

11/9/2019 command line - Why are the images produced by pdfimages different when using the -all flag? - Ask Ubuntu



[Why are the images produced by pdfimages different when using the -](https://askubuntu.com/questions/776679/why-are-the-images-produced-by-pdfimages-different-when-using-the-all-flag)

[all flag?](https://askubuntu.com/questions/776679/why-are-the-images-produced-by-pdfimages-different-when-using-the-all-flag)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Asked 3 years, 5 months ago | | | | | Active [8 days ago](https://askubuntu.com/questions/776679/why-are-the-images-produced-by-pdfimages-different-when-using-the-all-flag?lastactivity) | | | Viewed | | 4k times | | | |
|  | It's my understanding that | | | | | |  | | --- | | pdfimages -all | | | | extracts images from PDFs in their native formats. | | | | |
| 5 | Therefore, I expected that the JPG (lossy) images extracted from that command would have the | | | | | | | | | | | | |
| same pixel information as the .ppm and .pbm files produced without the | | | | | | | | | | | |  | | --- | | -all | | option, as well as the |
|  | PNG (lossless) files created when I right-click and save the image in Evince. | | | | | | | | | | | | |
| 1 | However, my use of the ImageMagick | | | | | | | |  | | --- | | compare | | | | command tells me that there are differences in the | | |
| images contained within the JPG files compared to the other options above. To reproduce,  download the PDF in this link ( <https://fccid.io/document.php?id=2149405>), use it as an argument for | | | | | | | | | | | | |
| |  | | --- | | pdfimages | | | | and | |  | | --- | | pdfimages -all | | | | and use the first .ppm file and the first .jpg file as arguments for | | | | | | |
| |  | | --- | | compare | | | . When I do this, it produces an image file containing red to indicate a difference in the | | | | | | | | | | | |

images.

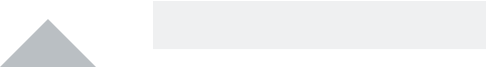
Is there something that I don't understand? Is creates .ppm and .pbm files?

|  |  |  |
| --- | --- | --- |
| |  | | --- | | pdfimages | | adding pixel information by default when it |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| command-line | | pdf | image-processing | | | | | | | | imagemagick | | | | | | | | | | | | | | | | |
| [edited Nov 1 at 17:15](https://askubuntu.com/posts/776679/revisions) | | | | | | | | | | | | | | | | | | | | asked May 24 '16 at 5:51 | | | | | | | |
|  | | | | | | | | | | | | [Zanna](https://askubuntu.com/users/527764/zanna) | | | | 15 | |  | 150 | |  | 253 | [Orion751](https://askubuntu.com/users/364609/orion751) | | | | |
| **53.9k** | | |  | **38** | 1 | 5 | | |
|  | Just how much difference is there between these images? Can you supply examples? – [John1024](https://askubuntu.com/users/221266/john1024) May 24 | | | | | | | | | | | | | | | | | | | | | | | | | | |
|  | '16 at 6:38 | | | | | | | | | | | | | | | | | | | | | | | | | | |
|  | @John1024 I'm trying to get images displaying the problem, but the PNG's seem to be too large for Stack | | | | | | | | | | | | | | | | | | | | | | | | | | |
|  | Overflow/Imgur. – | | | | | [Orion751](https://askubuntu.com/users/364609/orion751) May 24 '16 at 22:30 | | | | | | | | | | | | | | | | | | | | | |
|  | @John1024 Would a link to a PDF source be of any help? To reproduce, download it, use it as an | | | | | | | | | | | | | | | | | | | | | | | | | | |
|  | argument for | | pdfimages | | | | | and | | pdfimages -all | | | | and use the first .ppm file and the first .jpg file as | | | | | | | | | | | | | |
| arguments for | | | | compare | | | . When I do this, it produces an image file containing red to indicate a difference | | | | | | | | | | | | | | | | | | | | |
| in the images - [fccid.io/document.php?id=2149405](https://fccid.io/document.php?id=2149405) – | | | | | | | | | | | | | | | | [Orion751](https://askubuntu.com/users/364609/orion751) May 25 '16 at 3:36 | | | | | | | |  | | | |
|  | I tried that and saw some red. I also tried using | | | | | | | | | | | | |  | | --- | | convert | | | | | | to convert a jpg to ppm (no pdf involvement) | | | | | | | | | |
|  | and then running | | | | compare | | | | on the two; I still got differences. It might be that there are some rounding- | | | | | | | | | | | | | | | | | | |
| error issues with the conversion process that | | | | | | | | | | | | |  | | --- | | convert | | | | | | detects. – [John1024](https://askubuntu.com/users/221266/john1024) May 25 '16 at 7:18 | | | | | | | | | | |
|  | @John1024 Are you suggesting that these rounding-error issues may also apply to | | | | | | | | | | | | | | | | | | | | | | | |  | | --- | | pdfimages -all | | | | ? |
|  | Prior to using a newer version of | | | | | | | | | | pdfimages | for the | | | | -all | | | | flag, I was able to use | | | | convert | | on the | |

example to convert it from .ppm to .png without getting red, but I also got red when I tried converting to .jpg

|  |  |  |  |
| --- | --- | --- | --- |
| with it. – | Orion751 May 26 '16 at 17:04 |  |  |
| https://askubuntu.com/questions/776679/why-are-the-images-produced-by-pdfimages-different-when-using-the-all-flag | | | 1/3 |

11/9/2019 command line - Why are the images produced by pdfimages different when using the -all flag? - Ask Ubuntu

|

1 Answer

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 7 | pdfimages -all | | returns the *exact* file that was stored in the pdf. | | | |
| We can test this by doing a round-trip: starting with a jpg image, we add it to a pdf using LaTeX, | | | | | |
| extract it using | |  | | --- | | pdfimages -all | | | , and then compare it to the original. (The reason for using LaTeX | | |
|  | will be explained later.) | | | | | |
| I have the first jpg image as extracted from your link and I named it | | | |  | . Let's put it in a |
|  | |  | | --- | | device.jpg | |
|  | PDF file using LaTeX: | | | | | |

|  |
| --- |
| $ cat img.tex |

\documentclass{article}

\usepackage{graphicx}   
\begin{document}

\includegraphics[width=5in,keepaspectratio]{device}   
\end{document}

$ pdflatex img   
[...snip...]

Output written on img.pdf (1 page, 672455 bytes).

|  |  |  |
| --- | --- | --- |
| Transcript written on img.log. | | and compare it with the original: |
| Now, let's extract it using | |  | | --- | | pdfimages -all | |

|  |
| --- |
| $ pdfimages -all img.pdf img-all |

|  |
| --- |
| $ cmp device.jpg img-all-000.jpg |

|  |
| --- |
| $ |

The extracted jpg is *byte-for-byte identical* to the original.

**Footnote: the reason for using LaTeX**

The above test cannot be done using just any PDF creator. This is because not all PDF creators

|  |  |  |  |
| --- | --- | --- | --- |
| will put images into a PDF unmolested. For example, let's try ImageMagick's | |  | | --- | | convert | | : |

|  |
| --- |
| $ convert device.jpg device.pdf |

|  |
| --- |
| $ pdfimages -all device.pdf device-all |

|  |
| --- |
| $ cmp device.jpg device-all-000.jpg |

|  |  |
| --- | --- |
| device.jpg device-all-000.jpg differ: byte 4, line 1 | |
| |  | | --- | | convert | | re-sampled the image to a smaller size before placing it in the pdf. |

|  |
| --- |
| $ ls -1s device.jpg device-all-000.jpg |

|  |
| --- |
| 528 device-all-000.jpg |

|  |
| --- |
| 656 device.jpg |

Image accuracy was part of pdflatex's design goals. Other PDF creation software may, by default,

"optimize" images before placing them in the PDF.

https://askubuntu.com/questions/776679/why-are-the-images-produced-by-pdfimages-different-when-using-the-all-flag 2/3

11/9/2019 command line - Why are the images produced by pdfimages different when using the -all flag? - Ask Ubuntu

answered May 27 '16 at 19:40

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | [John1024](https://askubuntu.com/users/221266/john1024) | | |  | 40 |  |
| **11.1k** |  | 29 |
| https://askubuntu.com/questions/776679/why-are-the-images-produced-by-pdfimages-different-when-using-the-all-flag | | | |  |  | 3/3 |