

## Context

Recto verso registration can be used on documents suffering from bleed-through to :

- ▶ detect missing digitized pages,
- ▶ locate of the bleed-through defect over a page,
- ▶ measure the bleed-through amount by analyzing at the same time both the recto side and the verso side.

Several recto-verso registration are proposed :

- ▶ a parameter optimization method [1] aims to find the appropriate transformation matrix that minimizes the difference between the recto,
- ▶ the second method [2] uses a Fourier-Mellin transformation.

These methods have several drawbacks :

- ▶ high time computation costs,
- ▶ fail in some cases, where bleed-through is not high enough, without warning the user.

## Objectives

- ▶ Better time computation costs.
- ▶ Accuracy unchanged.
- ▶ Alert cases where registration cannot be done with enough accuracy.

## Over-all algorithm

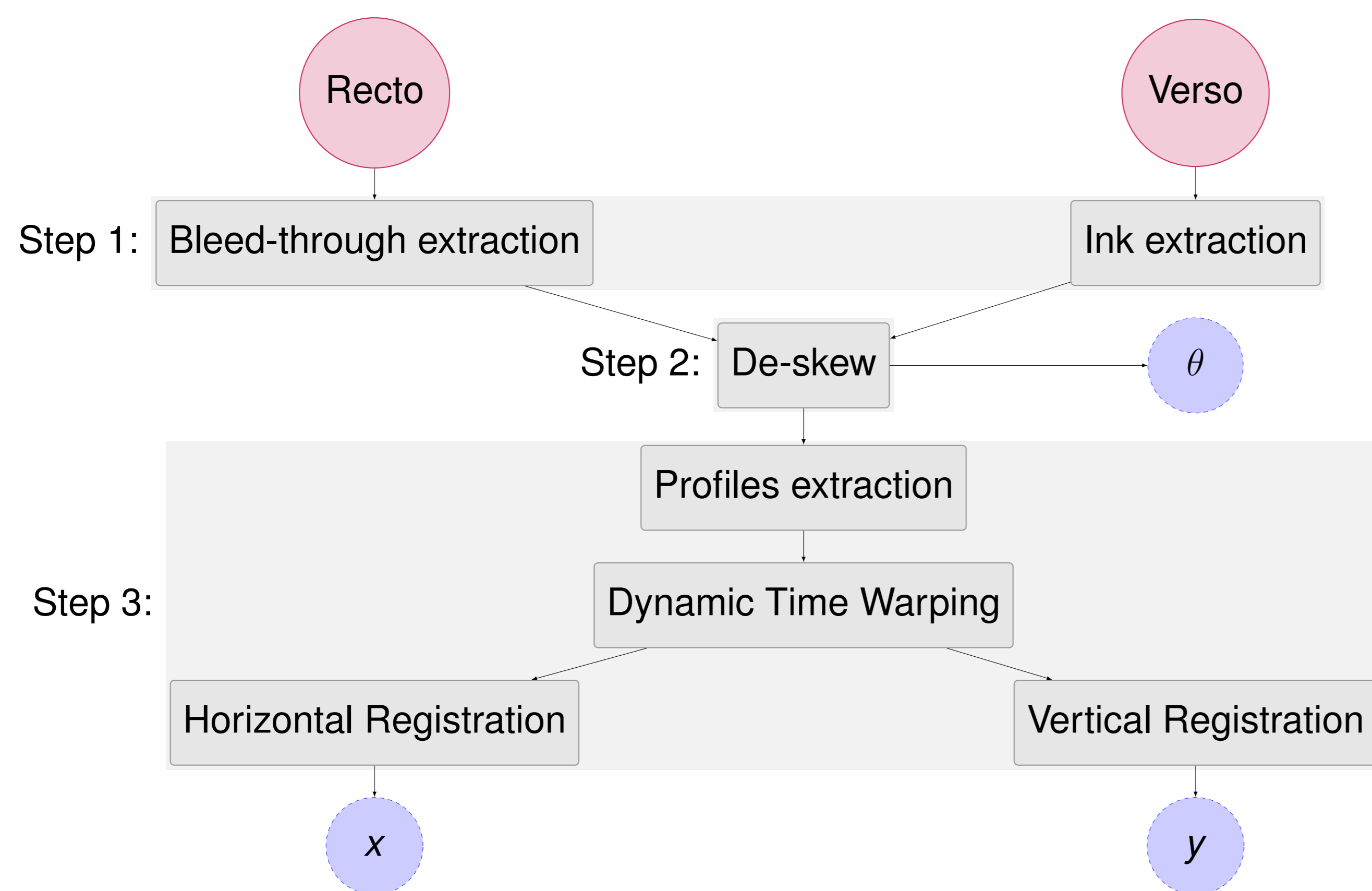


Figure: The overall registration method : dashed circles are the transformation parameters,  $\theta$  for the rotation,  $x$  for the horizontal shift and  $y$  for the vertical shift .

- ▶ Step 1 : Rough Trinarization of the recto and verso images.
- ▶ Step 2 : De-skewing of both pages :
  - Extracted bleed-through pixels on the recto.
  - Ink pixels on the verso.
- ▶ Step 3 : Profiles extractions and registration :  $x$  and  $y$  parameters are estimated by average shift.
  - Horizontal and vertical thresholded profiles extraction.

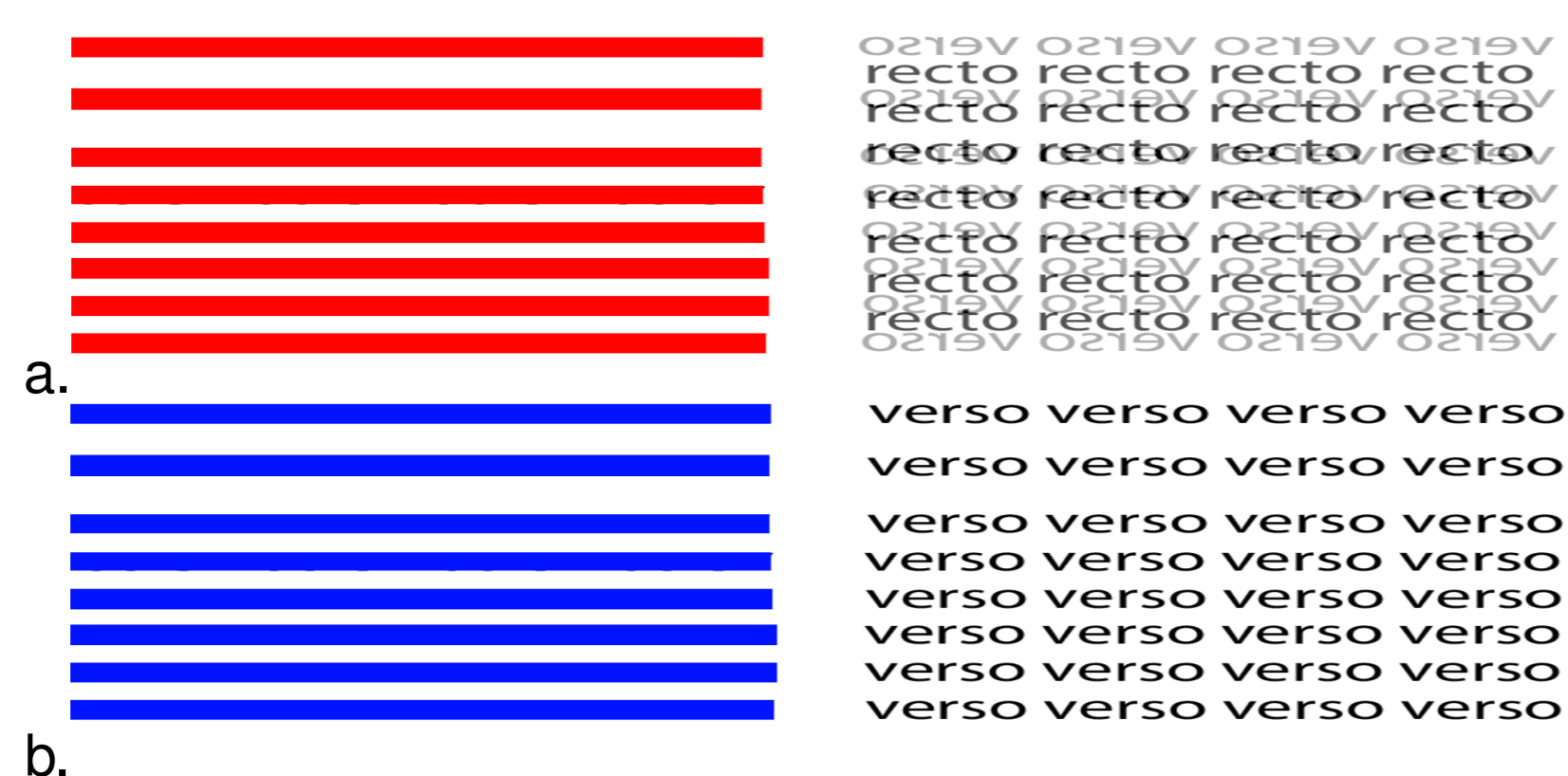


Figure: a. Horizontal profile of the recto's bleed through. b. Horizontal profile of the verso's ink

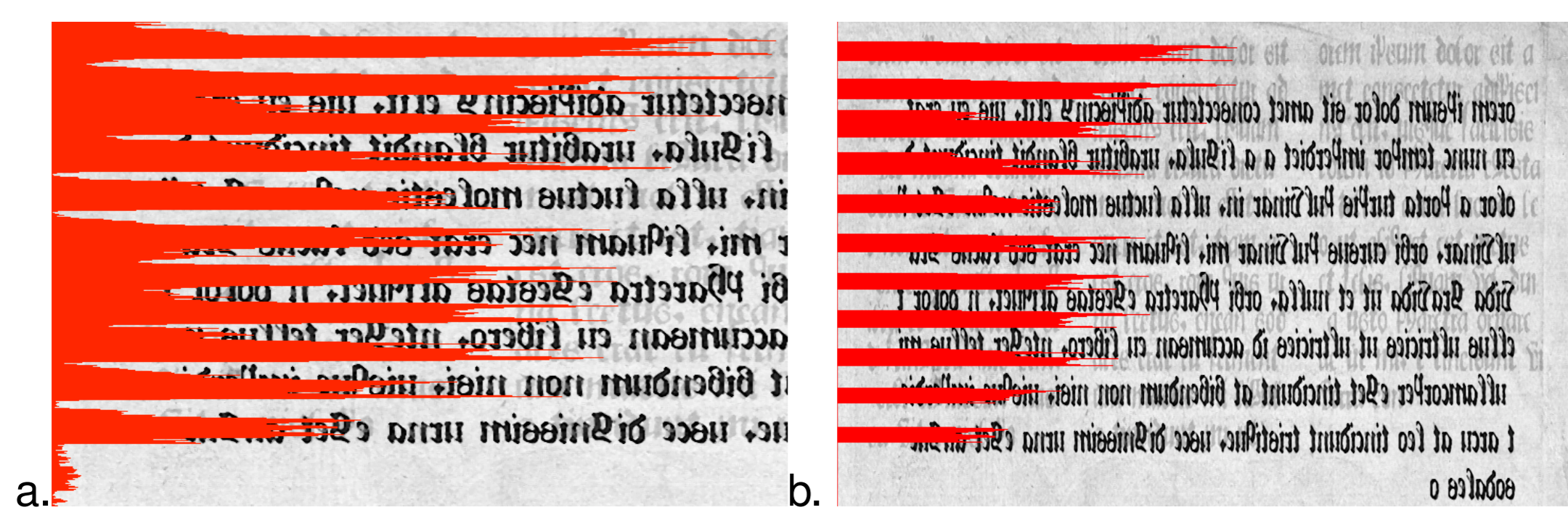


Figure: bleed-through lines extraction with vertical profile: a. the raw profile, b. the cleaned up profile: all irrelevant bins are removed.

- Registration of both profiles with a DTW.

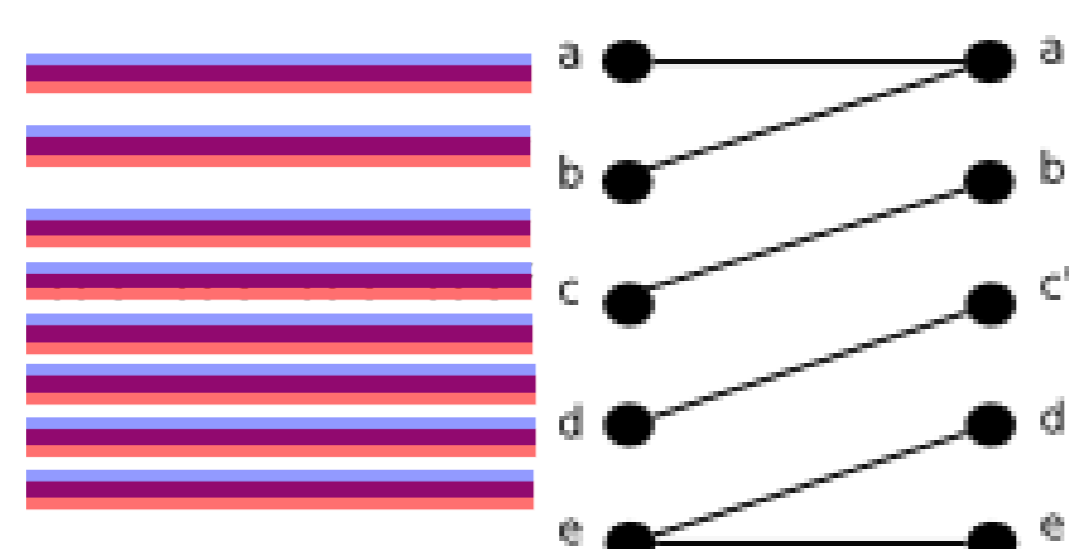


Figure: The DTW matching results in bins that are either *matched* :  $\{(b, a'), (c, b'), (d, c'), (e, d')\}$ , *inserted* :  $\{(a, a')\}$  or *deleted* :  $\{(e, e')\}$ . Circles on the left and on the right correspond respectively to bins on the first profile and the second profile.

- Accuracy measure :

- based on the normed DTW error.
- the higher the DTW error is the less bins have been matched.
- if the DTW error is higher than a given threshold, the registration can be cancelled.

## Experiments and Results

- ▶ Synthetic data set.

- With a defect model [3] we create a recto verso pair with 10 bleed-through levels.
- Then a random transformation is applied. ( $\theta = + - 20$ ,  $y = + - 20\%$ ,  $x = + - 20\%$ )
- Images are 2000x2829 px big.
- A total of 500 images.

- ▶ The registrations methods are implemented in C++ and tested on a computer with 8Go 1067 MHz DDR3 and an Intel Core i7 @ 2.8 Ghz.

Table: Registration methods accuracy comparison

Registration Method	Skew Angle Error			
	Max	Min	Mean	Standard Deviation
Our Method	0.25	-0.03	0.15	0.06
Dubois's Method	18	0	7.19	4.45
Registration Method	Horizontal Shift Error			
	Max	Min	Mean	Standard Deviation
Our Method	11	0	1.17	2.10
Dubois's Method	39	0	2.04	6.77
Registration Method	Vertical Shift Error			
	Max	Min	Mean	Standard Deviation
Our Method	1	0	0.51	0.53
Dubois's Method	38	0	1.81	5.04
Registration Method	Mean computation time			
Our Method	12s			
Dubois's Method	598s			

- ▶ Accuracy measure :

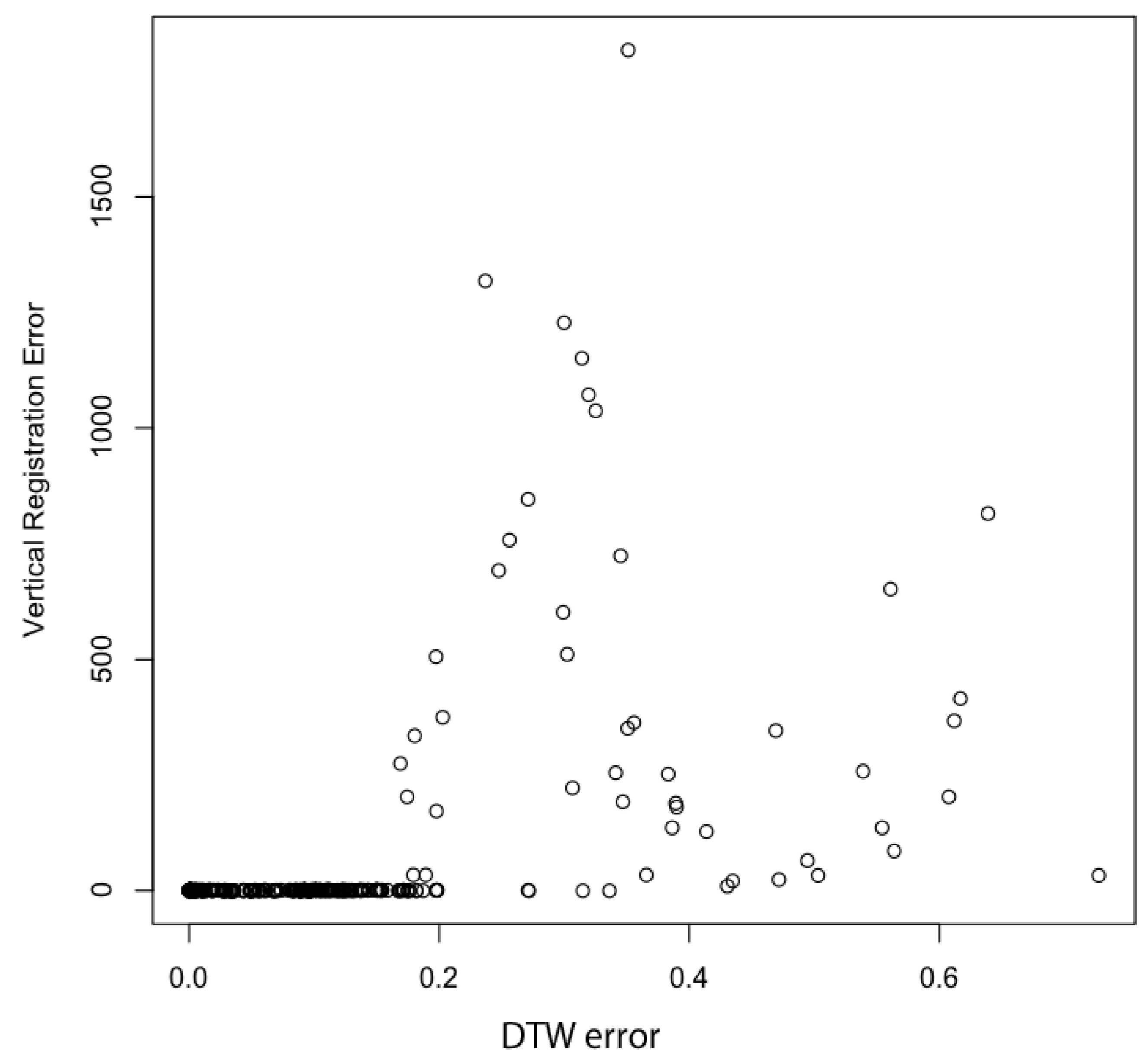


Figure: Relation between the DTW error (x axis) and the vertical registration error (y axis). Accurate registrations (vertical error close to 0) have a DTW error value lower than 0.19. When the DTW error is higher, we can not guaranty a pixel accurate registration. (500 images : 57 registrations failed (10%))

## Conclusion

- ▶ A new way to register a recto with its corresponding verso.
- ▶ Uses a dynamic time warping algorithm to match horizontal and vertical profiles of both the bleed-through on the recto and the ink on the verso.
- ▶ Accuracy unchanged and 50 times faster than existing techniques.

## Perspectives

- ▶ Applying this method to non-affine transformation.
- ▶ Complete the experiments on a larger data-set of real documents with their ground-truths.

## References

- [1] E. Dubois and A. Pathak, Reduction of bleed-through in scanned manuscript documents, in IS AND TS PICS CONFERENCE. SOCIETYFORIMAGINGSCIENCE&TECHNOLOGY, 2001, pp. 177180.
- [2] L. Hutchison , Fouriemellin registration of line delineated tabular document images, International Journal on Document Analysis and Recognition, Jan 2006.
- [3] R. Moghaddam and M. Cheriet, Low quality document image modeling and enhancement, International Journal on Document Analysis and Recognition, vol. 11, no. 4, pp. 183 201, 2009.