



HP Forum Archive 14

[[Return to Index](#) | [Top of Index](#)]

Scanning advice needed !

Message #1 Posted by [Valentin Albillo](#) on 23 Feb 2004, 6:37 a.m.

Hi all:

I've just got hold of a very good scanner and I'm thinking about scanning my HP materials of old to keep them safely stored in CD/DVD and eventually make them available on the net. This includes tons and tons of program listings (most of them unpublished), HP publications (Digest, Keynotes, Journal, brochures, marketing materials, courses, internal documentation), user publications (Australian PPC Technical Notes), User's Library programs, and hundreds of HP-calc related letters to and from other HP fans of old (like John McGechie's or Tom Cadwallader's), not to mention SHARP programs and materials.

As you may see, the material is very varied, so the print quality is also extremely varied as well, and I would very much appreciate if some of you already having hands-on experience in digitizing printed materials would be so kind as to give me some advice in how to properly do it, i.e: technique, dpi resolutions, B&W/grey-scale/color, color depth, useful hints, gotchas, what-nots, whatever.

The work involved is already utterly overwhelming as it is so I might as well do it properly at first try. The final result will surely be a great addition to the worldwide HP fan community pool of past knowledge.

Thanks in advance and best regards from V.

Re: Scanning advice needed ! (somehow long...)

Message #2 Posted by [Vieira, Luiz C. \(Brazil\)](#) on 23 Feb 2004, 9:54 a.m.,
in response to message #1 by Valentin Albillo

Hi, V;

I'm glad you're gonna share your treasures with us here, and I cannot help telling you I'm a lot curious about it all. Thanks!

For what I can tell you about my own experience scanning and OCR processing (not that much, but about 90% successful results), I enumerate some procedures of mine I take as useful; hope they help you, too.

- try separating originals as for typewritten, with and without graphics (and set graphics as single with lines and complex with gray-scaled images), hand written with colored or black pen;
- originals with high contrast give better results if directly scanned with B&W default; originals with fade images should be scanned with 256 gray scale OR color set. This is because it's easier to manage already existing images with higher number of colors so you can reduce later instead of having to scan them back many times till you find what's best;
- it's not unusual a set of "What.. if"s when you are first trying; for sure, after a few successful results, you'll set your own intuitive parameters, and that's gold;
- dpi resolution is another issue: 300 dpi is sort of "all weather" choice. Most of regular printed material will do fine with 300 dpi. Lower resolutions are a matter of "What... if". Higher resolutions might apply only if originals use small typesetting (like the 82104A Card Reader Owner's Handbook, pages with to HP67/97 compatibility keystrokes) or have details that do not look fine with selected resolution;
- KEEP ALL OF YOUR ORIGINALS IN SAFE and process copies of them; if you decide what's done is not what you want, there will be no need to scan them again; anyway, if whatever you do does not give the best resulting images, start from scratch and scan the originals with a different setting (higher dpi setting, gray scale instead of B&W, etc...); naming directories with a reference for dpi and color scheme will also help you choosing the best; do not wipe out all first-scanned images, instead keep them in a "trash" directory so you can be sure they are useless after the job is done; you know that unwrinkling some thrown-away paper sometimes save people's lives... or may condemn them ;^)
- final e-document should be set to the lower number of colors you decide that better express original images. In some circumstances, a 90% reduction in size may be observed without any reduction in quality, only by reducing # of colors. You may reduce the # of colors with the software you decide to use to process the scanned image. Although the number of colors is almost always a target, the dpi resolution should not be altered unless you actually need to. IF you want to reduce the dpi of an existing e-document (and so it's size), KEPP THE ORIGINAL IMAGE and compare results. Enlarging a reduced image back to its original size will never bring pixels back (of course, it does not apply if your image processor offers the UNDO command), but there's no way to "see" the results of reducing an image if you don't actually do it;
- about printing and viewing: most of the times, WYSIWYG (What You See Is What You Get) does not apply because there are monitors and monitors and so there are printers and printers. Maybe the same e-doc looks much better than prints in one system and shows the opposite in others. So, do what pleases you most in your system and ask for a tryout to some friends and let them give you a feedback. This is some other source of information.
- final format is another issue: images that look great may suffer some sort of quality reduction when saved under some compression techniques. Most extensions (like JPEG, PCX and others) that are fairly used to compose images allow compression rates that may simply "destroy" what you have with the original resolution. What I can tell you is that B&W images (one dot per bit) are smaller when saved as PCX or TIFF (I never tried compression with BMP extension... to be honest, I don't even know if it's possible); gray scaled images also fit well as PCX and GIF, although GIF results in larger files for the same source images; although many guys hate JPEG high-compression images (they tend to generate "minute particles" surrounding borders), you may find a reasonable balance between image quality, compression rate and file size;

About pdf generation and image storage: I think PDF is a great distribution format, and it somehow "protects" ownership. I found that the best Windows-based SW (I think) to compose images into a final PDF "booklet" is Imaging, the standard Windows image/scanner manager. I don't like to use Imaging to process images or scanners, but I like it to compose a set of images into pages of the same document. It shrinks and stretches images with different sizes to fit inside a default page size, shrunk images do not loose resolution in final PDF (if you have the MoHPC CD's, have a look at the Portuguese version of the 82104A Owner's Handbook, HP67/97 compatibility) and final size is fairly acceptable. I don't know image processors used in other platforms (Mac, Linux-base PC's, etc.), but I

know that generating a PDF from an original doc under Linux is a standard procedure (default), and you'll need some extra, non-standard plug-ins to generate PDF in Windows.

Wow! I wrote too much. When we write too many things, there is a potential margin of errors... Anyway, I think I did not forget the main themes.

I'd like to add that this is all based in an original text I prepared (in Portuguese) to my friend José Ernesto, and I'm not sure I actually sent it at the time I wrote it... Zé, if I did not, please forgive me... <:^(

Hope this helps you, Valentin.

Cheers.

Luiz

Re: Scanning advice needed ! - one correction

*Message #3 Posted by **Vieira, Luiz C. (Brazil)** on 23 Feb 2004, 10:04 a.m.,
in response to message #2 by Vieira, Luiz C. (Brazil)*

Hi, folks;

please, where you read:

KEEP ALL OF YOUR ORIGINALS IN SAFE and process copies of them;

read instead:

*KEEP ALL OF YOUR ORIGINAL FILES WITH ORIGINAL SCANNED IMAGES IN SAFE and process **only the files with** copies of them;*

Not so much a meaningful suggestion.

Cheers.

Luiz (Brazil)

Re: Scanning advice needed ! (somehow long...)

*Message #4 Posted by **Valentin Albillo** on 23 Feb 2004, 10:25 a.m.,
in response to message #2 by Vieira, Luiz C. (Brazil)*

Hi, Luiz:

Luiz posted: *"I'm glad you're gonna share your treasures with us here, and I cannot help telling you I'm a lot curious about it all."*

Thanks a lot for your comprehensive and detailed help, I'll try and follow your savvy advices and I'm sure the results will be excellent. I'm not very sure whether I'll keep each page as a standalone image file or else I'll group them into PDF documents. Both approaches have its pros & cons. Another possibility is to group them into .cbz files, for very easy and convenient reading using CDisplay, while also compressing them a little more still.

As for your curiosity, believe me, there are things very rarely seen, if at all, such as the long questionnaire that HP sent us HP-67 owners in order to survey for wish-list features to be incorporated in a next model (the HP-41C, for sure) .. a great many mostly unpublished HP-67/97, 25C, 34C, 41C, and 71B programs .. not to mention the incredibly interesting (if only for historical reasons) correspondence I had with many great PPC contributors at the very beginning of 41C's synthetics exploration. Reading them gives you the feeling of being transported back to the golden days of PPC , almost like reading new, unpublished issues :-)

The one and only problem (apart from knowing proper scanning techniques) is that of sheer volume. It'll take a **long** while to get a significant portion of it in electronic format, and there's the problem of where to put that many megabytes so that they're accessible on-line to everyone interested. Not an easy matter, though I can always resort to rotating contents.

Thanks again and best regards from V.

Re: Scanning advice needed !

*Message #5 Posted by [Terry Ingram](#) on 23 Feb 2004, 10:23 a.m.,
in response to message #1 by Valentin Albillo*

Hello Valentin,

From a user's point of view, include only one page per scan. It's tempting to double-up and save scanning time, as was done with the HP-71b Reference Manual on the MoHpc DVD.

Thanks for sharing your information. Looking forward to seeing the SHARP information.

Terry

Re: Scanning advice needed !

*Message #6 Posted by [Valentin Albillo](#) on 23 Feb 2004, 10:31 a.m.,
in response to message #5 by Terry Ingram*

Hi, Terry:

Terry posted: *"Thanks for sharing your information. Looking forward to seeing the SHARP information."*

Certainly, some 100+ unpublished programs for the SHARP PC-1211 (TRS-80 PC-1) and later models, many of them pretty curious indeed.

Thanks for your interest and best regards from V.

Re: SHARP models

Message #7 Posted by **Terry Ingram** on 23 Feb 2004, 10:57 a.m.,
in response to message #6 by Valentin Albillo

Hi Valentin,

I have the TRS-80 PC-1, PC-2, & PC-3. These are really outstanding machines, with the PC-1 a joy to use.

But on the topic of sharing.. I have the "Game", "Business Finance", & "Personal Finance" packs for the PC-2 (sharp 1500). These are supplied on cassette tapes with also paper booklets. Do you know if anyone has been granted permission to share these materials. I would like to also share this information, but concerned about infringement.

Terry

Re: SHARP models

Message #8 Posted by **Valentin Albillo** on 23 Feb 2004, 12:30 p.m.,
in response to message #7 by Terry Ingram

Terry posted:

"I have the TRS-80 PC-1, PC-2, & PC-3. These are really outstanding machines, with the PC-1 a joy to use."

Agreed. It was in many aspects much better than its egregious contemporary, the HP-41C, though sadly many HP fundamentalists of the time wouldn't touch it with a ten feet pole. Good for them.

"I have the "Game", "Business Finance", & "Personal Finance" packs for the PC-2 (sharp 1500) [...] Do you know if anyone has been granted permission to share these materials. I would like to also share this information, but concerned about infringement."

The following pacs are available for public download at [Sharp PC-1500 computer \(TRS-80 PC-2\) resource page](#):

PC-2 Business Finance
PC-2 Chemistry Math
PC-2 Math Pak I
PC-2 Math Pak II
PC-2 Math Plotter
PC-2 Personal Finance
PC-2 Statistics
PC2word

among lots of other software, so it seems safe to assume that either permission has been granted or that site's webmaster couldn't care less about any copyright left on such old, obsolete, monetarily worthless software. Personally, I wouldn't give a damn, too, and would make it available as well. The risk of being 'sued' for doing so seems to me far less than the risk of an asteroid striking dead center on my very roof and I can live with that, thank you very much.

Best regards from V.

Re: SHARP models

*Message #9 Posted by [Terry Ingram](#) on 23 Feb 2004, 12:52 p.m.,
in response to message #8 by Valentin Albillo*

Hi Valentin,

Thanks for the link to the sharp/trs-80 resource page. I'm impressed with the amount of information & code listings available.

I had also contemplated porting some of the sharp/trs-80 code listings into hp-71b code. The graphics/games may be a problem, but I'm guessing most of the business stuff would convert relatively easy. I have noticed the hp-71b handles strings in a non-conventional way, also some common basic statements are named differently.

Not really sure if there would be any interest... I'm way off the original topic now. I should post a query regarding anyone's interest I guess.

Thanks, Terry

Re: Scanning advice needed !

*Message #10 Posted by [Angel Martin](#) on 23 Feb 2004, 3:30 p.m.,
in response to message #6 by Valentin Albillo*

Valentin,

looking forward to *finally* seeing and enjoying the programs from "Matematica Avanzada"... if I could bias your selection, can you start with these first?

Animo con la tarea!

Best, AM

Re: Scanning advice needed !

*Message #11 Posted by **Dave Shaffer** on 23 Feb 2004, 11:11 a.m.,
in response to message #1 by Valentin Albillo*

Valentin,

I second Luiz's choices of file types. I almost always produce PCX files as my first output from a scan. They are lossless and can always be changed into something else later if necessary. Further (lossless) compression (up to 80% or so) of the PCX files seems to be possible with ZIP. Producing PDF files from PCX files also seems to provide compression (with, as far as I can see) no loss of detail. I use Adobe Acrobat 5 for this.

Under NO CIRCUMSTANCES use JPG compression. It's not too bad for general images, but horrible for text and line drawings (in B&W or color).

As to resolution: if these are to be archival scans, it might be worth the time and space to "overscan." While 300 dpi will usually give acceptable results, I almost always use 600 dpi. Again, as Luiz notes, you can always throw things away AFTER you have scanned.

Another point: I have found with all my scanners (mid-range from all kinds of manufacturers - HP, Epson, Canon, Microtek) that despite expectations, a B&W scan does not produce the same fine resolution as a gray scale scan. It must have something to do with the sensor. If I need the best resolution, I generally scan in gray and then adjust the contrast and brightness in post-processing to give an essentially B&W image (again, at the cost of scanning time and storage space).

The bottom line: make some practice scans, and play with your post-processing software to decide on an optimal strategy.

Thank you very much ! [no text]

*Message #12 Posted by **Valentin Albillo** on 23 Feb 2004, 12:11 p.m.,
in response to message #11 by Dave Shaffer*

Indeed !

Best regards from V.

Re: Scanning advice needed !

*Message #13 Posted by **Vieira, Luiz C. (Brazil)** on 23 Feb 2004, 12:36 p.m.,
in response to message #11 by Dave Shaffer*

Hello, Dave;

thank you for your complementary and valuable information.

After reading your post I noticed that I forgot to add some info about OCR, and that PCX B&W scanned images with 300 dpi or more (depending on sharpness and typeset size) may give you text files too close to the original information. I'm achieving good results on recovering text information from scanned images. And it's a fact that 600 dpi is a lot better, and most OCR software will generate fewer errors when applied to higher resolution images.

If information is the target and time to do it is not an issue, text documents generated from images with OCR are the ones that will occupy the smallest space, as you surely know about. And you can re-design the page as you wish. Anyway, if historical reasons are the target, then let's keep them as they are, with the highest resolution possible ;^)

Best regards and thank you again.

Luiz (Brazil)

Re: Scanning advice needed !

*Message #14 Posted by **Bill (Smithville, NJ)** on 23 Feb 2004, 12:13 p.m.,
in response to message #1 by Valentin Albillo*

Hi Valentin,

Here's some ideas that I use when scanning:

For Black-White scans, use 300 DPI for Letter size, 600 DPI for smaller size, or for letter size with small text, such as program listings.

Based on the originals, select a standard size paper space to scan to. That way all the scanned images are the same size.

Use an imaging program to clean up the scans. I use the Kodak Imaging program that comes with Windows. Zoom in on the scan, then delete any artifacts such as staple/punch holes/or dust specs. Although this takes quite a bit of time to do for each scanned page, it makes a world of difference to the finished product. I hate seeing black marks up the side of a scanned page. It also saves a lot of ink when it's printed out.

If scanning a manual, include the blank pages - no need to scan them - just make sure the PDF file includes them. That way the PDF file can be printed to a double sided printer and recreate the original with the pages numbers on the correct page edge.

If the original requires, scan to either gray or color. Use an imaging program to increase contrast/color balance. Then reduce the number of colors to a lower level.

I just finished scanning the Advantage module manual. Each page was scanned at 8.42" by 5.49", at 300 DPI color. Loaded each page into Microsoft Photo Editor, increased contrast to 60%. This made the background completely white. Then loaded each page into Kodak Imaging, erased the edge spiral binding marks. I left it at true color, but could have lowered it to 256 colors to save space. Added the blank pages and created PDF. The finished file when printed, using Adobe scale to Fit Page, creates a great Advantage Manual enlarged to letter size. I just bound it and use it for daily reference.

12345 to delete

Valentin, some of this is already scanned FYI

*Message #15 Posted by **Gene** on 23 Feb 2004, 1:32 p.m.,
in response to message #1 by Valentin Albillo*

Gene: HI. Thought I'd make sure you knew that some of this is already available ... might save you some effort. Of course, if you don't like how it has already been done, feel free to scan it better. :-)

You wrote: This includes...

- a) tons and tons of program listings (most of them unpublished),
- b) HP publications (Digest, Keynotes, Journal, brochures, marketing materials, courses, internal documentation),

GENE: All the keynotes are scanned and on disks offered by Jake Schwartz. He also has all (?) of the calculator/portable computer Journal articles scanned. I think the Digests are done.

- c) user publications (Australian PPC Technical Notes),

GENE: Jake did these last year. He had CD's with these on it at the HHC 2003 conference.

- d) User's Library programs, and

- e) hundreds of HP-calc related letters to and from other HP fans of old (like John McGeachie's or Tom Cadwallader's), not to mention SHARP programs and materials.

Re: Valentin, some of this is already scanned FYI

*Message #16 Posted by **Valentin Albillo** on 24 Feb 2004, 4:36 a.m.,
in response to message #15 by Gene*

Thanks for the caveat, Gene, much appreciated. I'll follow your advice and avoid re-scanning those materials, then.

Best regards from V.

Re: Valentin, some of this is already scanned FYI

Message #17 Posted by **Gene** on 24 Feb 2004, 7:30 a.m.,
in response to message #16 by Valentin Albillo

Well, take a look at how they were scanned and then decide if you like it or not. :-)

Can't wait to see the goodies you have.

Gene

Re: Scanning advice needed !

Message #18 Posted by **David Ramsey** on 23 Feb 2004, 5:20 p.m.,
in response to message #1 by Valentin Albillo

OK, before I start, here are my bona fides:

- * I was a programmer on both Macintosh and Windows versions of Caere's "OmniPage" OCR program for 8+ years.
- * I've been scanning old calculator programs for Gene since October of last year.

That said, here's my advice:

1. Forget OCR for the program listings. Most of the documents you scan will likely be complex enough to require extensive manual tweaking. Also, as you noticed, the quality varies dramatically and even for the good documents, the OCR engine's recognition dictionary won't have most of the terms anyway.

I've been scanning hundreds of old programs (ftp://ftp.neko.com/HP_Docs) and have simply reproduced each page as a 300dpi monochrome image (600dpi in a few cases).

2. PDF is the only way to go. You ABSOLUTELY don't want to use some proprietary solution or one that's only available on a specific platform.

3. A scanner with a sheet feeder is virtually a must. I've been using the HP 5550c, a nice little scanner that sells for \$300 in the US. It's inexpensive, reasonably fast (about 4 pages per minute sustained), and can handle double-sided originals.

BTW, feel free to copy and distribute or archive any of the stuff I've done so far.

Re: Scanning advice needed !

*Message #19 Posted by [Ángel Martín](#) on 24 Feb 2004, 3:18 a.m.,
in response to message #18 by David Ramsey*

David, quite an impressive archive! I can only start to *imagine* all the work you've put to it, a great achievement. Thanks for sharing it with all of us.-

Best, ÁM

Thank you very much for ALL your advices ! :-)

*Message #20 Posted by [Valentin Albillo](#) on 24 Feb 2004, 4:37 a.m.,
in response to message #1 by Valentin Albillo*

Much appreciated, indeed.

Best regards from V.

Re: Thank you very much for ALL your advices ! :-)

*Message #21 Posted by [Ángel Martín](#) on 24 Feb 2004, 8:33 a.m.,
in response to message #20 by Valentin Albillo*

Valentín, your project is intrinsically interesting to all of us, and thus the warm response.

In fact, I'm hardly containing my enthusiasm to finally see you unearth all those gems from their sleep. One thing that it's been popping up in my mind is making the *ultimate* compilation of math programs, for which I'm more than sure yours will more than qualify. Can't you already see the "ADV MATH" ROM label?? :-)

Does that intrigue you? I'll be extremely happy to be your ROM compiler if you feel like getting into this project! You can reply directly to my e-mail if you want.

Saludos, Ángel