

## Astronomical image compression

M. Louys<sup>1,2</sup>, J.L. Starck<sup>3</sup>, S. Mei<sup>1,4</sup>, F. Bonnarel<sup>1</sup>, and F. Murtagh<sup>1,5</sup>

<sup>1</sup> Observatoire de Strasbourg, Université Louis Pasteur, F-67000 Strasbourg, France

<sup>2</sup> Laboratoire des Sciences de l'Informatique, de l'Image et de la Télédétection, ENSPS, Université Louis Pasteur, F-67000 Strasbourg, France

<sup>3</sup> DAPNIA/SEI-SAP, CEA-Saclay, F-91191 Gif-sur-Yvette Cedex, France

<sup>4</sup> European Southern Observatory, Karl-Schwarzschild-Straße 2, D-85748 Garching, Germany

<sup>5</sup> School of Computer Science, The Queen's University of Belfast, Belfast BT7 1NN, Northern Ireland

Received December 14, 1998; accepted February 8, 1999

**Abstract.** We compare a range of powerful compression methods – fractal, wavelets, pyramidal median, JPEG – with compression tools dedicated to astronomy such as HCOMPRESS, FITSPRESS and Mathematical Morphology, and apply these to astronomical images. Quality is quantified from visual appearance, and from photometric and astrometric measurements. Computational requirements of each method are discussed. We also review the implications of Web-based storage and transmission, stressing what we term progressive vision. In summary, no method is perfect, but the PMT method is the best compromise for general astronomical images, combining acceptable photometric and positional precision with good compression capabilities. JPEG is still an excellent method for compression factors less than 40 and has the advantage of being very widely available.

**Key words:** data analysis — image processing — astronomical data bases: miscellaneous

### 1. Introduction

From year to year, the quantity of astronomical data increases at an ever growing rate. In part this is due to very large digitized sky surveys in the optical and near infrared, which in turn owes its origin to the development of digital imaging arrays such as CCDs. The size of digital arrays continually increases following the demands of astronomical research for obtaining larger quantities of data in shorter time periods. Currently, projects such as the European DENIS and American 2MASS infrared sky surveys, or the Franco-Canadian MEGACAM Survey and the American SLOAN Digital Sky Survey, will each produce of

the order of 10 Tbytes of image data. The routine and massive digitization of photographic plates have been made possible by the advent of automatic plate scanning machines (MAMA, APM, COSMOS, SuperCOSMOS, APS, PMM, PDSs) (Richter 1998). 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. Clearly the storage and manipulation of astronomical data always requires the latest innovation in archiving techniques (12" or 5¼" WORM in the past, CD WORMS or even magnetic disks with RAID technology now, hopefully DVD in the near future). In addition, the simple transfer of such amounts of data over computer networks becomes too cumbersome and in some cases practically impossible. The transmission of a high resolution Schmidt plate image over the Internet would take of the order of 50 hours. Facing this extraordinary increase in pixel volumes, and taking into account the fact that catalogues produced by extraction of information from the pixels can always be locally wrong or incomplete, the needs of the astronomer follow two very different directions:

- On one side the development of Web technology creates a need for fast access to informative pixel maps, which are more intuitively understandable than the catalogues alone.
- On the other side, quantitative work often requires accurate refinement of astrometry and photometry, or effective redetection of missed objects.

Thus, the astronomical community is confronted with a rather desperate need for data compression techniques. Several techniques have in fact been used, or even developed, in the field of astronomy. Véran & Wright (1994) studied lossless techniques. White et al. (1992) developed HCOMPRESS, based on the H-transform, Press et al. (1992) developed FITSPRESS based on the Daubechies wavelet

transform. JPEG, a general purpose standard has been tested by us. Compression based on the multiresolution pyramidal median transform (PMT) algorithm has been developed by Starck et al. (1996). Huang & Bijaoui (1991) introduced *MathMorph* for astronomical image processing.

This issue of data compression has become more important to us at the Centre de Données astronomiques de Strasbourg (CDS) in the context of the ALADIN project, an all-sky “clickable map” service provided by Strasbourg Observatory (Bartlett et al. 1996), which has been developed with the goal of storing and providing network access to a full sky archive of digitized Schmidt plates. Currently we provide access to this archive on the Web, allowing users of the CDS services to retrieve images of objects starting from occurrence of names in astronomical catalogues, the SIMBAD database, and bibliographical references such as titles or abstracts.

In previous papers (Starck et al. 1996; Murtagh et al. 1998) we reported our findings concerning the effects of three compression algorithms on astrometry and photometry. The three methods considered were *HCOMPRESS*, *FITSPRESS* and the video standard *JPEG*. Concerning the signal-to-noise ratio, the photometry and astrometry, *JPEG* and *HCOMPRESS* produce images of equivalent quality, but *FITSPRESS* is worse than the other two methods. The conclusion of this study is that the standard *JPEG* method was, ultimately, not so bad, even if block artifacts appear. However such block artifacts certainly appeared to be prohibitive for compression rates above 40.

In the present work, we extend our research to more realistic astronomical settings by considering stars with pixel values closer to the background, and by examining different criteria for position and magnitude determination. In addition to the three previous compression methods we studied two other methods: *PMT*, based on multiresolution analysis, and another one based on mathematical morphology, both implemented in the *MR/1* package (MR/1 1998). We used a sample of nearly 2000 stars from an ESO Schmidt plate (not one, in fact, belonging to the survey plates) centred on the globular cluster M 5. The results indicate that *PMT* can give compression ratios of up to 5 times the maximum ratio obtained from the other methods, when the regions are not too dense.

The next section contains a brief and general description of image compression techniques, and of the four compression software packages, *FITSPRESS*, *HCOMPRESS*, *JPEG* and *PMT*. This is followed in Sect. 3 by a presentation of the data and calibrations used for our study (and a discussion of our approach to testing the astronomical quality assessments of the compressed images), and a presentation of our results. We then conclude in Sect. 4.

## 2. Image compression methods

### 2.1. The principle

Numerical image information is simply coded as an array of intensity values, reproducing the geometry of the detectors used for the observation or the densitometer used for plate digitization. The object signal will be stored with noise, background variations, and so on. The relevant information depends on the application domain, and represents what the astronomer wants to study. The information of relevance reflects the limits of the observing instrument and of the digitization process. Reducing the amount of data to be coded will require that the relevant information be selected in the image and that the coding process be reorganized so that we emphasize the relevant information and drop noise and non-meaningful data. For this, we can focus on the region of interest, filter out noise, and quantize coarsely to take into account the limits of our human visual system if the images are only used for browsing.

Furthermore, the usual pixel array representation associated with images stores a lot of redundant information due to correlation of intensity values between nearby pixels and between different scales for large image structures or slowly varying background. A good compression scheme should aim at concentrating on the meaningful information in relation to the scientific purpose of the imaging (survey) project and code it efficiently, thereby limiting as much as possible the redundancy.

For this study, we examined the major available image compression packages, and compared their strategies with respect to these goals.

### 2.2. Compression packages

Methods used in astronomy include *HCOMPRESS* (White et al. 1992), *FITSPRESS* (Press 1992), and *JPEG* (Furht 1995). These are all based on linear transforms, which in principle help to reduce the redundancy of pixel values in a block and decorrelate spatial frequencies or scales. Two other methods have also been proposed for astronomical image compression: one using mathematical morphology, and another based on the pyramidal median transform (a nonlinear transform). A specific decompression method has also been developed in Bijaoui et al. (1996) in order to reduce artifacts relative to the *HCOMPRESS* method. In the signal processing domain, two other recent approaches are worthy of mention. The first is based on fractals, and the second uses a bi-orthogonal wavelet transform.

We first briefly review all of these methods, and then compare them in the framework of astronomical images.

## HCOMPRESS

HCOMPRESS (White et al. 1992) was developed at Space Telescope Science Institute (STScI, Baltimore), and is commonly used to distribute archive images from the Digital Sky Survey DSS1 and DSS2. It is based on the Haar wavelet transform. The algorithm consists of

1. applying a Haar wavelet transform to the data,
2. quantizing the wavelet coefficients linearly as integer values,
3. applying a quadtree to the quantified value, and
4. using a Huffman coder.

Sources are available at

<http://www.stsci.edu/software/hcompress.html>

## HCOMPRESS with iterative decompression

Iterative decompression was proposed in Bijaoui et al. (1996) to decompress files which were compressed using HCOMPRESS. The idea is to consider the decompression problem as a restoration problem, and to add constraints on the solution in order to reduce the artifacts.

## FITSPRESS

FITSPRESS (Press 1992) uses a threshold on very bright pixels and applies a linear wavelet transform using the Daubechies-4 filters. The wavelet coefficients are thresholded according to a noise threshold, quantized linearly and runlength encoded. This was developed at the Center for Astrophysics, Harvard. Sources are available at <ftp://cfata4.harvard.edu/pub/fitspress08.tar.Z>.

## JPEG

JPEG is the standard video compression software for single frame images (Furht 1995). It decorrelates pixel coefficients within  $8 \times 8$  pixel blocks using the discrete cosine transform (DCT) and uniform quantization.

## Wavelet

Various wavelet packages exist which support image compression, leading to more sophisticated compression methods. The wavelet transform we used is based on a bi-orthogonal wavelet transform (using Antonini-Daubechies 7/9 coefficients) with non-uniform coding (Taubman & Zakhor 1994), and arithmetic encoding. Source code is available at <http://www.cs.dartmouth.edu/~gdavis>

## Fractal

The image is decomposed into blocks, and each block is represented by a fractal. See Fisher (1994) for more explanation.

## Mathematical morphology

This method (Starck et al. 1998), denoted **MathMorph** in this work, is based on mathematical morphology (erosion and dilation). It consists of detecting structures above a given level, the level being equal to the background plus three times the noise standard deviation. Then, all structures are compressed by using erosion and dilation, followed by quadtree and Huffman coding. This method relies on a first step of object detection, and leads to high compression ratios if the image does not contain a lot of information, as is often the case in astronomy.

## Pyramidal median transform

The principle of this compression method (Starck et al. 1996; Starck et al. 1998), denoted **PMT** here, is to select the information we want to keep, by using the pyramidal median transform, and to code this information without any loss. Thus the first phase searches for the minimum set of quantized multiresolution coefficients which produce an image of “high quality”. The quality is evidently subjective, and we will define by this term an image such as the following:

- there is no visual artifact in the decompressed image; and
- the residual (original image – decompressed image) does not contain any structure.

Lost information cannot be recovered, so if we do not accept any loss, we have to compress what we take as noise too, and the compression ratio will be low (3 or 4 only).

### 2.3. Remarks on these methods

The pyramidal median transform (PMT) is similar to the mathematical morphology (**MathMorph**) method in the sense that both try to understand what is represented in the image, and to compress only what is considered as significant. The PMT uses a multiresolution approach, which allows more powerful separation of signal and noise. The latter two methods are both implemented in the MR/1 package (see <http://ourworld.compuserve.com/homepages/multires>).

Each of these methods belongs to a general scheme where the following steps can be distinguished:

1. Decorrelation of pixel values inside a block, between wavelength, scales or shape, using orthogonal or non-linear transforms.

**Table 1.** Description and comparison of the different steps in the compression packages tested

Software	Transform	Coefficient Quantization	Coefficient Organisation	Geometrical Redundancy reduction	Statistical Redundancy reduction
JPEG	DCT 8×8 pixels	Linear	Zigzag sequence	Runlength coding	Huffmann
HCOMPRESS	Haar 2×2 pixels	Linear	Pyramidal	Quadtree on bitplanes	Huffmann
FITSPRESS	Wavelets Daubechies-4	Linear	Increasing Resolution	Runlength coding	Huffmann
MR/1 PMT	Pyramidal Median Transform	Linear / Noise estimation	Decreasing Resolution	Quadtree on bitplanes	Huffmann
MR/1 Math. Morph.	Erosion/ Dilation	Linear / Noise estimation	–	Quadtree on bitplanes	Huffmann

2. Selection and quantization of relevant coefficients.
3. Coding improvement: geometrical redundancy reduction of the coefficients, using the fact that pixels are contiguous in an array.
4. Reducing the statistical redundancy of the code.

How each method realizes these different steps is indicated in Table 1.

Clearly these methods combine many strategies to reduce geometrical and statistical redundancy. The best results are obtained if appropriate selection of relevant information has been performed before applying these schemes.

For astronomical images, bright or extended objects are sought, as well as faint structures, all showing good spatial correlation of pixel values and within a wide range of greylevels. Noise background, on the contrary, shows no spatial correlation and fewer greylevels. The removal of noisy background helps in regard to data compression of course. This can be done with filtering, greylevel thresholding, or coarse quantization of background pixels. This is used by FITSPRESS, PMT and MathMorph which divide information into a noise part, estimated as a Gaussian process, and a highly correlated signal part. MathMorph simply thresholds the background noise estimated by a 3-sigma clipping, and quantizes the signal as a multiple of sigma (Huang & Bijaoui 1991). FITSPRESS thresholds background pixels and allows for coarse background reconstruction, but also keeps the highest pixel values in a separate list. PMT uses a multiscale noise filtering and selection approach based on noise standard deviation estimation. JPEG and HCOMPRESS do not carry out noise separation before the transform stage.

#### Identifying the information loss

Apart from signal-to-noise discrimination, information losses may appear after the transforms at two steps: coefficient selection and coefficient quantization. The interpretable resolution of the decompressed images clearly depends upon these two steps.

If the spectral bandwidth is limited, then the more it is shortened, the better the compression rate. The coefficients generally associated with the high spatial frequencies related to small structures (point objects) may be suppressed and lost. Quantization also introduces information loss, but can be optimized using a Lloyd-Max quantizer for example (Proakis 1995).

All other steps, shown in Table 1, such as reorganizing the quantized coefficients, hierarchical and statistical redundancy coding, and so on, will not compromise data integrity. This statement can be made for all packages. The main improvement clearly comes from an appropriate noise/signal discrimination and the choice of a transform appropriate to the objects' signal properties.

#### Identifying the needs for compression

Following the kind of images and the application needs, different strategies can be used:

1. Lossy compression: in this case the compression ratio is relatively low ( $< 5$ ).
2. Compression without visual loss. This means that one cannot see the difference between the original image and the decompressed one. Generally, compression ratios between 10 and 20 can be obtained.
3. Good quality compression: the decompressed image does not contain any artifact, but some information is lost. Compression ratios up to 40 can be obtained in this case.
4. Fixed compression ratio: for some technical reasons, one may decide to compress all images with a compression ratio higher than a given value, whatever the effect on the decompressed image quality.
5. Signal-to-noise separation: if noise is present in the data, noise modelling can allow very high compression ratios just by including some type of filtering in the wavelet space during the compression.

Following the image types, and the selected strategy, the optimal compression method may vary. The main interest

**Table 2.** List of criteria for comparison of compression methods for various types of astronomical image-based applications

Application type	Quick view	Catalog overlay: cross-correlation	Source extraction and Cross-identification	Deep detection	Recalibration
quality: visual	medium	high	medium	indifferent	high
quality: precision	low	medium	high	very high	high
Transfer + computation speed	very fast	fast	medium	slow	medium
Progressive vision	yes	yes	no	no	no

in using a multiresolution framework is to avail of progressive information transfer and visualization.

### 3. Comparison

#### 3.1. Quality assessment

In order to compare the different compression methods, we can use several characteristics, with constraints on these characteristics depending on the type of applications (see Table 2).

The progressive vision aspect is very useful in the context of quick views (for example on the Web) and catalogue overlays, where the user can decide when the quality of a displayed image is sufficient. On the contrary, this feature is not required for more quantitative tasks.

The requirement of speed of display (transfer + processing time) is usually critical for applications related to progressive vision.

The estimation of the quality of a compression method and rate compared to others is based on the quality of restitution of the relevant information, which is always relative to the type of application. For good quality quick views of a given area, catalogue and database overlays, and cross-correlation of sources at various wavelengths, the required quality will be essentially qualitative: good geometry of the objects, no visual artifacts, good contrast, etc.

For cross-identification processes, and any situation where recalibration to improve astrometry and photometry is needed, or reprocessing of object detections where some were obviously missed, star/galaxy discrimination or separation of distinct objects falsely merged, the quality estimation must be a quantitative process. The loss of information can be measured by the evolution of “relevant parameters” varying according to compression rate and method.

Quality criteria which can be retained for estimating the merits and performances of a compression method fall under these headings:

1. Visual aspect.
2. Signal-to-noise ratio.
3. Detection of real and faint objects.

4. Object morphology.
5. Astrometry.
6. Photometry.

Very few really quantitative studies have been carried out up to now in astronomy, in order to define which compression method should be used. Two studies were carried out in the framework of the ALADIN project. One was in 1993-94, when JPEG, FITSPRESS, and HCOMPRESS were evaluated (Carlsohn et al. 1993; Dubaj 1994) and another in 1996-1997 (Starck et al. 1997a; Starck et al. 1997b), when JPEG and PMT were compared (see Sects. 3.3 and 3.4).

#### 3.2. Visual quality

A quick overview was obtained of each method produced by running all compression algorithms on two images. The first was a  $256 \times 256$  image of the Coma cluster from an STScI POSS-I digitized plate, and the second was a  $1024 \times 1024$  image, extracted from the ESO 7992V plate digitized by CAI-MAMA (described in more detail in the next section). The visual quality was estimated from the visual aspect of the decompressed image, and the quality of the residual (original image – decompressed image). Conclusions relative to this study are:

- FITSPRESS leads to cross-like artifacts in the residual image, a loss of faint objects and a decrease in objects’ brightness.
- JPEG cannot be used at compression ratios higher than 40. Above this, artifacts become significant, and furthermore astrometry and photometry become very bad.
- The fractal method cannot be used for astronomical data compression. There are boxy artifacts, but the main problem is that object fluxes are modified after decompression, and the residual contains a lot of information (stars or galaxies can be easily identified on the residual map).
- MathMorph leads to good compression ratios, but the background estimation is delicate. For the Coma cluster, the result was relatively bad, due to the difficulty of finding the background. More sophisticated algorithms can certainly be used to do this task. Another

**Table 3.** Detection of faint objects in a digitized patch of a Schmidt plate image, using the MIDAS detection routines SEARCH/INVENTORY on original and compressed/decompressed images at different compression rates: 4:1, 10:1, 20:1 and 40:1. With comparable detection parameters, and depending on compression method, faint objects can be lost or spurious objects corresponding to local maxima can be found. Visual inspection is necessary to confirm real detected objects

Method	compression ratio	Real objects detected	Lost objects	False objects detected	Number of detection errors	Percentage of detection errors
JPEG	4	15	1	2	3	19
	10	14	2	2	4	25
	20	14	2	4	6	38
	40	13	3	5	8	50
Hcompress	4	14	2	3	5	31
	10	14	2	5	7	44
	20	11	5	3	8	50
	40	11	5	5	10	53
Fitspress	4	15	1	1	2	13
	10	13	3	0	3	19
	20	10	6	0	6	38
	40	5	11	0	11	69

drawback of this method is the bad recovery of the contours of the object, which leads also to a loss of flux.

- **HCOMPRESS** produces artifacts. Iterative reconstruction allows them to be suppressed, but in this case the reconstruction takes time. However this approach should be considered when the archived data are already compressed with **HCOMPRESS** (e.g. HST archive).
- The wavelet method produces very good results for the Coma cluster (compression ratio of 40). For the second image, where a compression ratio of more than 200 is obtained with the **PMT** or by mathematic morphology, artifacts appear if we try to achieve the same high performances. This method can be used, but not for very high compression ratios.
- The pyramidal median transform produces good quality results for both images. The compression ratio, similarly to the mathematical morphology method, depends on the content of the image. The fewer the pixels of objects in the image, the higher the compression ratio.

An interesting feature of the wavelet method is that the compression ratio is a user parameter. For **PMT**, and **MathMorph**, the compression ratio is determined from noise modelling. For other methods, a user parameter allows the compression ratio to be changed, and consequently the image quality, but only iterations can lead to a given compression ratio, or to a given quality.

### 3.3. First ALADIN project study

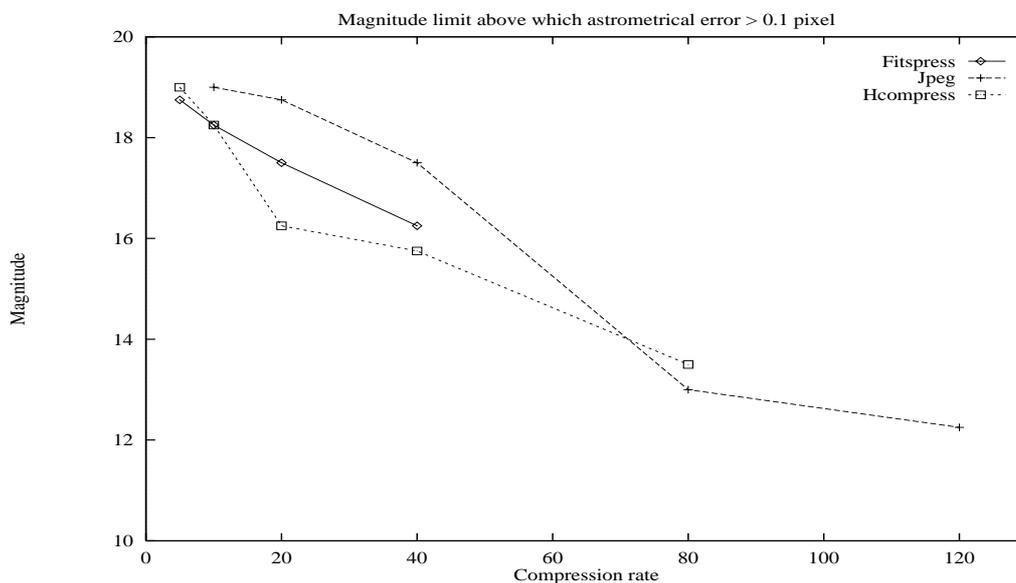
We conducted two quantitative studies at CDS, within the scope of the ALADIN project, focusing on a small num-

ber of methods. The effects of compression for a Schmidt photographic plate in the region of M 5 (numbered ESO 7992v), scanned with the CAI-MAMA facility, were examined. The digitized image is a mosaic of  $28 \times 28$  subimages, each of  $1024 \times 1024$  pixels. Sampling is 0.666 arcsec per pixel. This region was chosen because of the availability of a catalogue (Ojha et al. 1994) obtained from the same plate digitization, where positions and blue magnitudes had been estimated for 20 000 stars or galaxies of magnitude 10-19. The position of each object was ascertained by Ojha et al. by marginal Gaussian fitting to the intensity distribution. Magnitude was determined using 120 photometric standards, which allowed the magnitude-integrated density calibration curve to be specified.

To carry out our tests in a reasonable time and to avoid plate boundary effects, we analyzed 25 adjacent subimages, located at the centre of the photographic plate. We stress that these test images are real and not simulated. They are representative of the images distributed by the CDS's reference image service, ALADIN. The central region used for the astrometry and photometry measurements contains about 2000 objects whose magnitude distribution (from 14 for the brightest objects, to 19 for the faintest objects) illustrate that of the global population of the catalogue (Dubaj 1994).

Detection experiments (Carlsohn et al. 1993) have been performed to study the effect of compression on the preservation of faint objects. These were carried out on a test region, where 16 sources of estimated magnitude close to 21 were identified. Results are described in Table 3.

Detection errors (loss, and false detections) clearly increase with the compression rate. **FITSPRESS** loses the most objects, **HCOMPRESS** creates the most false objects. **JPEG** slightly better preserves faint objects but only below



**Fig. 1.** Comparison of the ability of the different packages to recover the position of objects according to object magnitude: astrometrical error increases with magnitude. We recorded the limit of magnitude above which the position error exceeds the catalogue precision: 0.1 pixel

compression rate of 40:1. In Carlsohn et al. (1993), we also compared the three methods with respect to the signal-to-noise ratio, positional and brightness error of known objects. We conclude that JPEG is better than HCOMPRESS at low signal-to-noise ratios, and is relatively similar at higher levels. Concerning the signal-to-noise ratio, the astrometry, and photometry, JPEG and HCOMPRESS produce images of equivalent quality, but FITSPRESS is again worse than the other two methods.

The first astrometrical tests were undertaken by Dubaj and are summarized in Fig. 1.

Star/galaxy discrimination was assessed by measuring the mean density by pixel, and considering the deviation of this quantity relative to its mean value for stars with the same integrated density as the object. Sufficiently low values are considered associated with galaxies. Applying this criterion to a subsample of around 1000 objects known a priori as stars or galaxies led to a contamination rate of 18% on the original image and 21% to 25% with compressed/uncompressed images (compression factor 40, for the three methods). This shows at least that morphological studies can be made on compressed/uncompressed images without substantial degradation.

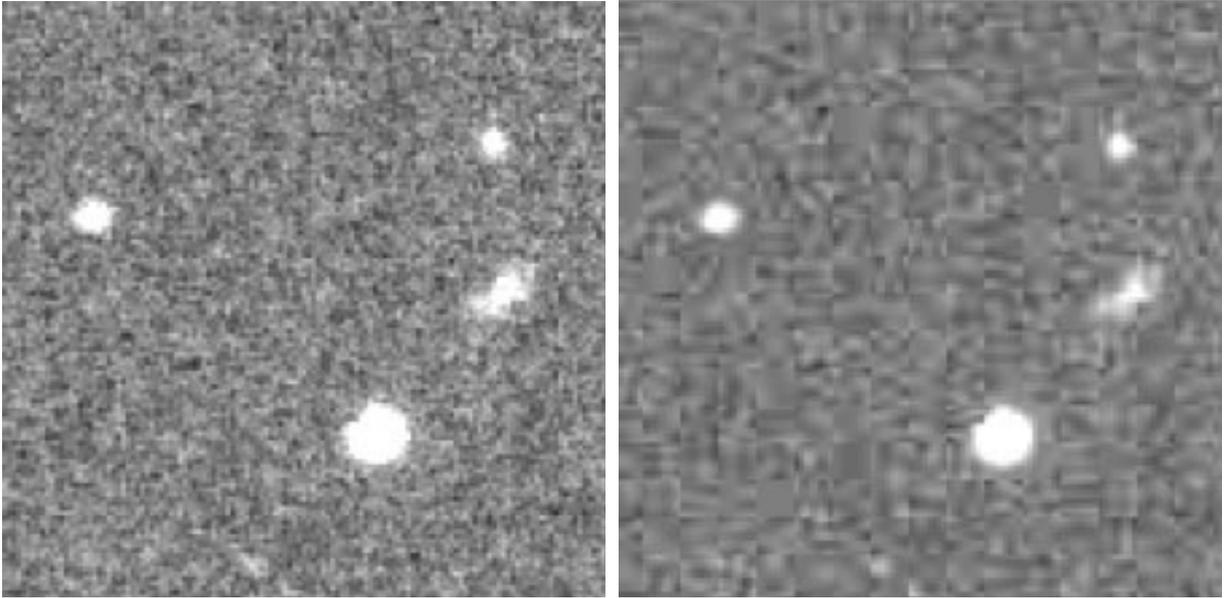
The general conclusion of this first study was that none of these methods could provide good visual quality above compression rates of 40:1 and that the standard JPEG method was ultimately not so bad, even if block artifacts appear. The available software (i.e., HCOMPRESS and FITSPRESS) developed in astronomy was not convincing in the framework of the ALADIN project. When the PMT method was proposed (Starck et al. 1996; Starck et al. 1998), a second study was carried out in order to compare

JPEG and PMT. In the meantime, MathMorph was implemented and underwent the same tests before its integration into the MR/1 package.

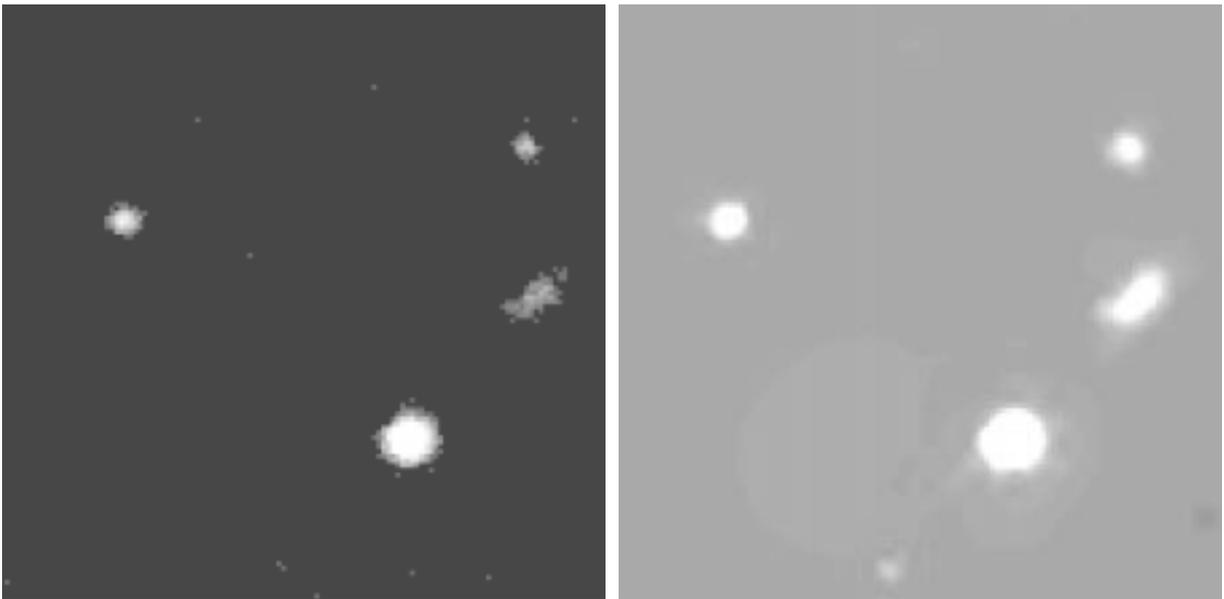
### 3.4. Second ALADIN project study

For the two compression methods studied here (JPEG and PMT), each implying loss of information, we have to look for good compromise between compression rate and visual quality. In the case of JPEG, various studies (Carlsohn et al. 1993; Dubaj 1994) confirm that beyond a compression rate of 40:1 this method of compression, when used on 12 bit/pixel images, gives rise to “blocky” artifacts. For PMT, as described in this article, the reconstruction artifacts appear at higher compression rates, beyond a rate of 260 in the particular case of our images. Figure 2 allows the visual quality of the two methods to be compared, for test image 325. A subimage of the original image is shown in Fig. 2.

To estimate the influence of compression algorithms on astrometrical precision of the objects, we studied the error in the position of the object in the original image compared to the position in the compressed/uncompressed image. This was done for each object in the catalogue. To determine the position of the objects, we used the MIDAS (ESO 1995) software routines based on fitting of two marginal Gaussian profiles as used originally by Ojha et al. (1994) for creating the catalogue. Knowing the catalogue magnitude of the objects, we can represent the mean positional error as a function of the object magnitude. This was done for magnitude intervals of 0.25 for the 2000 objects of the dataset used. Figure 4 allows the



**Fig. 2.** Left: Original image, subimage extracted from  $1024 \times 1024$  patch, extracted in turn from the central region of ESO7992v. Right: JPEG compressed image at 40:1 compression rate



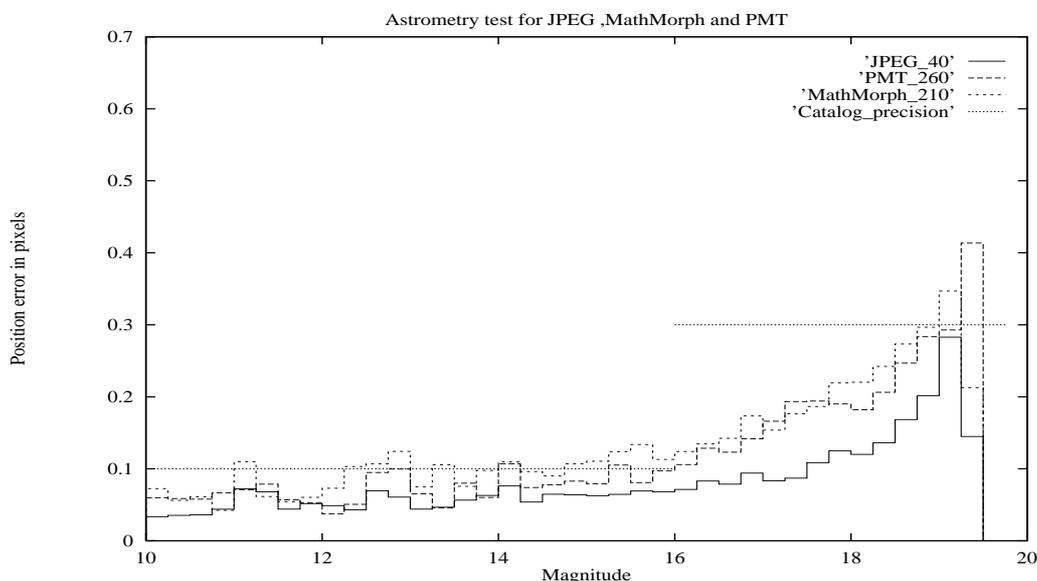
**Fig. 3.** Left: MathMorph compressed image of the same patch, at 203:1 compression rate. Right: PMT-compressed image at 260:1 compression rate

performances of JPEG and PMT to be compared. We note that for the two methods, the error is below the systematic error of the catalogue, in particular in the interval from object magnitudes 13 to 19 where we have sufficient objects to warrant asserting a significant result. Outside that interval, our dataset does not contain enough objects to establish a mean error in the astrometry.

Conservation of photometric properties is also a fundamental criterion for comparison of compression algorithms. We compared the integrated density of the objects in the 25 original images with the

corresponding integrated densities from the images compressed/uncompressed with PMT and with JPEG. This study was carried out in three stages:

- Detection of objects in the original image, and in the reconstructed image, and calculation of the integrated densities. This stage of the processing therefore gives a list of objects characterized by  $(x_o, y_o, d_o)$ , with  $(x_o, y_o)$  the coordinates of the barycentre, and  $d_o$  the logarithm of the integrated density. Similarly,  $(x_r, y_r, d_r)$  represents the list of objects detected under similar conditions in the reconstructed image.



**Fig. 4.** Mean error in astrometry, by interval of 0.25 magnitude, for images compressed 40 times by JPEG, 260 times by PMT, and 210 times for MathMorph

- Magnitude calibration of the original image and of the reconstructed image.
- Calculation of the error in the logarithm of the integrated density, by magnitude interval.

With each detected object, we associate its nearest neighbour in the catalogue according to the following rule, and we assign the corresponding catalogue magnitude,  $M_c$ , to the detected object: a detected object,  $(x, y)$ , is associated with the closest catalogue object  $(x_c, y_c)$  subject to their distance being less than or equal to 3 pixels. This finally provides two object lists:  $(x_o, y_o, d_o, M_{c_o})$ , for the original image, and  $(x_r, y_r, d_r, M_{c_r})$  for the reconstructed image. In a similar manner, the magnitude and logarithm of the integrated density association curves,  $M_{c_r} = f(d_r)$ , are studied for the JPEG- and PMT-reconstructed images. To verify the stability of the photometric values in spite of the compression, we hope to obtain curves, and thus to calibrate the reconstructed images, and to find dispersion around an average position which stays close to the dispersion obtained on the calibration curve of the original image. In fact, for varied compression methods, a systematic lowering of integrated densities of images is noted (Dubaj 1994), which results in the average calibration function (fitted by an order 3 polynomial) being slightly translated relative to the calibration function of the original image. To estimate the behaviour of the dispersion of the calibration curve for both compression methods, we proceeded thus:

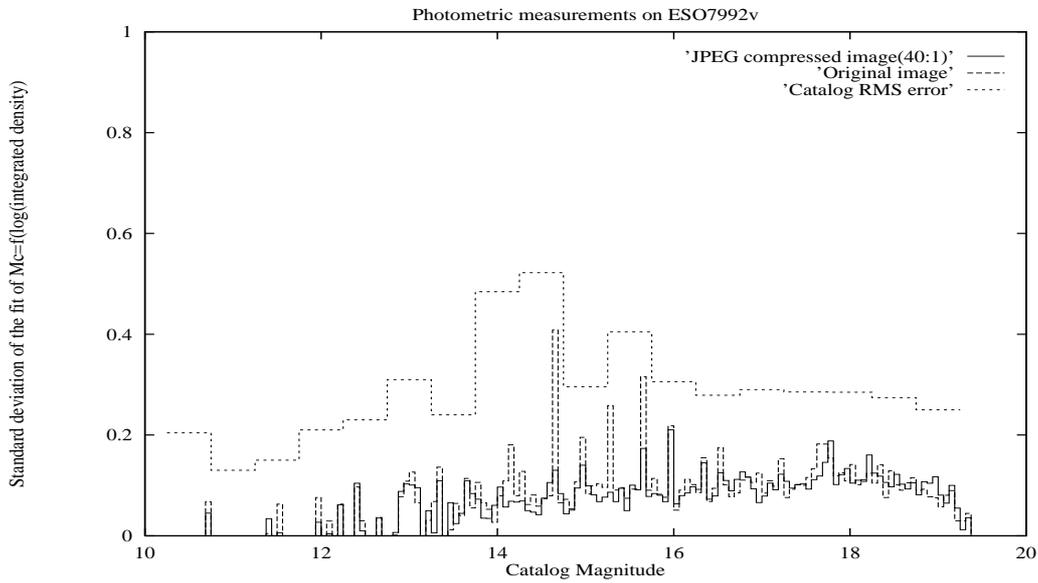
- Approximation by polynomial (degree 3) regression of the calibration function.  $M_c = f(d)$ .
- Calculation of the mean calibration error by magnitude interval, for the set of objects detected on the 25 subimages, i.e. about 2000 objects in all.

Thus we measure the photometric stability of the objects following compression, relative to their representation in the original image. The corresponding error curves are shown in Figs. 5, 6 and 7. The JPEG curve shows a slight increase for magnitudes above 18, and a smoothing effect for brighter objects between 14 and 16. For PMT, an increase in dispersion is noticed for high magnitudes, which corresponds to the problem of the detection of faint objects. Lowering the detection threshold from  $4\sigma$  to  $3\sigma$  does not change this. We note that the number of intervals below 14 is too small to allow for interpretation of the behaviour of very bright objects. Even if PMT brings about greater degradation in the photometry of objects, especially when the objects are faint, the errors stay close to that of the catalogue, and as such are entirely acceptable. Of course we recall also that the compression rate used with PMT is 260:1, compared to 40:1 for JPEG.

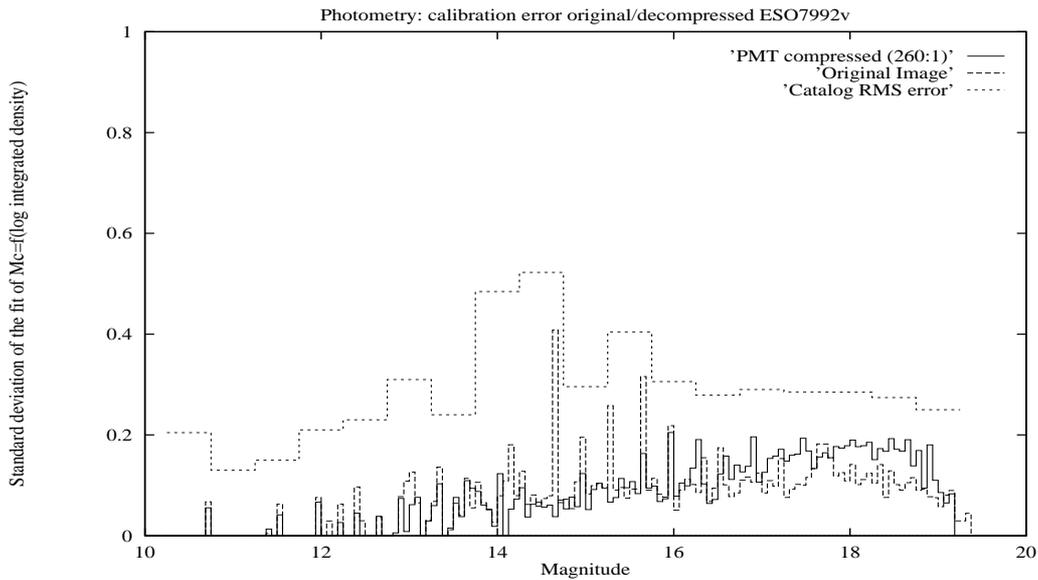
### 3.5. Computation time

Table 4 presents the computation time required for compression and decompression on a specific platform (Sun Ultra-Enterprise, 250 MHz and 1 processor). With the JPEG, wavelet, and fractal methods, the time to convert our integer-2 FITS format to a one-byte image is not taken into account. Depending on the applications, the constraints are not the same, and this table can help in choosing a method for a given project. The last column indicates if software already exists for progressive image transmission.

Thinking from a Web server point of view we would like to compare the performances of the different packages considering two scenarios:



**Fig. 5.** Comparison of the calibration error, by 0.0625 magnitude intervals, between the uncompressed image using JPEG, the original image, and the reference catalogue



**Fig. 6.** Comparison of the calibration error, by 0.0625 magnitude intervals, between the uncompressed image using PMT, the original image, and the reference catalogue

- Archive original and compressed images and distribute both on demand.
- Compress the data before transferring them and let the end-user decompress them at the client side.

This situation has been studied and is illustrated in Fig. 8. Considering a network rate of 10 Kbits/second and an image of 2 Mbytes, we measured the time necessary to compress, transmit and decompress the image. Methods are ordered from top to bottom according to increasing visual quality of the decompressed image. If we consider 20 seconds to be the maximum delay the

end-user can wait for an image to be delivered, only HCOMPRESS and PMT succeed, with less artifacts for PMT.

#### 4. Conclusion

The **MathMorph** method reproduces the image pixels up to a given threshold. The quality of the image depends on the estimate of the noise standard deviation before the application of **MathMorph** transformations. The method has good performance on uncrowded astronomical fields.

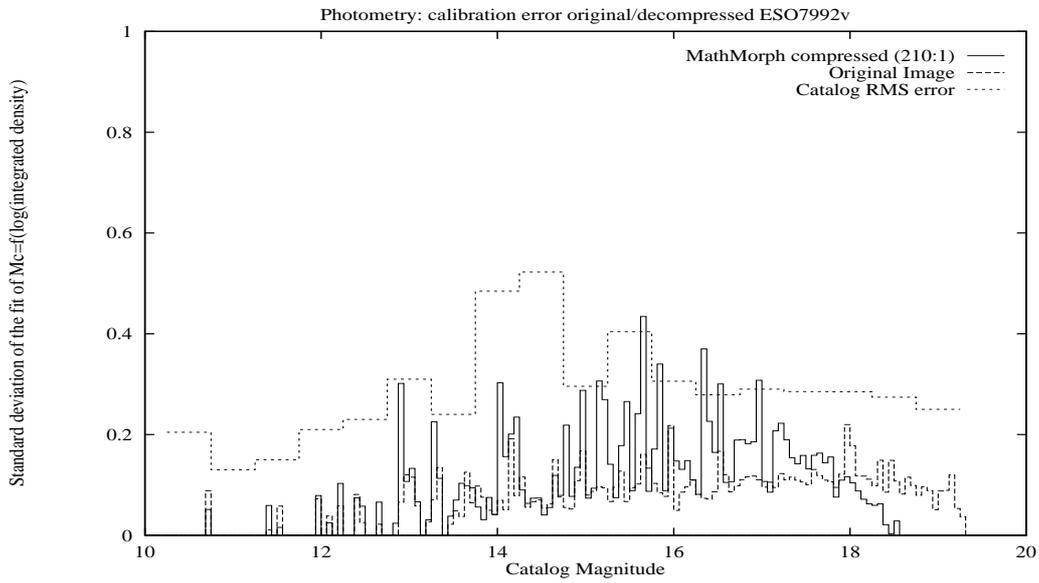


Fig. 7. Comparison of the calibration error, by 0.0625 magnitude intervals, measured on the uncompressed image using MathMorph, the original image, and the reference catalogue

Table 4. Compression of an 1024 × 1024 integer-2 image. Platform: Sun Ultra-Enterprise; 250 MHz and 1 processor

	Comp. time (sec)	Decomp. time (sec)	Artifact	Comp. ratio	Progressive transmission
JPEG	1.17	4.7	Y	<40	Y (in C)
Wavelet	45	7.1	Y	270	N
Fractal	18.3	9	Y	< 30	N
Math. Morpho.	13	7.86	N	< 210	N
Hcompress	3.29	2.82	Y	270	Y (in C)

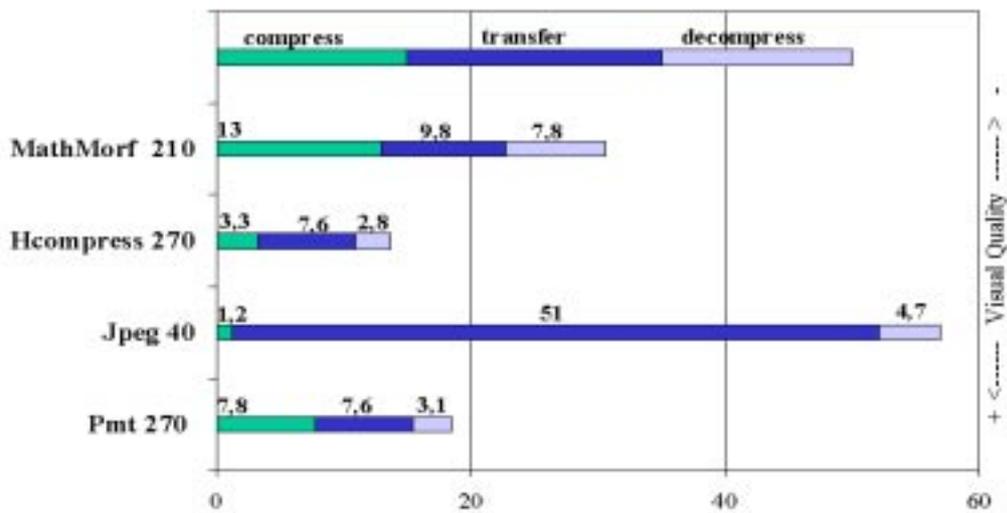


Fig. 8. Comparison of the overall time for compression, transmission and decompression for distribution of astronomical images using the Web: the network rate is supposed to be 10 Kbits/second, and the image size is 2 MBytes (1024 × 1024 × 2 bytes). The best preserving codecs with respect to visual quality are shown at the bottom of the graph

When a crowded field or an extended object is present in the image, the compression rate becomes much lower than the one obtained with the pyramidal median transform and, with traditional estimation of noise standard deviation, the faint extensions of objects and faint objects are lost in the compression.

The PMT method provides impressive compression rates, coupled with acceptable visual quality. This is due to the progressive noise suppression at successive scales. Nevertheless, on some crowded regions the PMT cannot compress more than 50:1, because a lot of objects' information is to be coded, in a small amount of image scales.

It is robust and can allow for certain image imperfections. On a Sun Ultra-Enterprise (250 MHz, 1 processor), compressing a  $1024 \times 1024$  image takes about 8 seconds (CPU time), with subsequent very fast decompression.

The decomposition of the image into a set of resolution scales, and furthermore the fact they are in a pyramidal data structure, can be used for effective transmission of image data (Percival & White 1996). Current work on Web progressive image transmission capability has used bit-plane decomposition (Lalich-Petrich et al. 1995). Using resolution-based and pyramidal transfer and display with Web-based information transfer is a further step in this direction. Java code implementing this functionality for PMT compressed images, is also available.

In the scope of the ALADIN project (Genova et al. 1999), we finally implemented the distribution of JPEG compressed images as a first step, due to the portability of JPEG among the different Web browsers. We are preparing to distribute PMT compressed images using the progressive resolution property in the near future.

*Acknowledgements.* We wish to thank J. Guibert, MAMA, for providing the test images, D. Dubaj and A. Schmidtstuhl for extensive tests on images, L. Huang and A. Bijaoui for helping in the tuning of the *MathMorph* approach, O. Bienaymé for providing the calibrated object catalogue, and J. Bartlett for helpful discussions. S. Mei acknowledges support from CNR, Italy. We are grateful to the referee, M. Albrecht, for his comments.

## References

- Bartlett J., Bonnarel F., Egret D., et al., 1996, in: *Astronomical Data Analysis Software and Systems V*, Jacoby G., Barnes J. (eds.), p. 489
- Bijaoui A., Bobichon Y., Huang L., 1996, *Vistas Astron.* 40, 587
- Carlsruhn M., Paillou P., Louys M., Bonnarel F., 1993, in: *Computer Vision for Space Applications*. Antibes, INRIA-ESA-Aérospatiale (ed.), p. 389
- Dubaj D., 1994, DEA Technical Report. Strasbourg Observatory
- ESO, 1995, MIDAS, Munich Image Data Analysis System. European Southern Observatory
- Fisher F., 1994, *Fractal Image Compression: Theory and Applications*. Springer-Verlag, Berlin
- Furht D., 1995, *Real-Time Imaging* 1, 49
- Genova F., Fernique P., Bartlett J., et al., 1999, in: *Astronomical Data Analysis Software and Systems VIII*, Jacoby G., Barnes J. (eds.) (in press)
- Huang L., Bijaoui A., 1991, *Exper. Astron.* 1, 311
- Lalich-Petrich V., Bhatia G., David L., 1995, in: *Third Int. WWW Conf.*, Holzappel (ed.), p. 159
- MR/1, 1998, *Multiresolution Image and Data Analysis Software Package, Version 1.0*, Multi Resolutions Ltd
- Murtagh F., Starck J.L., Louys M., 1998, *Int. J. Imag. Syst. Techn.* 9, 38
- Ojha D., Bienaymé O., Robin A., Mohan V., 1994, *A&A* 290, 771
- Percival J., White R., 1996, in: *Astronomical Data Analysis Software and Systems V*, Jacoby G., Barnes J. (eds.), p. 108
- Press W., 1992, in: *Astronomical Data Analysis Software and Systems I*, Worrall D., Biemesderfer C., Barnes J. (eds.), p. 3
- Proakis J., 1995, *Digital Communications*. McGraw-Hill
- Richter G. (ed.), 1998, *IAU Commission 9 Working Group on Sky Surveys Newsletter*. AIP Potsdam
- Starck J.L., Murtagh F., Bijaoui A., 1998, *Image and Data Analysis: the Multiscale Approach*. Cambridge University Press, Cambridge
- Starck J.L., Murtagh F., Louys M., 1997a, in: *Advanced Techniques and Methods for Astronomical Data Handling*, Bijaoui A. (ed.). Observatoire de la Côte d'Azur
- Starck J.L., Murtagh F., Louys M., 1997b, in: *Data and Information Fusion, Integration and Classification*, Molina R. (ed.), special issue, *Vistas in Astronomy*
- Starck J.L., Murtagh F., Pirenne B., Albrecht M., 1996, *PASP* 108, 446
- Taubman D., Zakhor A., 1994, *IEEE Transactions on Image Processing*, 3
- Véran J., Wright J., 1994, in: *Astronomical Data Analysis Software and Systems III*, Worrall D., Biemesderfer C., Barnes J. (eds.), p. 40
- White R., Postman M., Lattanzi M., 1992, in: *Digitised Optical Sky Surveys*, MacGillivray H., Thompson E. (eds.). Kluwer, Dordrecht, p. 167