is then maximal. Removing the i-th layer of the j-th stream implies an increase in distortion and a decrease in rate, and we get:

$$\lambda_{i,j} = \frac{\Delta d_{i,j}}{\Delta r_{i,j}} = \frac{d_{i-1,j} - d_{i,j}}{r_{i,j} - r_{i-1,j}}$$

The successive singular values of λ are obtained by ordering the $\lambda_{i,j}$ in a decreasing manner and by considering all the streams.

The first step consists in computing a new distortion / rate curve for each new compressed image to store: respecting the layers order and the rates from JPEG2000 PCRD-opt, but using our perceptual metric. As a consequence the new $d^*(r)$ curves are not necessary convex and the previous bit allocation process has to be adapted to our problem.
As long as the memory is not full, the compressed images are stored with all their layers so with the best quality. When the physical storage area becomes full, the goal is to remove the layers which induce the largest decrease in rate and the smallest increase in distortion. For each new image addition in the memory, the basic allocation method steps are:
1. calculate the set of $\lambda^*_{i,j} = \Delta d^*_{i,j}/\Delta r_{i,j}$ and store them in a list conserving the layers order of the jpeg stream. This list will be used as a "FILO", where the first accessible item corresponds to the lambda of the last jpeg layer;
2. considering the FILO lists of the jpeg streams already stored and the FILO list of the new compressed image:
until the rate criterion of equation 1 is satisfied, find successively the smallest accessible item λ^*, remove it from its list and delete the corresponding jpeg layer;
We impose another constraint such as the system guarantees a minimal quality to the stored images. Thus it is forbidden to continue to remove a layer of a jpeg stream once this threshold is reached. In practice for the allocation method, the corresponding lambdas are simply retired from the list relative to the suppressible layers. Obviously when the memory is full with compressed images having reached all this threshold of minimal quality, it is impossible to add another picture (without erasing completely beforehand a stored jpeg stream).

3 Experimental Results

A first use consists of the insertion of six 512x512 colour images in the flexible memory with the perceptual quality criterion. The JPEG2000 coding is made from the Kakadu library[2] functions. The setup fixes a max rate to 1 bpp, a initial layers number to 9 and a perceptual score threshold to 4 for a memory size of 100 kB. Table 1 illustrates the various stages of the operation while indicating to each new insertion (ni): the number of layers, the perceptual score, the number of extra layers for each image inserted as well as the use of the memory in term of flexibility (a number of releasable bytes corresponding to the total number of extra layers) and of availability compared to the inserted images set. The first 3 insertions of 'Lena', 'Peppers' and 'Air-Force Plane' do not pose any problem because the memory is sufficient: no quality layers are removed. If one preserves in the state the whole of the layers only 1.8 % of the memory is available. On the other hand, one has a budget of around 50 kB (half of the memory) compared to the acceptable perceptual score. The insertion of a new image generates the optimization of the memory by suppression of extra layers ('Lena' loses 2 layers, 'Peppers', 'Air-Force' and 'Barbara' 1 layer). The budget of 20 kB is sufficient for the insertion of the House image which extracts all the extra layers of the other images except Lena which preserves 2 extra layers. These last bring a very significant quality compared to the bytes used. It does not allow the insertion of the sixth image ('Fruits') taking into account the thresholds of the flexible memory. For the 5 images stored, the perceptual scores are quite homogeneous. The goal of this second test set is the comparison with a PSNR based scheme. Five images of the previous data set are stored. The setup is the same as before, except: the coding max rate is 1.5 bpp, and the quality threshold is 25 dB for a PSNR and 3 for a perceptual distortion use. This parameters are choosen for increasing the extra layer number and for highlighting our results when storing the image 'Mandril'. The curves of the figures 3 (for which, after bit allocation, the choosen $PSNR(r)$ or $d^*(r)$ points are marked by circles) show indeed that for this high frequency image, the classical distortion gives low PSNR values and no layer can be removed. Our perceptual metric produces objective distortion measures: some layers can be deleted from the 'Mandril' stream without visual artifacts, and the saved bits give more rate to store the other images with a homogeneous visual quality.

4 Conclusion

We have introduced a new method for the flexible storage of still images. The method is based on the use of a perceptual metric which produces objective quality scores when coding the images with JPEG2000. An overview of the system is given. The perceptual distortion computation and the memory management are detailed. Results are given with a basic usage and a PSNR comparison. They

[2] avalaible at http://www.kakadusoftware.com

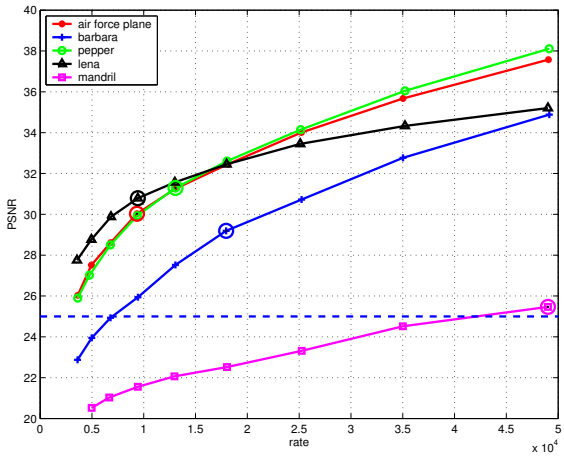
Table 1. Basic usage of flexible storage with successive new insertions (ni) of five images.

Inserted images	# layers	Percept. score	Extra layers	Extra bytes	Availability (%)
Lena (ni)	9	5	4	22030	67.2
Lena	9	5	4	36200	34.5
Peppers (ni)	9	5	2		
Lena	9	5	4	50251	1.8
Peppers	9	5	2		
Air-Force (ni)	9	4.89	2		
Lena	7	4.91	2	19729	7.7
Peppers	8	4.77	1		
Air-Force	8	4.58	1		
Barbara (ni)	8	4.60	1		
Lena	7	4.91	2	7932	0.8
Peppers	7	4.30	0		
Air-Force	7	4.10	0		
Barbara	7	4.58	0		
House (ni)	7	4.03	0		

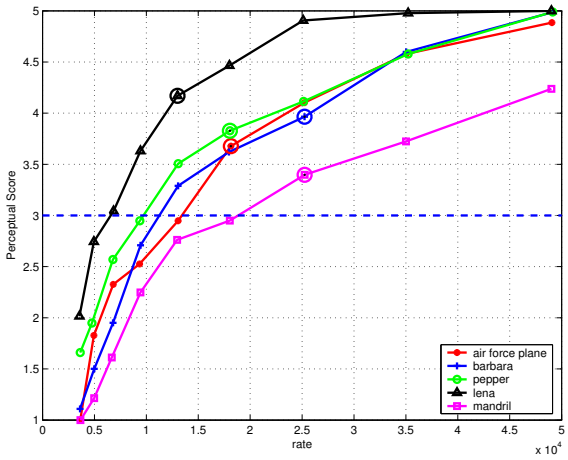
show how the objective quality evaluation allows to share more efficiently the bits between the stored layers of the JPEG2000 streams.

References

1. Van Der Vleuten, R., Kleihorst, R., Hentschel, C.: Flexible storage of images with application to digital cameras. In: Proc. of Int. Conf. on Image Proc. ICIP, Thessaloniki, Greece (2001) 1097–1100
2. Carnec, M., Le Callet, P., Barba, D.: An image quality assessment method based on perception of structural information. In: Proc. of Int. Conf. on Image Proc. ICIP, Singapore (2004)
3. Krauskopf, J., Williams, D., D.W., H.: Cardinal directions of color space. Vision Research **22** (1982) 1123–1131
4. Sheikh, H., Wang, Z., Cormak, L., A.C., B.: Live image quality assessment database. http://live.ece.utexas.edu/research/quality (2003)
5. Shoham, Y., Gersho, A.: Efficient bit allocation for an arbitrary set of quantizers. IEEE Transactions on Acoust. Speech Signal Processing **36** (1988) 1445–1453



$PSNR(r)$ curves.



$d^*(r)$ curves.

Fig. 3. Distortion / rate curves obtained when storing 5 images in the flexible memory. The circles mark the retained points after bit allocation.