

Limitations of Open-Source OCR Software versus OmniPage

Accuracy

Generally, Open-Source OCR is not as accurate as more advanced commercial solutions.

- Open-Source OCR solutions typically either completely lack image pre-processing algorithms or require significant trial and error to implement pre-processing algorithms that work well for most documents. These include:
 - Adaptive binarization (to deal with shaded backgrounds)
 - Deskewing (including 3D deskewing for photographs).
 - Adaptive binarization.
 - Adaptive resolution enhancements.
- Open-Source OCR solutions are typically slower than commercial offerings.

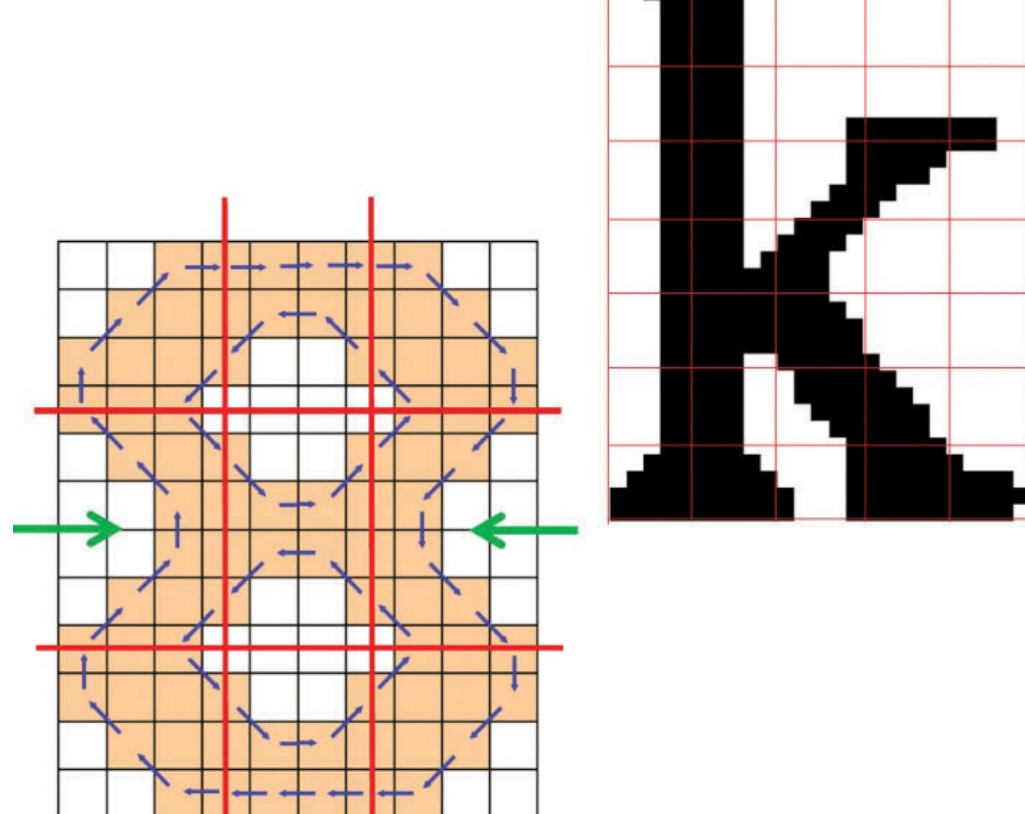


Figure 1: Optical Character Recognition is the process of recognizing images of characters and converting these into text. The OmniPage solution contains three separate (fully Tungsten owned) character classifiers.

Key Takeaway:

The OmniPage solution offers a more feature rich offering as a whole in comparison to open-source.

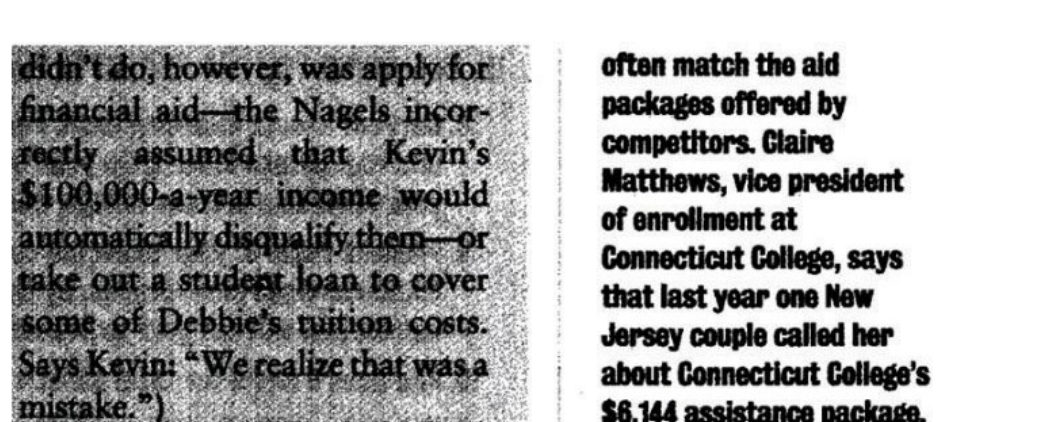


Figure 2: Example of image pre-processing that improves the quality of input images

Support

- Open-Source OCR does not offer any support. Any issues that arise become the responsibility of your development team.
- There is no design assistance offered with Open-Source OCR it is up to your development team to figure out how to best implement it.

Key Takeaway:

Open-source won't give you any support. If issues arise they are your problem, which requires more resources and needed supported.



Figure 3: Page Decomposition is the process of breaking up the input image into one or more logical, homogenous zones

Customer Success

- Your marquee customers will need to be informed that you are downgrading a key component of your solution. This may prompt them to either ask for discounts or investigate other vendors.
- Many of your competitors are using commercial OCR solutions, switching could put you at a competitive disadvantage.
- Change of OCR engines will increase your technical support calls. Things that worked before no longer function.

Key Takeaway:

A shift from OmniPage is a downgrade in the solution. The result? Added support calls plus opening the door for a competitive disadvantage.

Development Costs

Ripping out your current commercial OCR solution and replacing it with Open-Source will require significant development effort. This effort could instead be spent improving other aspects of your products.

- Getting Open-Source OCR solutions fine-tuned to perform similarly to your current solution is a large task. You will need to invest resources into this project and may eventually discover that you cannot meet your customer's requirements with Open-Source. All of the invested efforts will have then been wasted.

Key Takeaway:

With a rip and replace there are added costs for development. To fine tune open source to your document there are necessary resources required. OmniPage has the experience in many use cases.

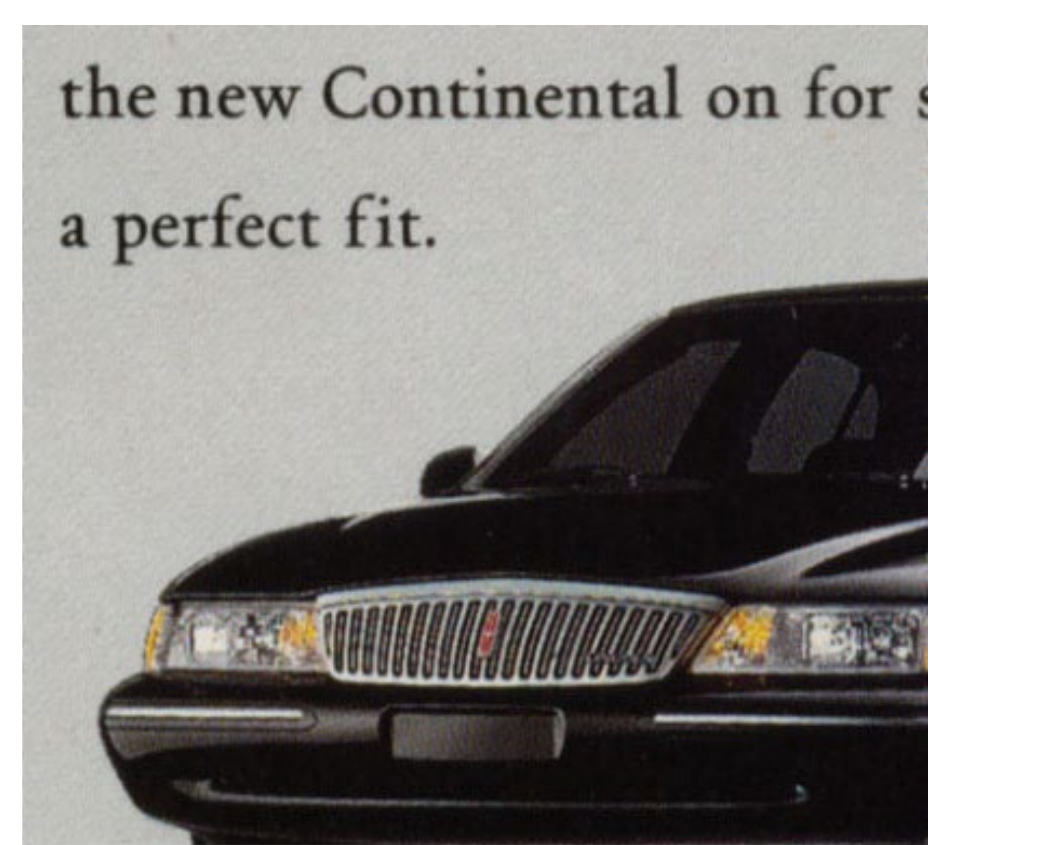


Figure 4: As you can see, the image quality of the car is almost identical to the original and the text is even cleaner with Level 5 MRC compression.

Security

Open-Source solutions do not follow a Secure Software Development Life Cycle (SSDLC). This is often a requirement for large corporate/government accounts.

- If any security vulnerabilities arise in your Open-Source solution, they become your security vulnerabilities. Your development team will need to halt other projects so that they can address these issues. With a commercial offering, your OEM partner will take care of addressing any security issues that arise.

Key Takeaway:

Open Source does not guarantee adherence to secure processes and is not held to any particular compliance. Any vulnerabilities are solely your responsibility which opens the door to many issues.

To learn more about the value OCR can add to your solutions, download our white paper,

Tech Talk: How OCR Adds Value to Your Organization