Proc. IV Serbian-Bulgarian Astronomical Conference,
Belgrade 21–24 April 2004,
eds. M. S. Dimitrijević, V. Golev, L. Č. Popović, M. Tsvetkov,
Publ. Astron. Soc. "Rudjer Bošković" No 5, 2005, 95 - 100 *Invited lecture*

WEB ACCESS AND IMAGE PROCESSING IN ASTROPHYSICAL DATABASES

D. KALAGLARSKY

Space Research Institute, Bulgarian Academy of Sciences
E–mail damyan@skyarchive.org

Abstract. Web technologies motivate needs for fast access to informative pixel arrays, which are more intuitively understandable than the stellar catalogues alone. The transfer of huge astrophysical images over the web at present is unrealistic in view of hours needed at a typical user bandwidth rate. The current paper aims to preset the development of a system for web-access to image in the Wide-Field Plate Database (WFPDB) (Tsvetkov and Tsvetkova, 1999). The basics of the image compression methods are discussed. The object signal will be stored with noise, background variations, and so on. Different packages of wide-filed astrophysical images compression as Hcompress FitsPress, Independent JPEG Group (IJG) software are mentioned. The access to the WFPDB images has to solve different problems and face various requirements. The effective compression is the key feature to enabling the users to performe the desired image previews as fast as possible. Besides, some basic on-line image processing like noise-filtering, brightness/contrast enhancement should be also applied. The WFPDB users have to be able to see an image at a desired resolution, as well as to get just a part of the image. They can require a preview that is smaller or bigger, with higher level of detailness, depending on their purpose and network bandwidth. One can also get a high level resolution image of a specified region of the plate without having to transfer the whole amount of data.

1. NEED FOR COMPRESSION

The digitization of photographic plates is done by various automatic plate scanning machines (COSMOS, SuperCOSMOS, APS, PMM, PDSs, Flatbed scanners, etc.). These machines allow for the quantification of the truly enormous amount of useful astronomical data represented in a photograph of the sky, and they have realized the full potential of the large area photographic sky surveys. Image information is coded as an array of intensity values, reproducing the geometry of the scanner used for plate digitization.

The storage and simple transfer of such amounts of data over computer networks becomes too cumbersome and in some cases practically impossible. A typical scan in the Wide-Field Plate Database (WFPDB), for example is about 300-600 MB. Its transfer over the Internet would take hours at a typical user bandwidth rate.

The WFPDB contains data for various plate catalogues. It consists both of structured data, describing the plate from astronomical point of view, like center of filed, observation time, filter used, etc... along with scans of the plate itself. Often the user of the WFPDB on-line search system would like just to have a look at a preview of the plate, rather than to download the whole image, which is usually hundreds of megabytes in size. Thus one can quickly get an idea of the quality of the plate and its usefulness for his/her purposes. Since the WFPDB plate archives consist of wide-filed plate images mostly, it is also very convenient to be able to retrieve an image of only a region of the whole plate, thus not bothering yourself with data you are not interested in. Some on-line image processing techniques would also be of use to the user of the system. It is nice to be able to enhance the brightness/contrast of your image, to apply different filters without using additional software.

2. BASIC PRINCIPLES OF IMAGE COMPRESSION AND STANDARD METHODS USED IN ASTRONOMY

The compression methods make use of the redundancy contained in the image data in order to reduce the number of bits needed to code it. Since astrophysical images show good spatial correlation it is often possible to achieve good compression ratios without almost any visual loss of quality.

2.1. THE TYPICAL COMPRESSION STEPS

1. Apply a transformation to another domain, where in general the entropy of the coefficients is smaller (e.g. using wavelet transform, discrete cosine transform, etc.).
2. Quantize the coefficients obtained.
3. Code the quantized values by a loss-less method (i.e. Huffman tree, RLE, arithmetic coder, etc.) (Stark, TUTORIAL).

The first step is very essential. Choosing the proper transform determines to a great extent the effectiveness of your method for the class of images that are targeted. If we have an image I which pixels are distributed among L intensity levels with probability of the i$^{\text{th}}$ level to appear in the imagep_i, the entropy is given by:

$$H = \sum_{i=1}^{L} -p_i \log(p_i)$$

The smaller the entropy is, the more "compact" your data are.

Step 2, unlike 1 and 3 is irreversible. The distortion of the reconstructed image will depend mostly on the way in which the coefficients are quantized. That's why we have to compromise between the number of bits used for coefficient quantization and the image quality. The desired quality is quite subjective and will depend on the particular application. Since the images in the WFPDB on-line system have to serve for visual previews mostly, without specific astronomical tasks like astrometry and photometry being applied on them, we could make a decent trade-off in this step of the compression.

2.2. SOME STANDARD SOFTWARE PACKAGES

- JPEG: A standard compression algorithm developed by the Joint Photographic Experts Group (JPEG). It decorelates pixels coefficient within 8 x 8 blocks using the Discrete Cosine Transform (DCT) and uniform quantization. There is a free software library by the Independent JPEG Group (IJG).
- FITSPRESS: based on wavelet transform with Daubechies-4 filters.
- HCOMPRESS: based on the Haar wavelet transform.

All of these packages are in some way inappropriate to the purposes of the WFPDB on-line system, since they are not oriented towards very large images (VLI). The handling of VLI involves specific memory management and other software considerations. These packages are also not quite suitable for some WFPDB tasks like partial image reconstruction, visualization of regions of interest, etc. The compression ratios image quality achieved by them are also not quite satisfactory too.

3. METHODS CONSIDERED

Having to face the development of a specific image compression and visualization, different on-line methods have been considered. They have been estimated according to the results shown by their implementations in terms of compression ratio, image quality, computational complexity, resources usage, etc.

3.1. MULTIRESOLUTION APPROXIMATION

By manipulating the resolution of a signal for the purposes of a specific task, one can inspect only the necessary details. The multiresolution pyramid (Mallat, 1999) allows to process an image at a low resolution, and subsequently to increase it, if needed.

Formally, the approximation of a signal f at resolution 2^{-j} is defined as an orthogonal projection of f in a sub-space V_j. V_j consists of all possible projections of approximations of f at resolution 2^{-j}. The orthogonal projection of f is a function $f_j \in V_j$ that minimizes the norm $\|f - f_j\|$. The following formal definitions describe the properties of multiresolution sub-spaces:

1. $\forall (j,k) \in \mathrm{Z}_2, f(t) \in V_j \Leftrightarrow f(t - 2^j k) \in V_j$
2. $\forall j \in \mathrm{Z}, V_j + 1 \subset V_j$
3. $\forall j \in \mathrm{Z}, f(t) \in V_j \Leftrightarrow f(\frac{t}{2}) \in V_j + 1$
4. $\lim\limits_{j \to +\infty} V_j = \bigcap\limits_{j=-\infty}^{+\infty} V_j = \{0\}$
5. $\lim\limits_{j \to -\infty} V_j = \mathrm{Hull}(\bigcup\limits_{j=-\infty}^{+\infty} V_j) = L_2(\mathrm{R})$

These properties are the basis of an apparatus that would make possible to reconstruct a compressed image at resolution 2^{-j}, which would be quite applicable to the purposes of the WFPDB on-line system.

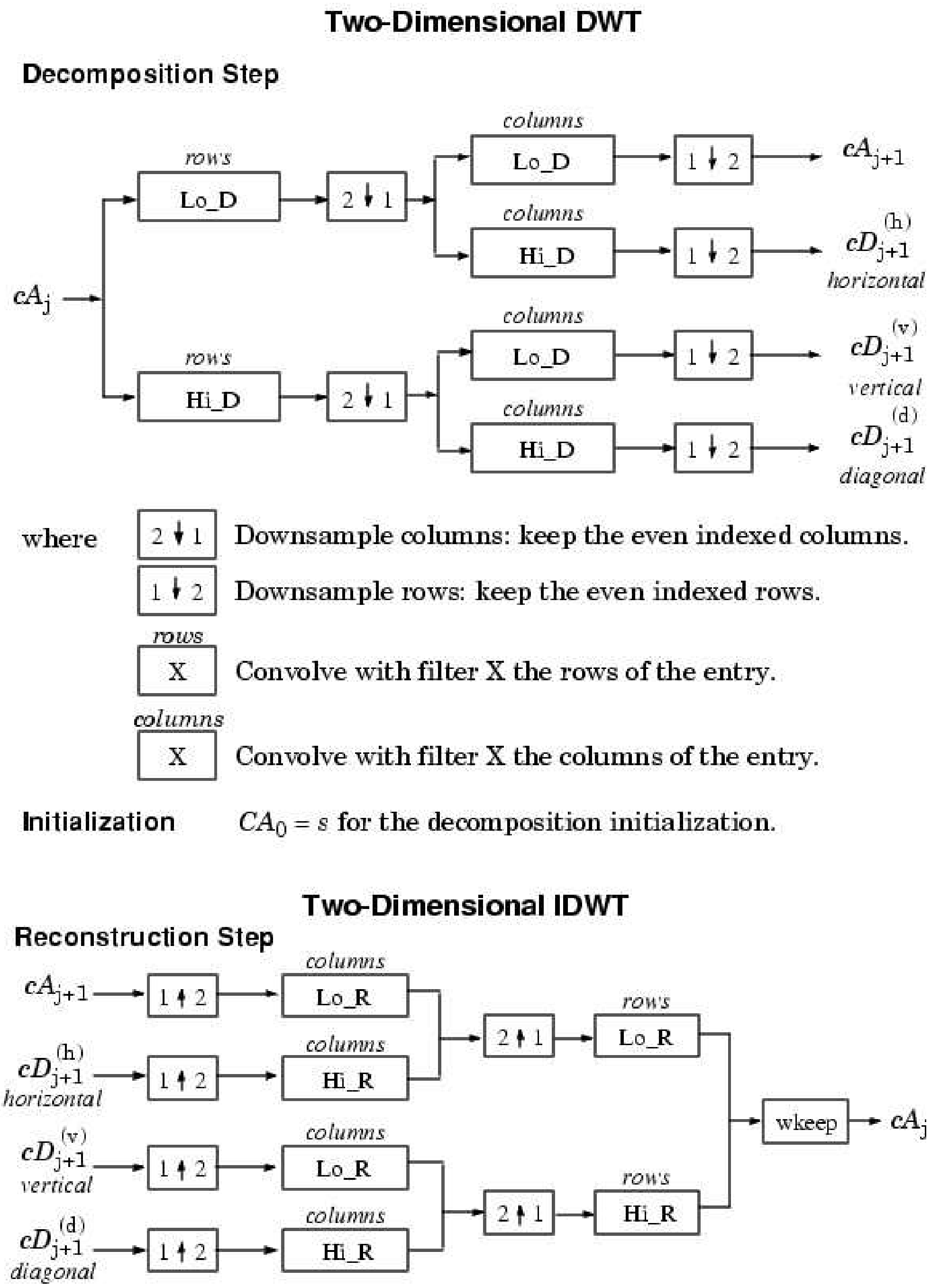
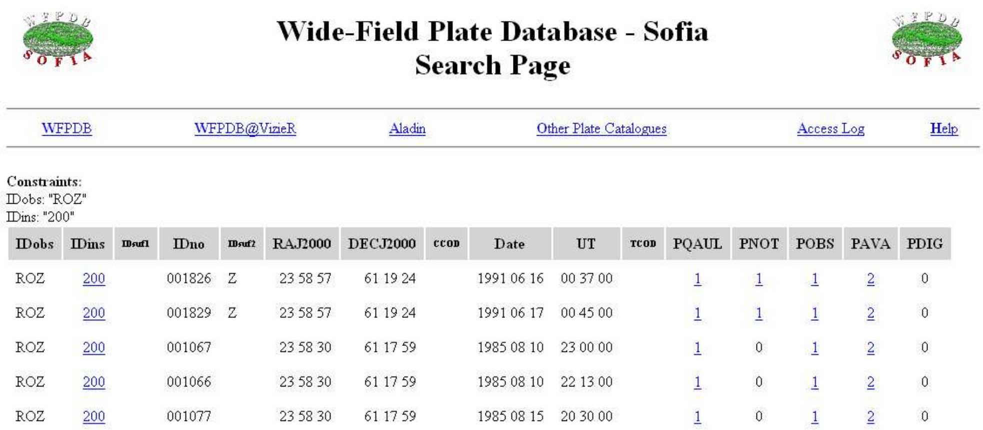


Figure 1: Discrete Wavelet Transform.



Wide-Field Plate Database - Sofia
Search Page

WFPDB | WFPDB@VizieR | Aladin | Other Plate Catalogues | Access Log | Help

Constraints:
IDobs: "ROZ"
IDins: "200"

IDobs	IDins	IDsuf1	IDno	IDsuf2	RAJ2000	DECJ2000	CCOD	Date	UT	TCOD	PQAUL	PNOT	POBS	PAVA	PDIG
ROZ	200		001826	Z	23 58 57	61 19 24		1991 06 16	00 37 00		1	1	1	2	0
ROZ	200		001829	Z	23 58 57	61 19 24		1991 06 17	00 45 00		1	1	1	2	0
ROZ	200		001067		23 58 30	61 17 59		1985 08 10	23 00 00		1	0	1	2	0
ROZ	200		001066		23 58 30	61 17 59		1985 08 10	22 13 00		1	0	1	2	0
ROZ	200		001077		23 58 30	61 17 59		1985 08 15	20 30 00		1	0	1	2	0

Figure 2: Plate search results (web-browser screenshot).

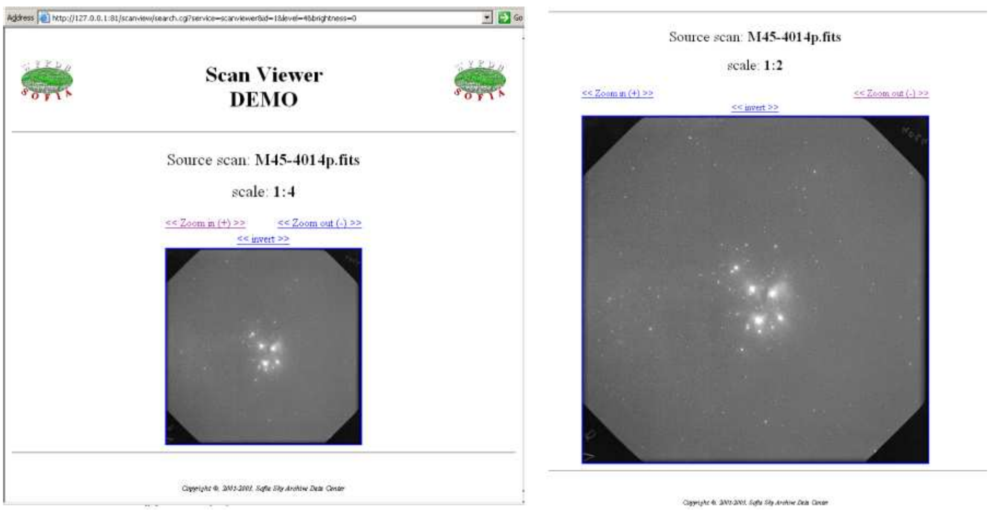


Figure 3: Scan images at different resolution (web-browser screenshot).

3.2. FAST WAVELET TRANSFORM (FWT)

The FWT algorithm presents recursively each approximation $PV_j f$ of a given signal f by a coarser approximation $PV_{j+1} f$ plus the wavelet coefficients $PW_{j+1} f$. The wavelet transform can be computed by consequently convolving the signal with a high-pass h and a low-pass g filters and downsampling. Thus no data redundancy is introduced. N input samples will result in N wavelet coefficients. The reconstruction goes analogously by upsampling the coefficients (Fig. 1).

3.3. CURVELET TRANSFORM

The digital curvelet transform is developed by Candès and Donoho for the needs of image analysis (Donoho and Duncan, 1999). The transform has been designed to represent effectively edges and other singularities along curves.

However, the transform is strongly redundant, thus not very appropriate for the purposes of image compression in WFPDB. The redundancy factor is equal to 16J + 1 whenever J scales are employed in the implementation.

4. DEVELOPMENT OF SOFTWARE

The image presentation system is currently being developed as a module of the WFPDB on-line search system. Having located the plate you are interested in through the interface for searching / navigating in the results (Fig. 2), you can request different previews of the scans of a plate (if available, of course). The presentation module will follow the software architecture of the search system (CGI) and will interact with the other modules of the system. You can choose the resolution of the image you get according to the level of details necessary for your tasks (Fig 3).

5. CURRENT PROBLEMS AND DIRECTIONS FOR FUTURE WORK

Various problems have to be solved in the future development of the system. The implementation has to face plenty of challenges, including enhanced memory management, speed optimization, more profound compression rates analysis. Another open issue is the form in which the generated image is to be transferred to the client node (web-browser). For the moment, JPEG is used. Maybee a native code should to be developed, so that image will not have to be recoded. Some on-line image processing techniques have also to be implemented. Identification of an image with reference to star catalogues is also a problem to be faced.

References

Donoho, D.L., Duncan, M.R.: 1999, Digital Curvelet Transform: Strategy, Implementation and Experiments, `http://www-stat.stanford.edu/~donoho/Reports/1999/DCvT.pdf`.

Mallat, St.: 1999, Wavelet tour, Second Edition, Academic Press.

Stark, J.L: TUTORIAL: Multiscale Methods and Applications, `http://jstarck.free.fr/cours.htm`.

Tsvetkov, M., Tsvetkova, K.: 1999, Treasure-Hunting in Astronomical Plate Archives (THAPA), Proceedings of the International Workshop, Sonneberg, March 4-6, 1999, 88.