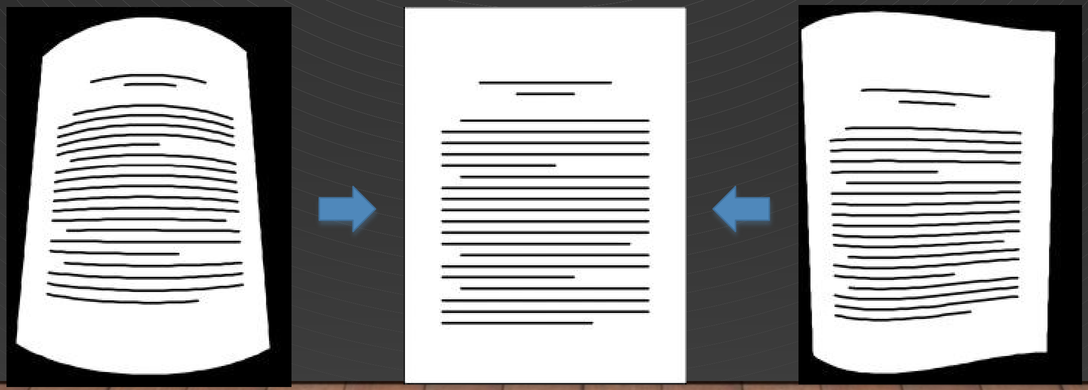
DEWARPING CURVED DOCUMENT IMAGES WITH MATLAB, COURSE PROJECT

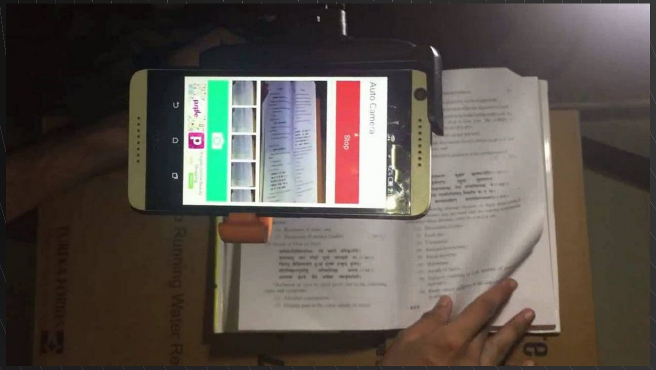
CPSC 635 - IMAGE ANALYSIS AND COMPUTER VISION - FALL 2017

CHI ZHANG



M. Zucker, “Page dewarping,” 2016. [Online]. Available: https://mzucker.github.io/2016/08/15/page-dewarping.html. [Accessed: 26-Oct-2017].

INTRODUCTION: DIGITIZING DOCUMENTS WITH CELL PHONE CAMERA



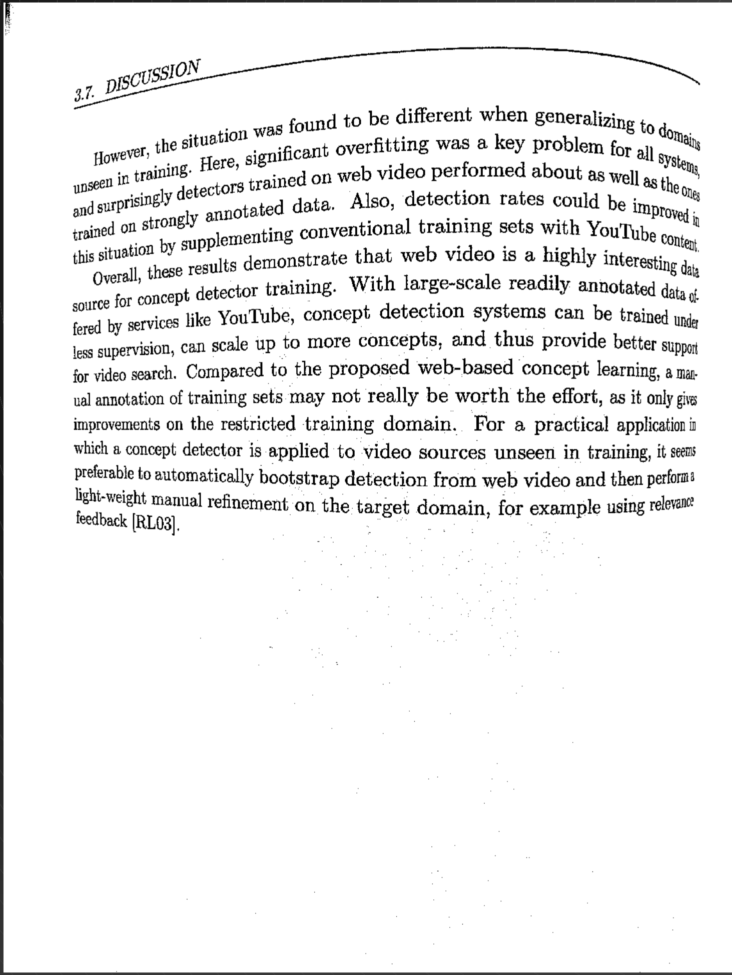
Digitizing books

Foreign language translation

Archiving rare documents

MOTIVATION: MOBILITY VS ACCURACY

DISTORTED IMAGE



3.7. DISCUSSION

However, the situation was found to be different when generalizing to domains unseen in training. Here, significant overfitting was a key problem for all systems, and surprisingly detectors trained on web video performed about as well as the ones trained on strongly annotated data. Also, detection rates could be improved in this situation by supplementing conventional training sets with YouTube content.

Overall, these results demonstrate that web video is a highly interesting data source for concept detector training. With large-scale readily annotated data offered by services like YouTube, concept detection systems can be trained under less supervision, can scale up to more concepts, and thus provide better support for video search. Compared to the proposed web-based concept learning, a manual annotation of training sets may not really be worth the effort, as it only gives improvements on the restricted training domain. For a practical application in which a concept detector is applied to video sources unseen in training, it seems preferable to automatically bootstrap detection from web video and then perform a light-weight manual refinement on the target domain, for example using relevance feedback [RL03].

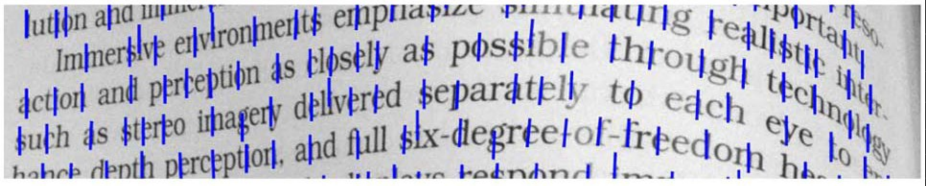
OCR RECOGNIZED TEXT

trained 011
_ A1 , detection rates co ld '
gly annotated data S0 u be Improvedin
_ _ 1 t . . t .
plementing conventiona raining se s with YouTube content.
ults demonstrate that Web Video IS a highly interesting data
tector training. With large-scale readily annotated data0f_
d by Services like YouTube, concept detection systems can be trained under
supervision, can scale up to more c0I1C€3Pl'{Sa and thus Provide better Support
for video search. Comparedto the proposed Web-based concept learning, a man.
ual annotation of training sets rnayrnot really be Worth the effort, as it only gives
improvements on the restricted training domain', For a practical applicationin
which a concept detector is applied to 'video sources unseen in training, it seems
trai
. - u
this situation by 5 P
Overall, these res

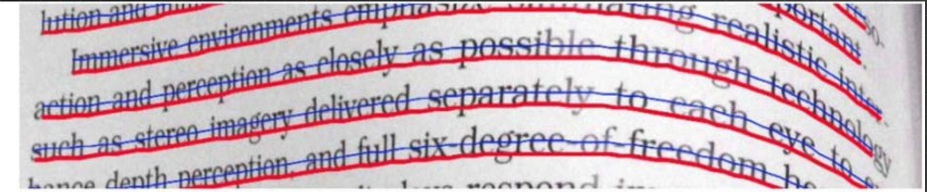
BACKGROUND: METHODS

Warp: to bend or twist out of shape, especially from a straight or flat form

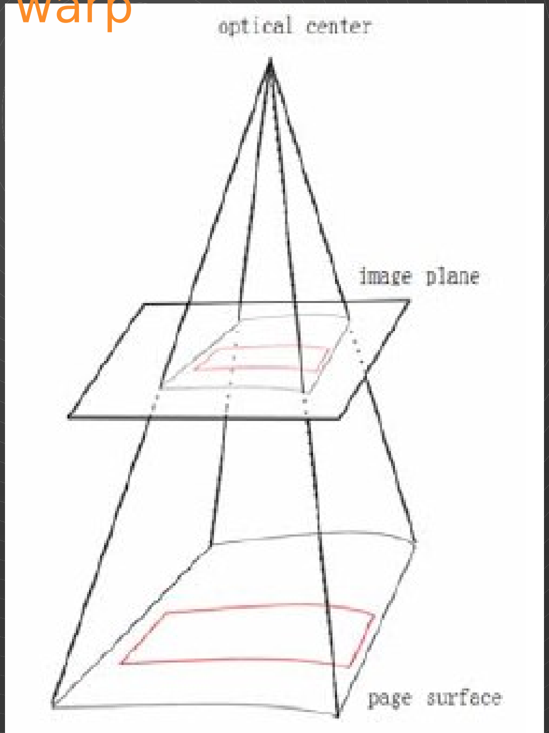
Dewarp: reverse process of warp



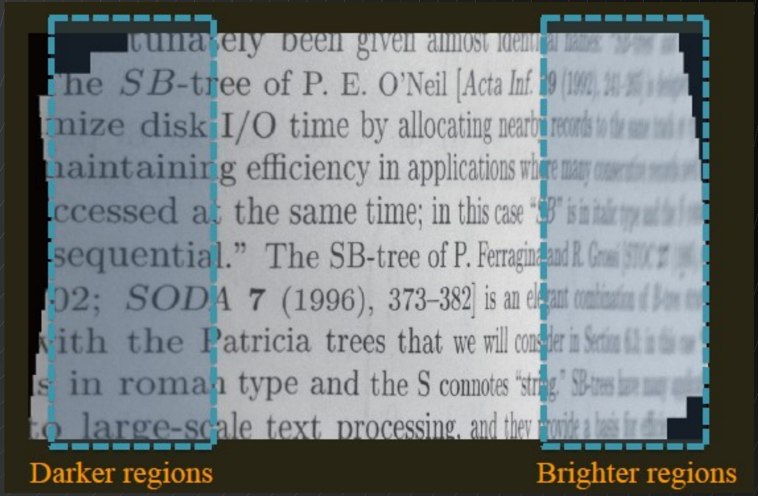
Stroke analysis



Baseline detection



3D transform model



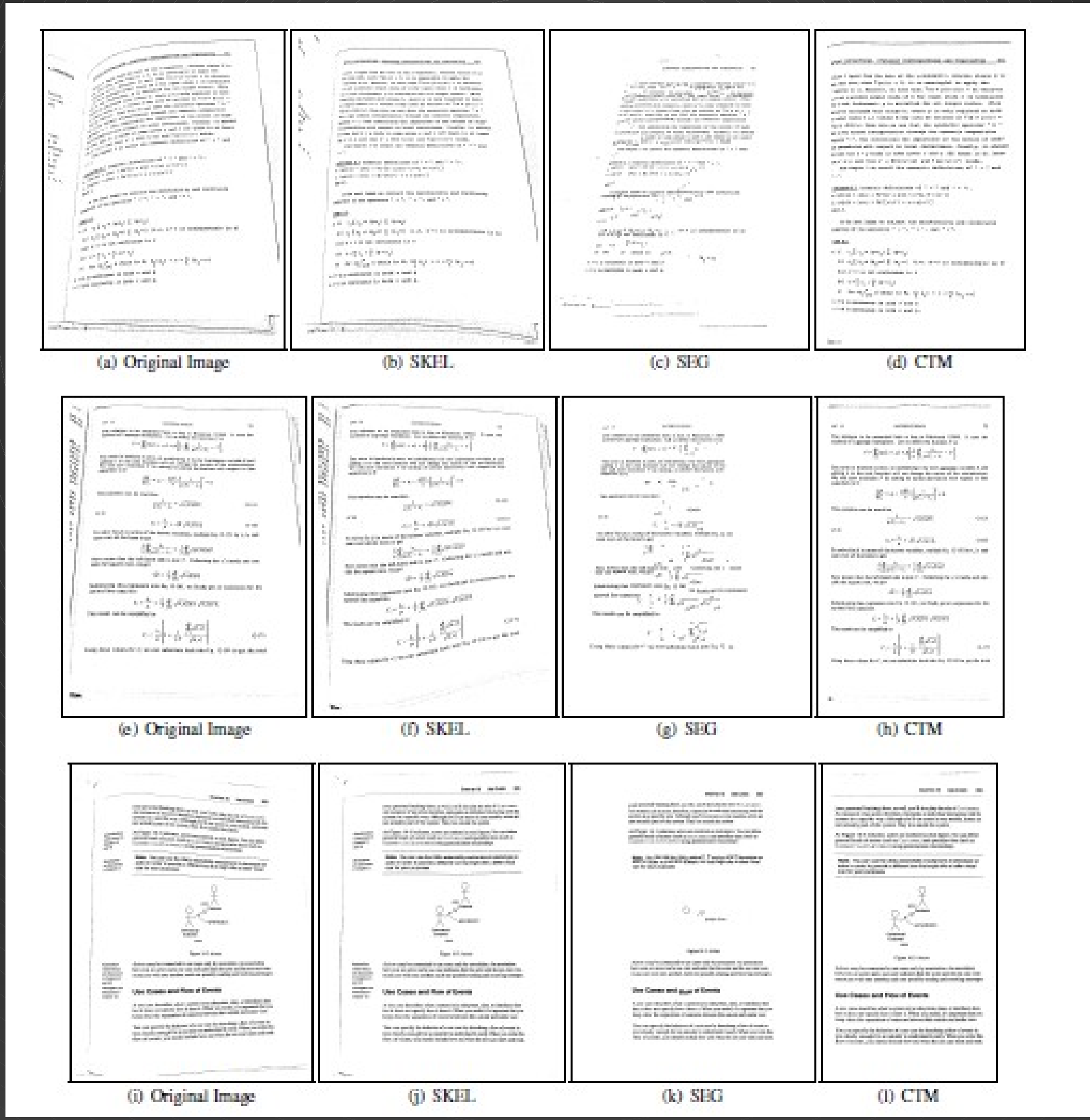
Brightness contrast

Y. Tian and S. G. Narasimhan, “Rectification and 3D reconstruction of curved document images,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 377–384, 2011.

B. Fu, M. Wu, R. Li, W. Li, Z. Xu, and C. Yang, “A model-based book dewarping method using text line detection,” Proc. 2nd Int. Work. Camera ..., pp. 63–70, 2007.

BACKGROUND: RESULTS

5



Best reported results: OCR accuracy 95-100%

F. Shafait and T. Breuel, “Document Image Dewarping Contest,” *2nd Int. Work. Camera-Based Doc. Anal. Recognit. (CBDAR-2007), Sept. 22, Curitiba, Brazil.*, pp. 181–188, 2007.

METHODOLOGY (1)

- Objective: dewarp document images with Matlab implementation; no external libraries
- IUPR 2011 Dataset (images of warped and scanned documents)
- http://didcontest2011.blogspot.ca/
- Image dimensions: 2592 pixels x 3456 pixels
- Assumptions
    1. images of individual pages from book
    2. focal plane almost normal to page surface
    3. text areas are in upright orientation (preprocessed through rotation)

METHODOLOGY (2)

Step 0. PREPROCESSING: Complement and erode (remove salt and pepper noise)

Step 1. Dilate horizontally and erode vertically to turn each textline into a connected component

Refinement

Step 2. Connected component analysis; textline fitting with 3rd order polynomial

Step 3. Text area left/right border fitting with 2nd order polynomial

Refinement

Step 4. Calculate intersection points between textlines and left/right borders

Step 5. Calculate reference points (local min/max of textlines, borders)

Step 6. Generate grid points (moving points)

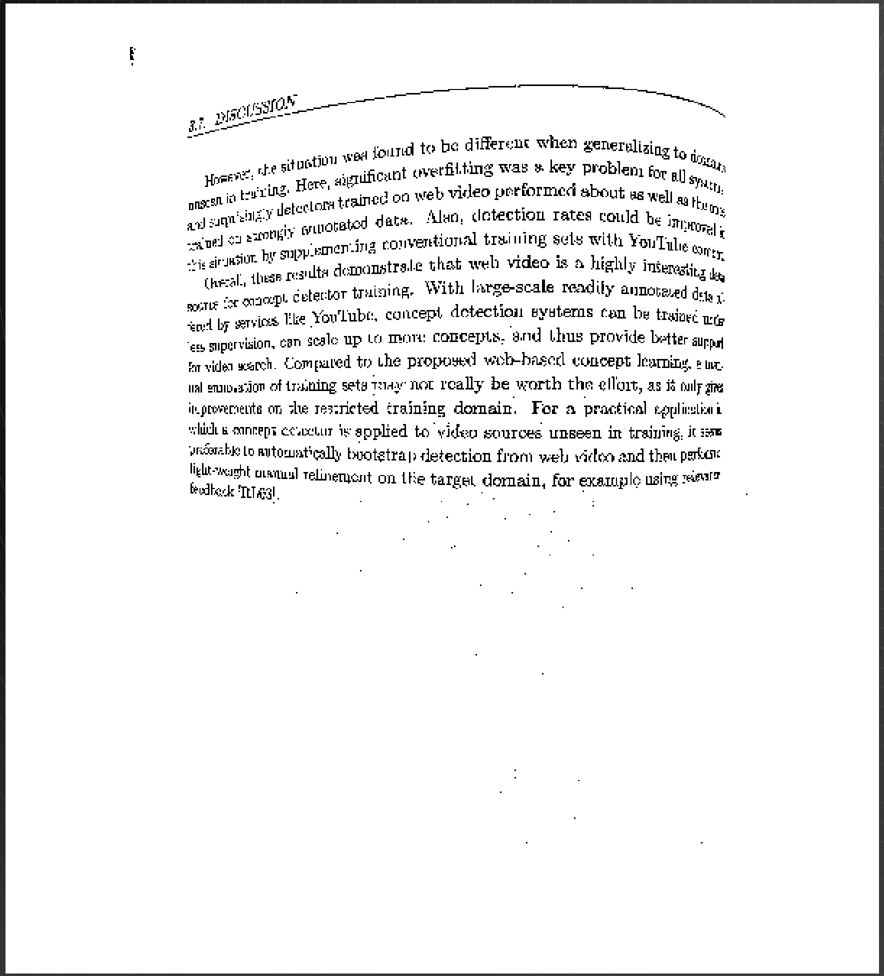
Step 7. Generate rectified grid points (fixed points)

Step 8. Polynomial transformation

STEP 0. COMPLEMENT AND ERODE TO REMOVE SALT AND PEPPER NOISE

8

BEFORE

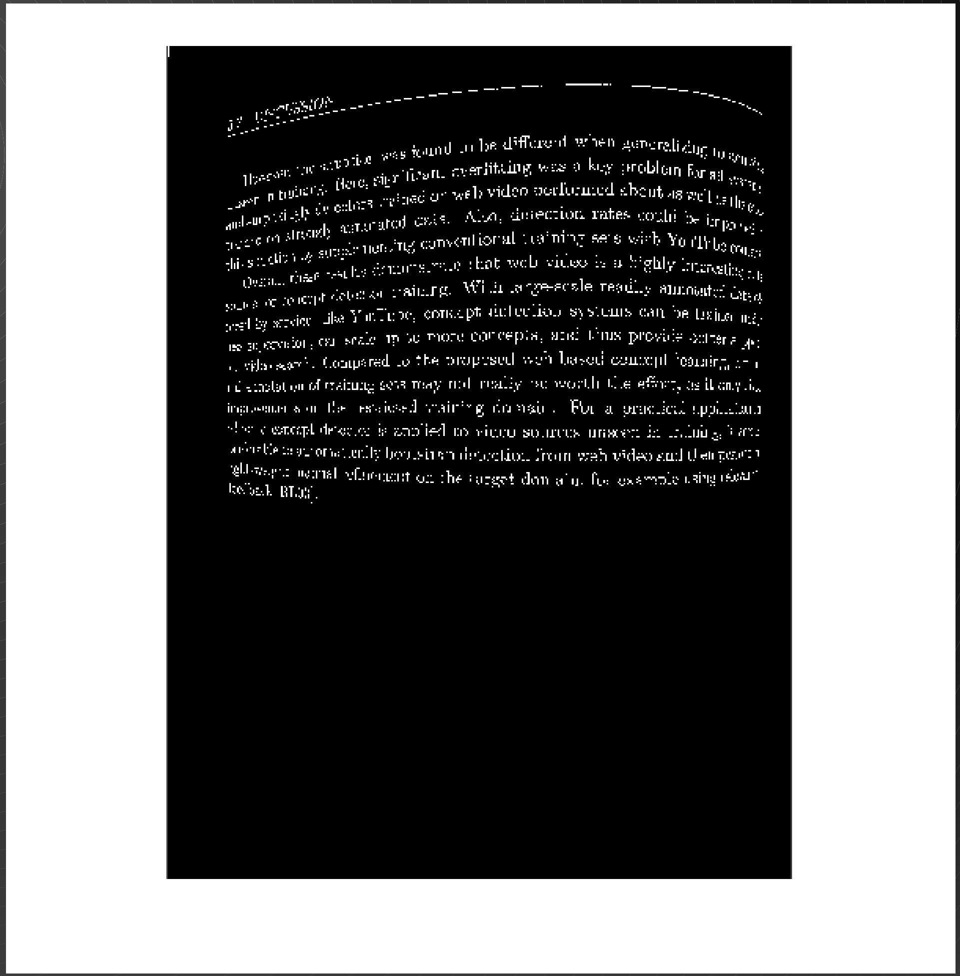


3.7 DISCUSSION

However, the situation was found to be different when generalizing to domains unseen in training. Here, significant overfitting was a key problem for all systems, and surprisingly detectors trained on web video performed about as well as the ones trained on strongly annotated data. Also, detection rates could be improved in this situation by supplementing conventional training sets with YouTube content.

Overall, these results demonstrate that web video is a highly interesting data source for concept detector training. With large-scale readily annotated data provided by services like YouTube, concept detection systems can be trained under less supervision, can scale up to more concepts, and thus provide better support for video search. Compared to the proposed web-based concept learning, a manual annotation of training sets may not really be worth the effort, as it only gives improvements on the restricted training domain. For a practical application in which a concept detector is applied to video sources unseen in training, it seems preferable to automatically bootstrap detection from web video and then perform light-weight manual refinement on the target domain, for example using relevance feedback [TL63].

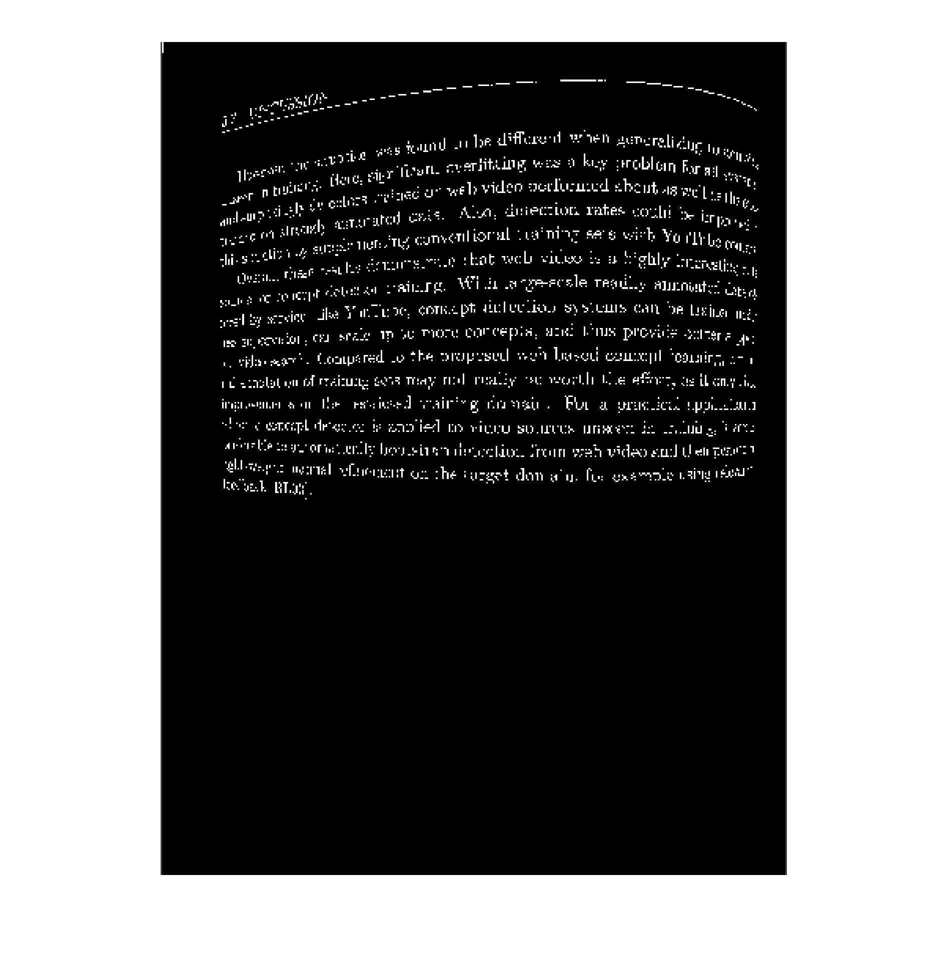
AFTER



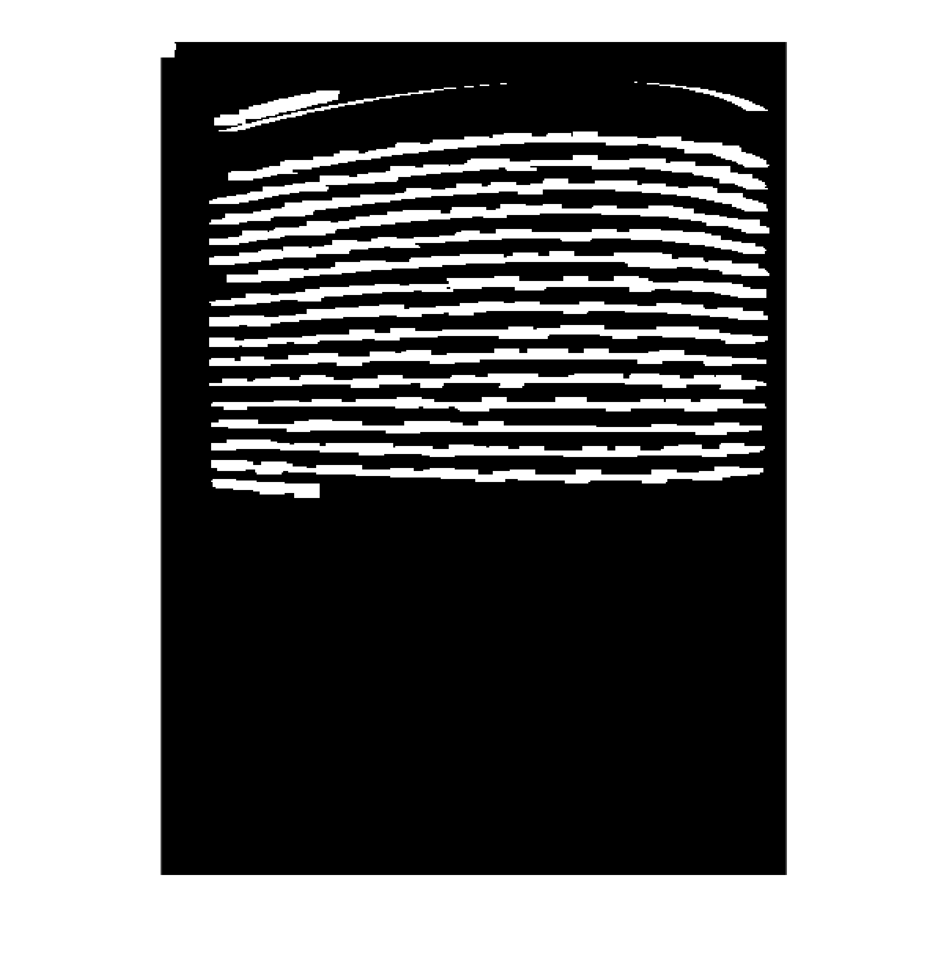
STEP 1. DILATE WITH HORIZ. ELEMENT / ERODE WITH VERT. ELEMENT

9

BEFORE

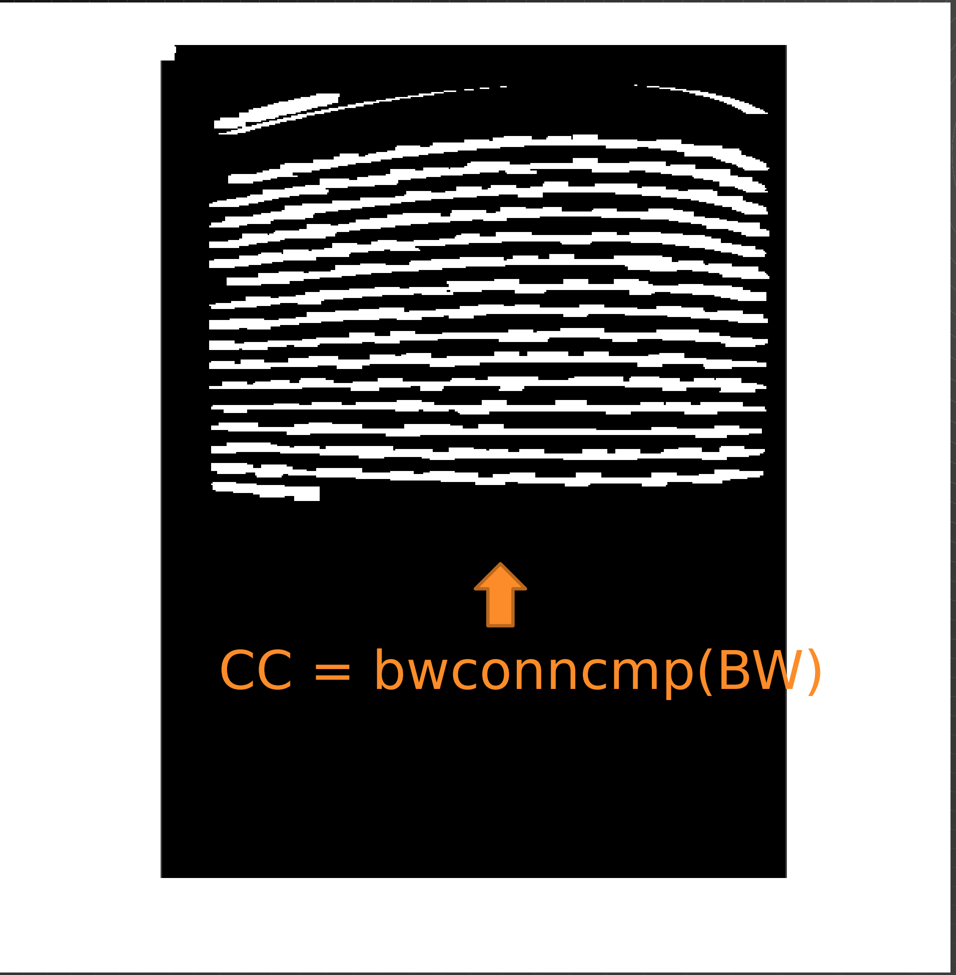


AFTER



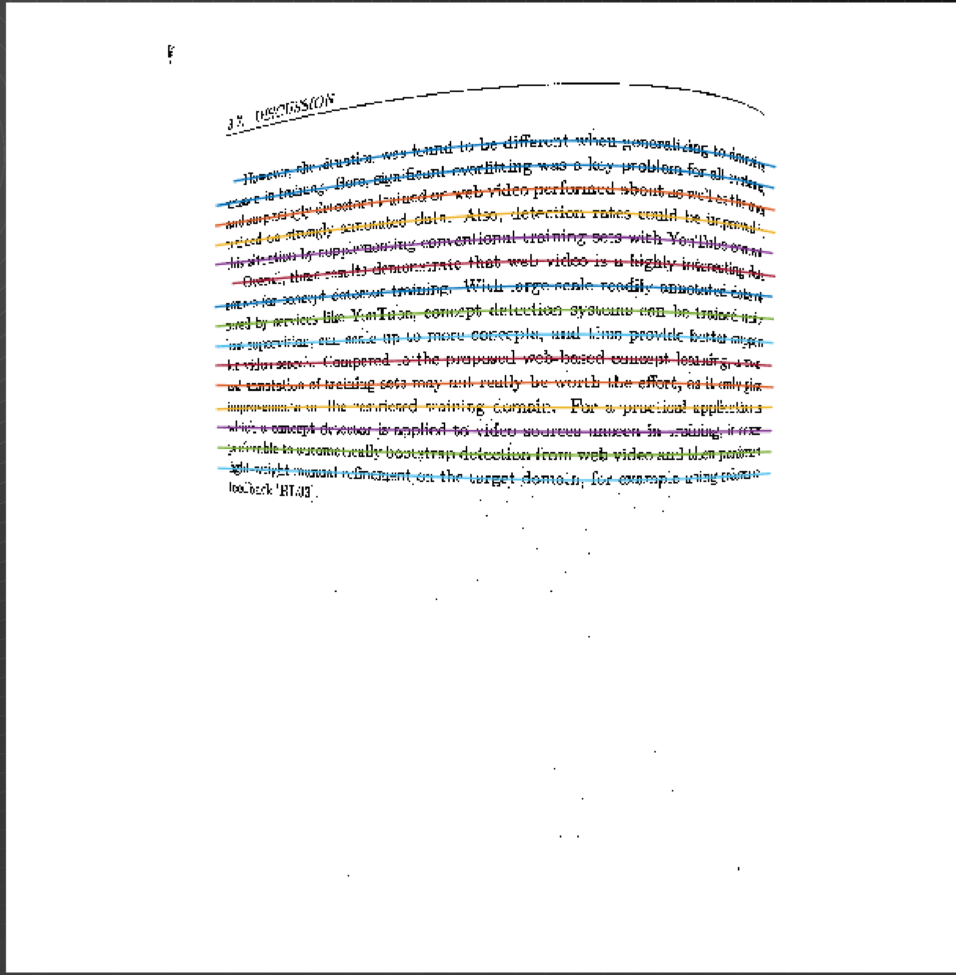
STEP 2. CONNECTED COMPONENTS ANALYSIS; TEXTLINE FITTING WITH CUBIC (3RD ORDER) POLYNOMIAL

BEFORE



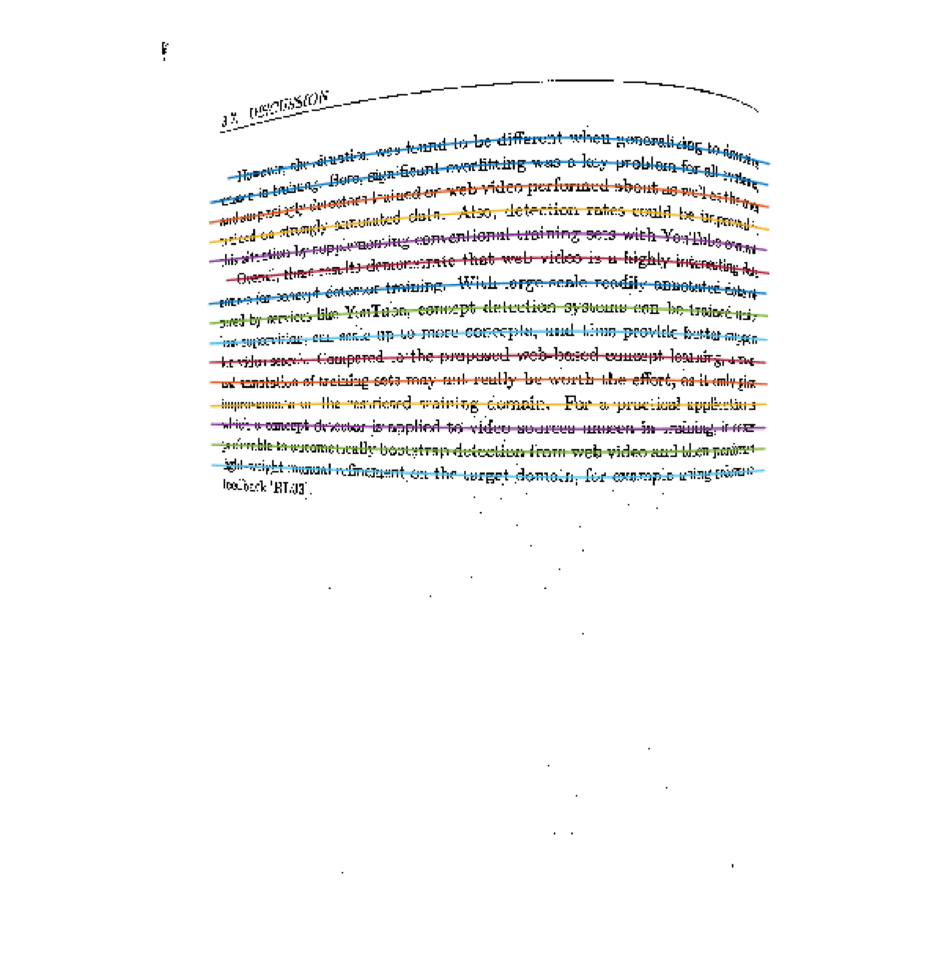
CC = bwconncmp(BW)

AFTER

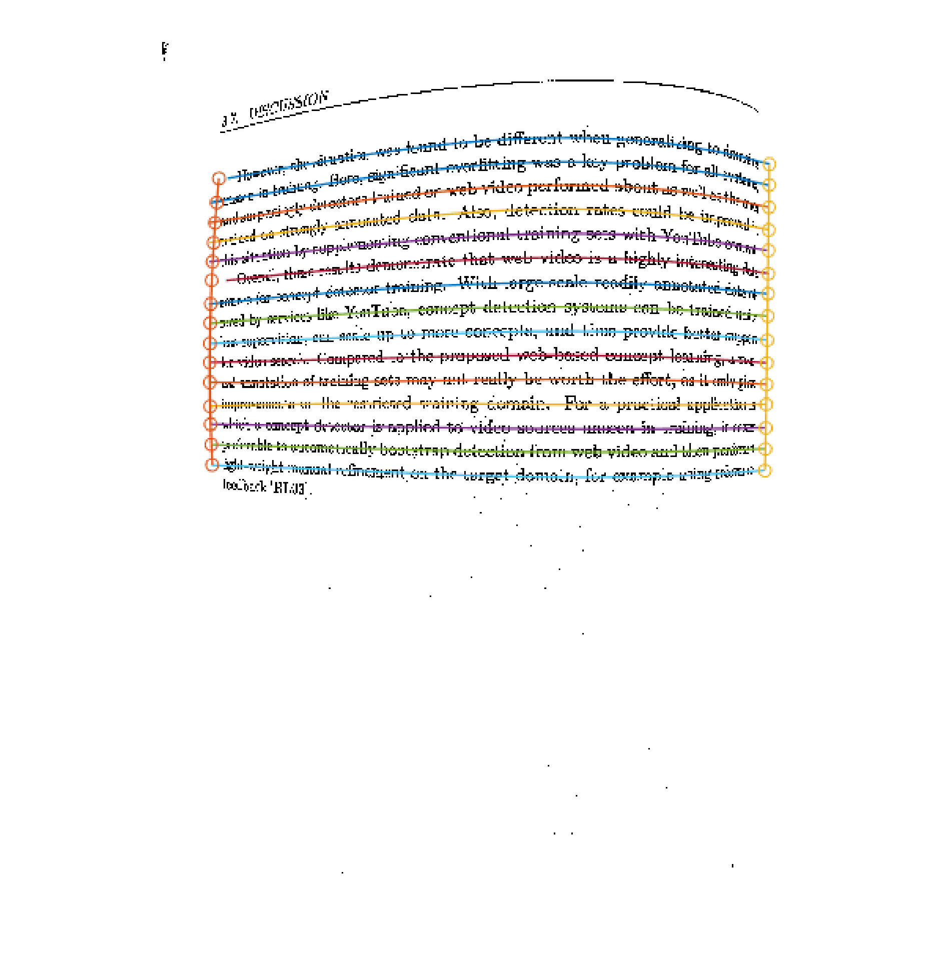


STEP 3. LEFT, RIGHT BORDER FITTING WITH QUADRATIC (2ND ORDER) POLYNOMIAL

BEFORE

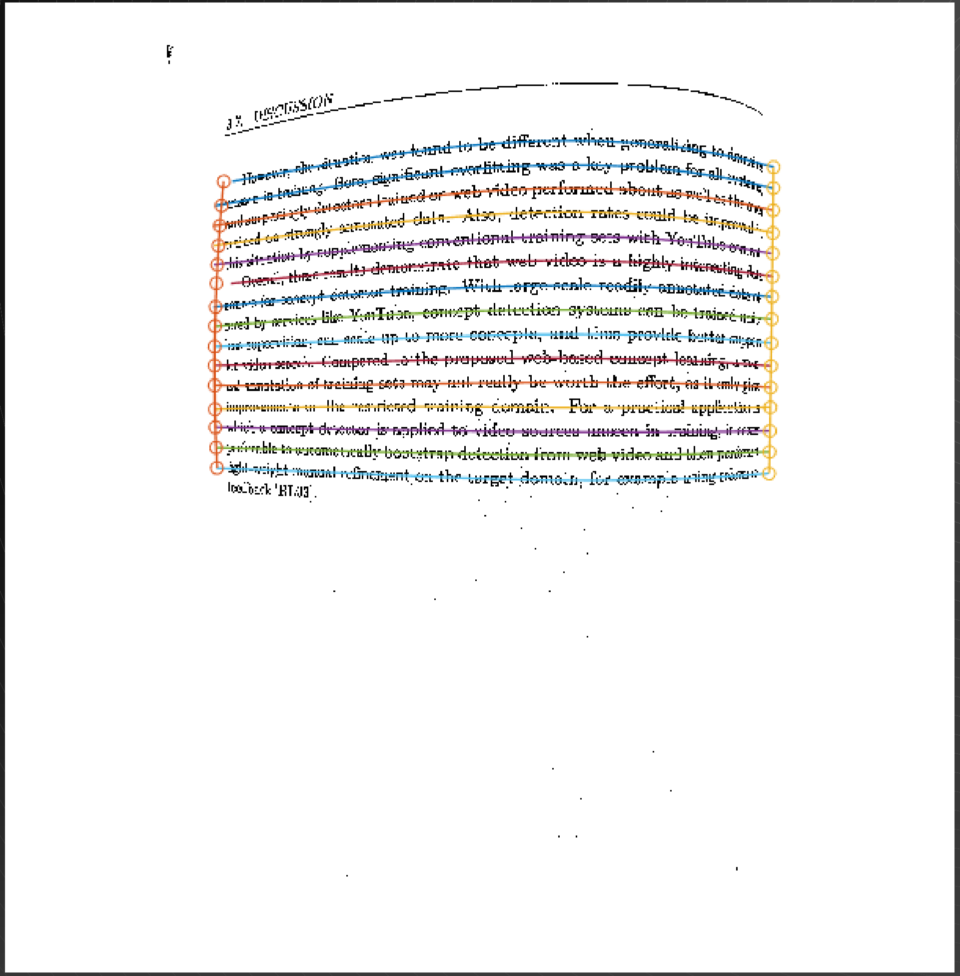


AFTER

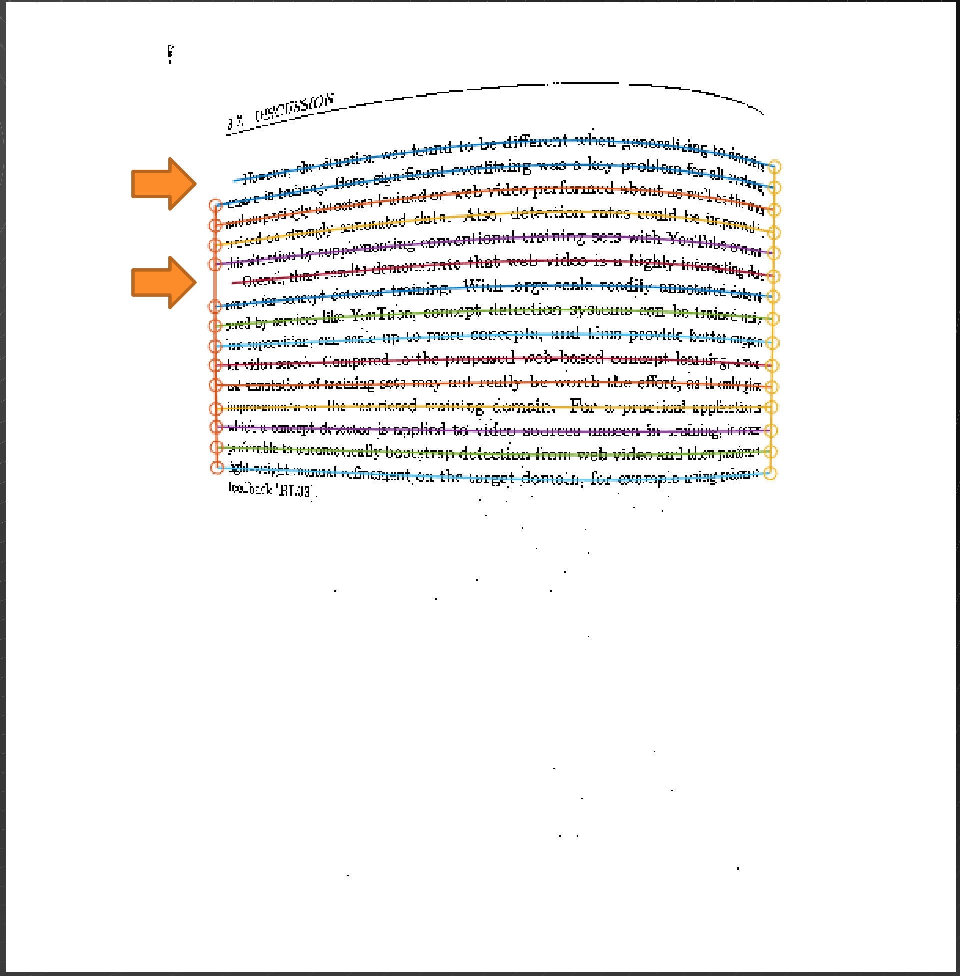


STEP 3A. REFINING LEFT BORDER (REMOVE INDENTATION)

BEFORE



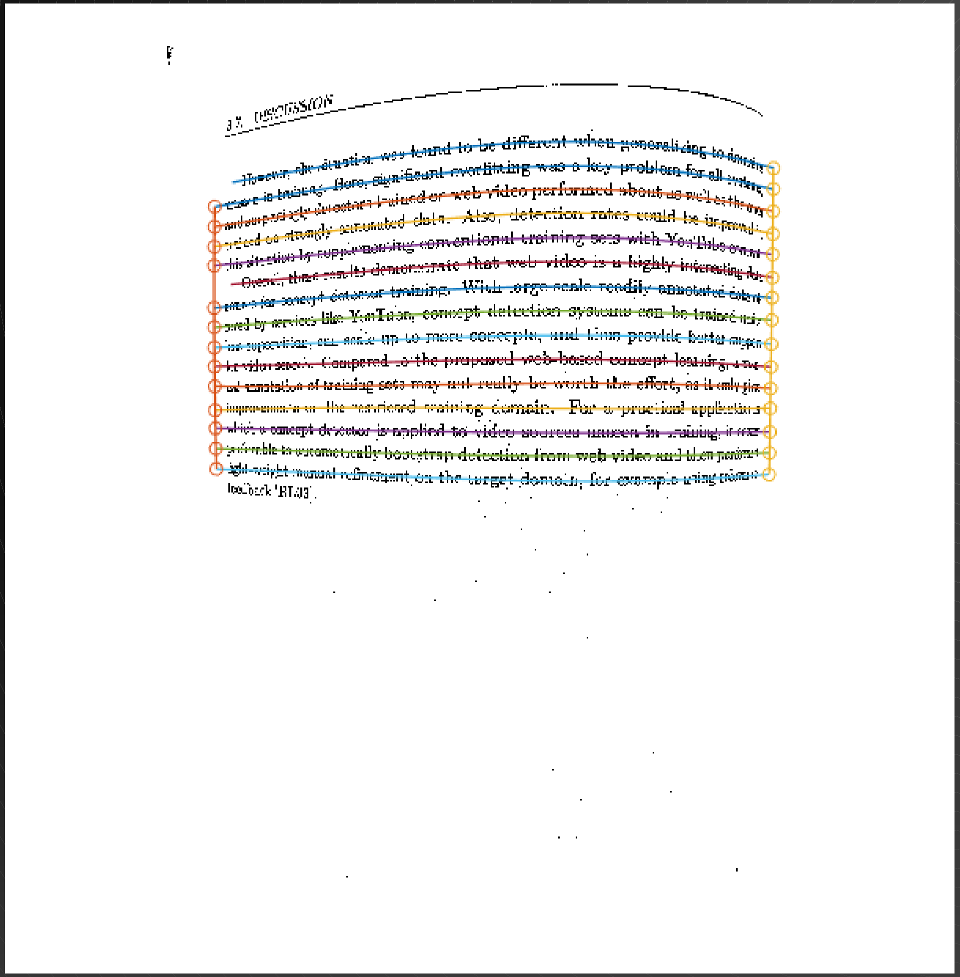
AFTER

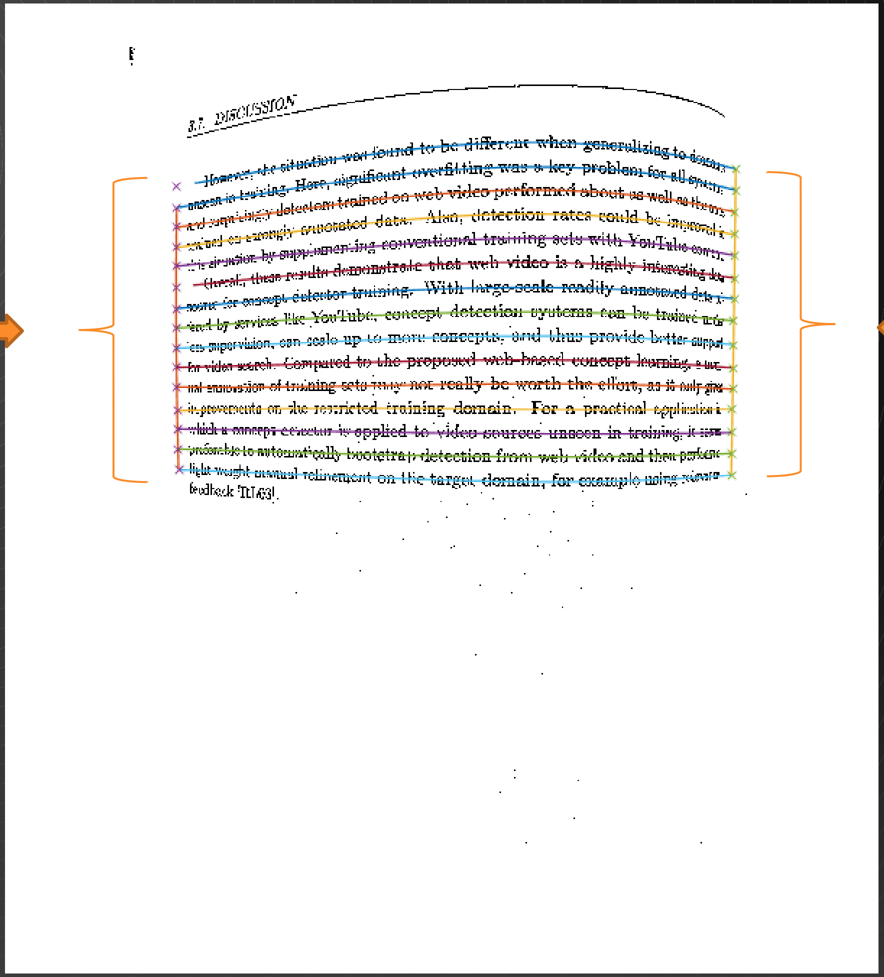


STEP 4. INTERSECTION POINTS BETWEEN LEFT/RIGHT BORDERS AND TEXTLINES

BEFORE

AFTER

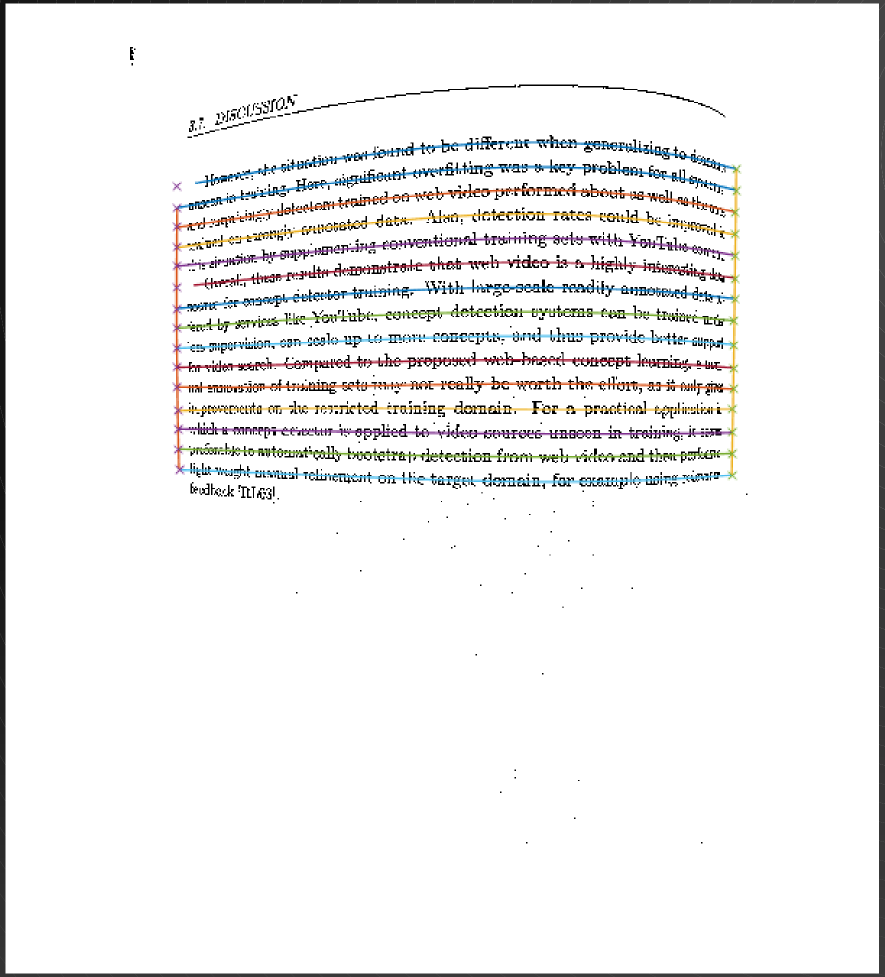


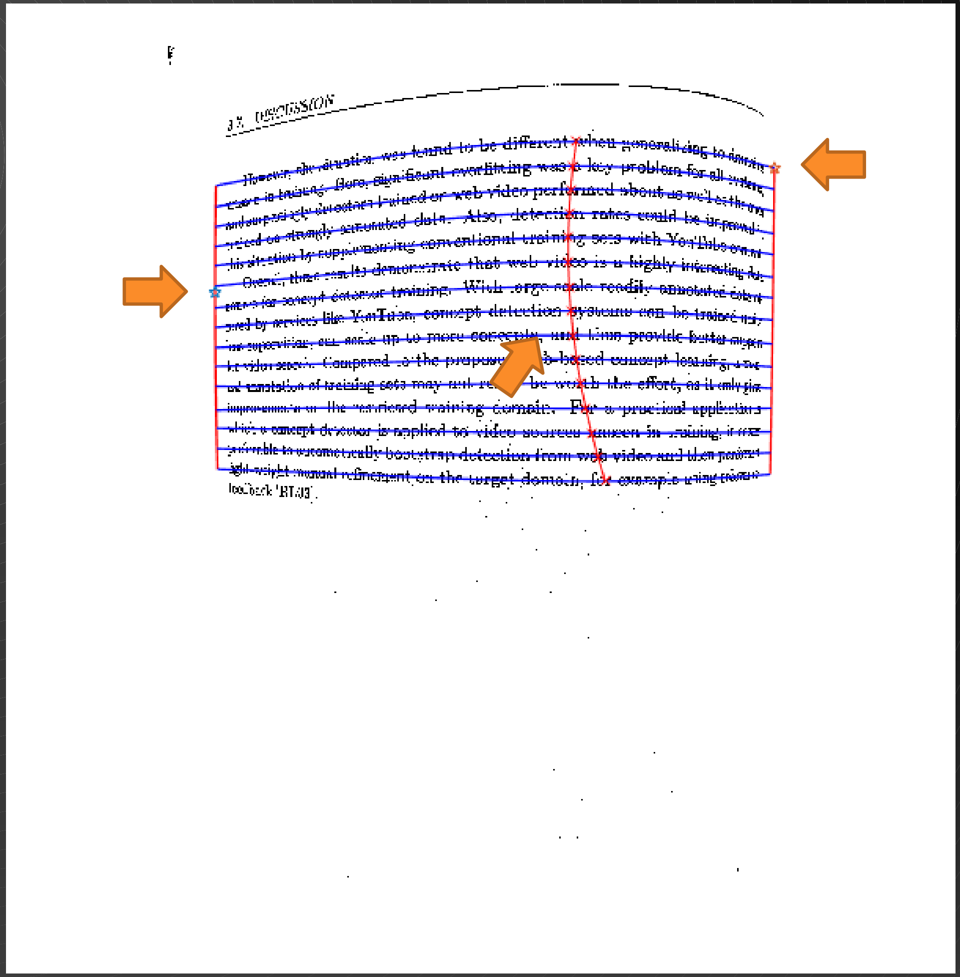


STEP 5. CALCULATE REFERENCE POINTS (LOCAL MIN/MAX OF TEXTLINES, LEFTMOST POINT ON LEFT BORDER, RIGHTMOST POINT ON RIGHT BORDER)

BEFORE

AFTER

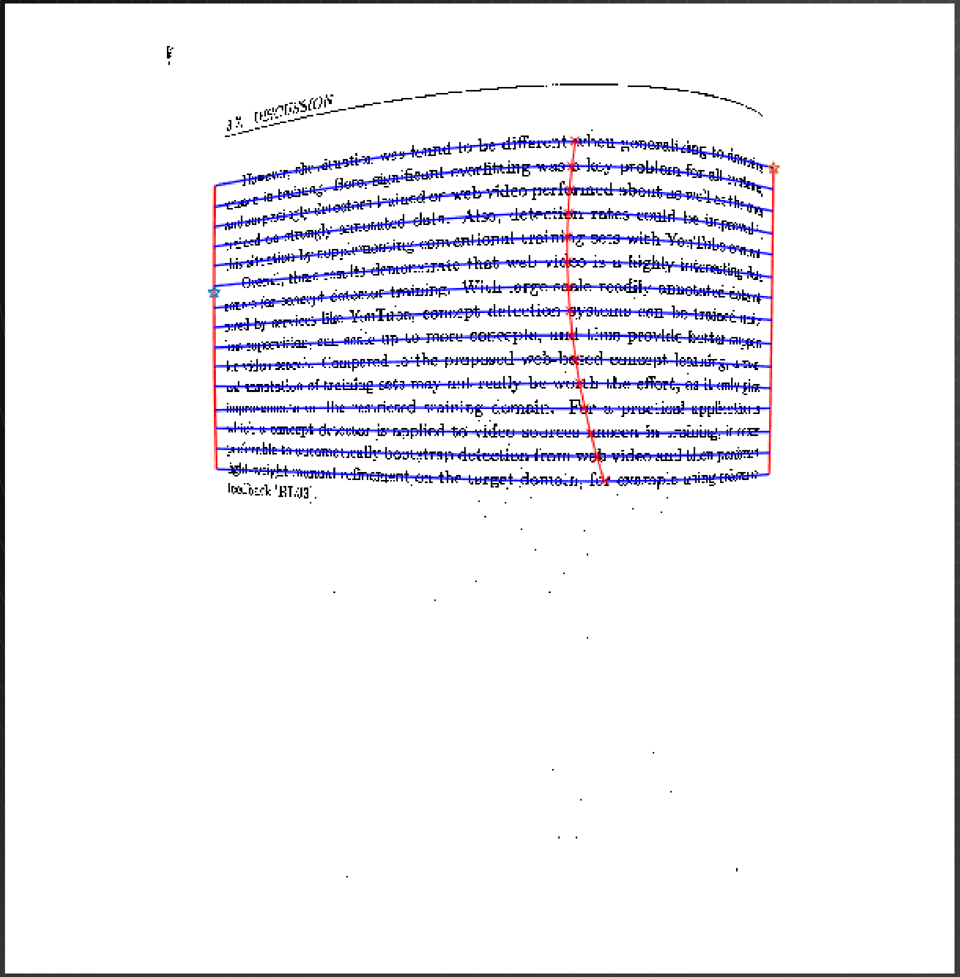




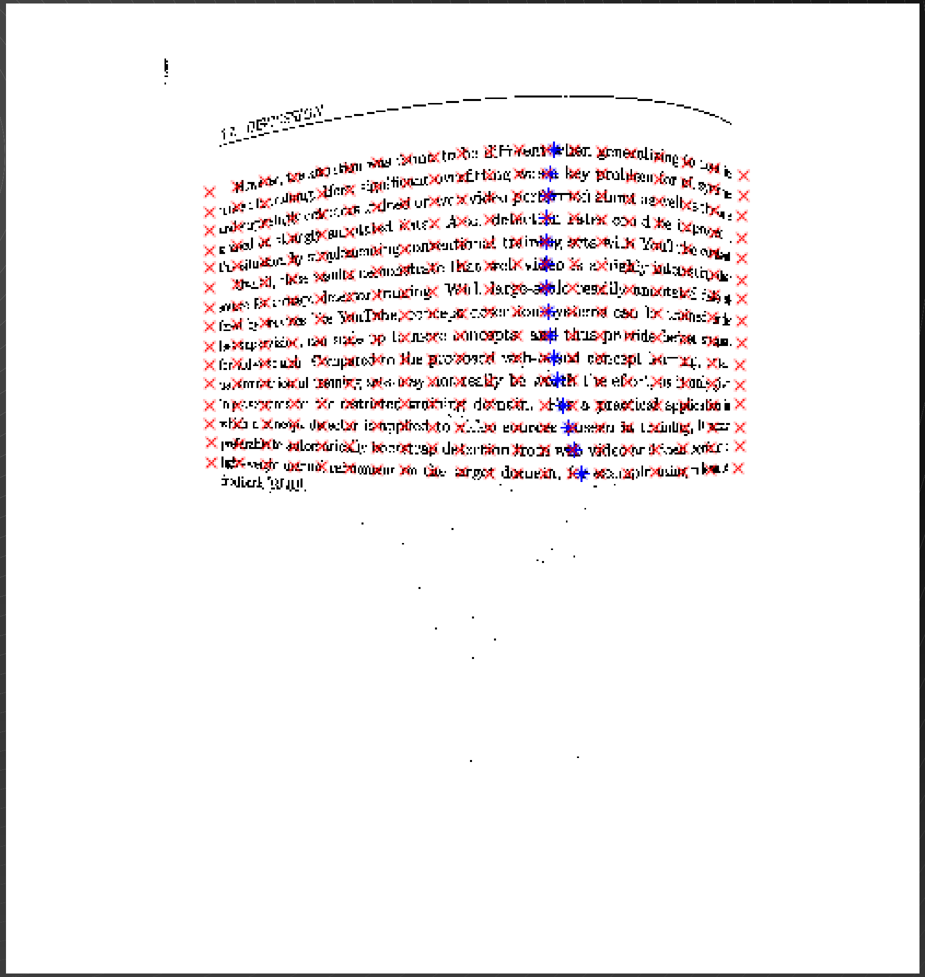
STEP 6. GENERATE MOVING POINTS (INITIAL GRID POINTS) ON TEXTLINES

15

BEFORE

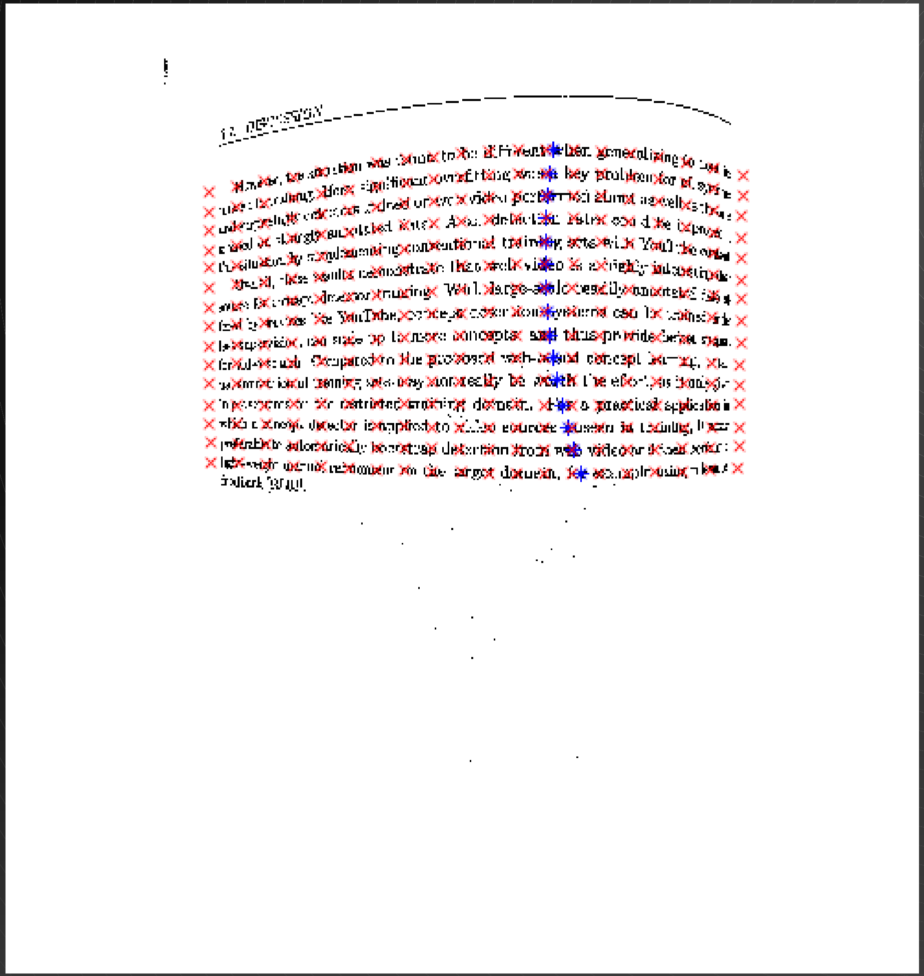


AFTER

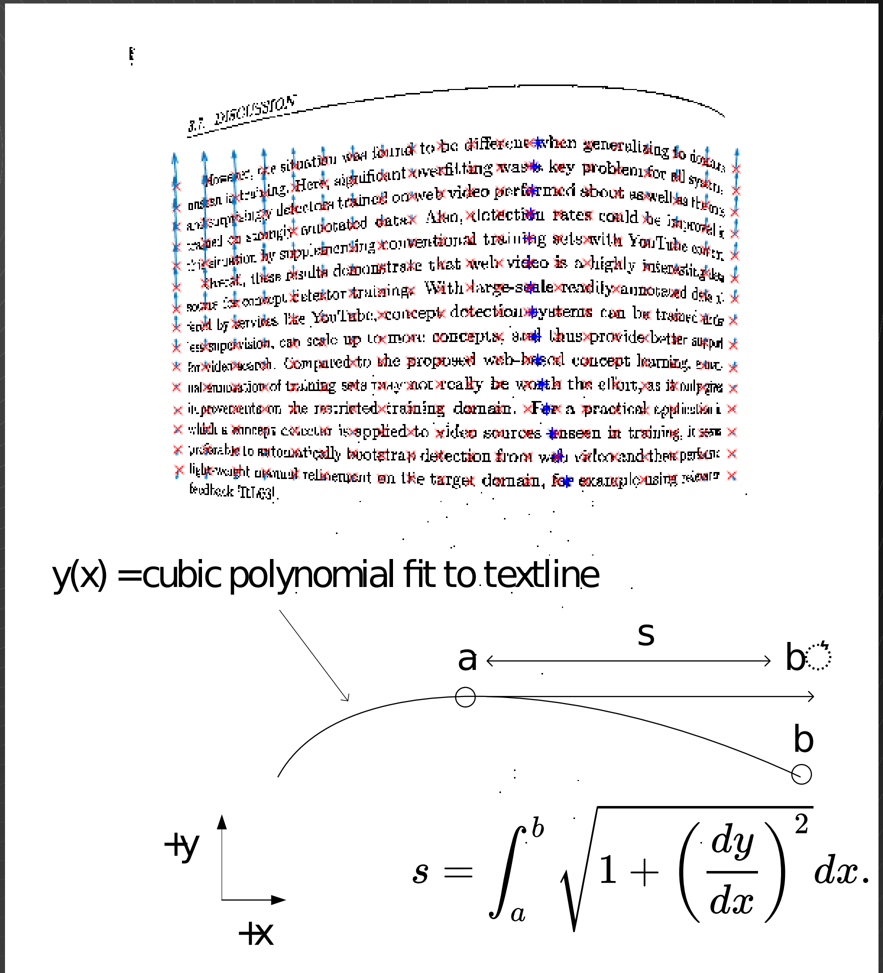


STEP 7. CALCULATE FIXED POINTS (MODIFIED GRID POINTS)

BEFORE



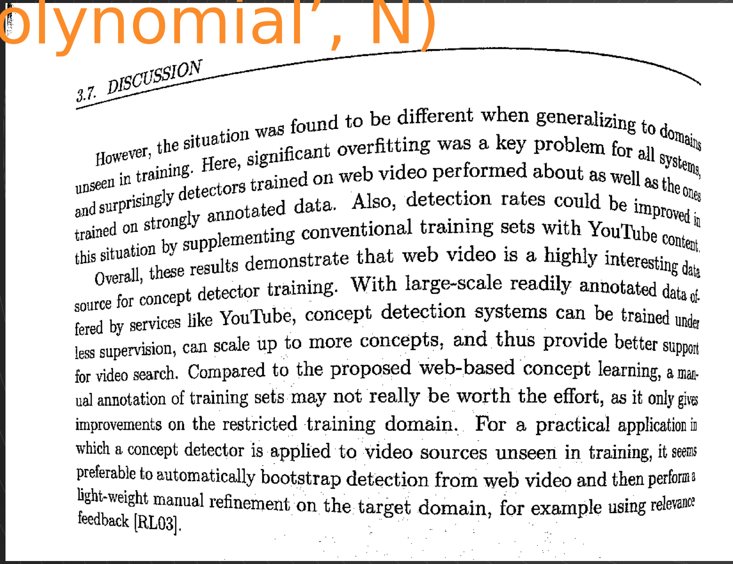
AFTER



RESULTS (1): DEWARPED IMAGES

tform = fitgeotrans(movingPoints, fixedPoints, ‘polynomial’, N)

Original

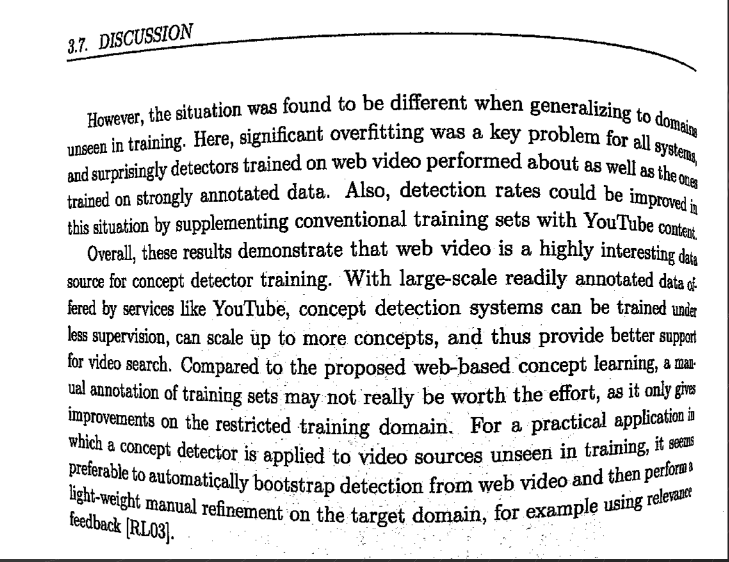


3.7. DISCUSSION

However, the situation was found to be different when generalizing to domains unseen in training. Here, significant overfitting was a key problem for all systems, and surprisingly detectors trained on web video performed about as well as the ones trained on strongly annotated data. Also, detection rates could be improved in this situation by supplementing conventional training sets with YouTube content.

Overall, these results demonstrate that web video is a highly interesting data source for concept detector training. With large-scale readily annotated data offered by services like YouTube, concept detection systems can be trained under less supervision, can scale up to more concepts, and thus provide better support for video search. Compared to the proposed web-based concept learning, a manual annotation of training sets may not really be worth the effort, as it only gives improvements on the restricted training domain. For a practical application in which a concept detector is applied to video sources unseen in training, it seems preferable to automatically bootstrap detection from web video and then perform a light-weight manual refinement on the target domain, for example using relevance feedback [RL03].

N = 2

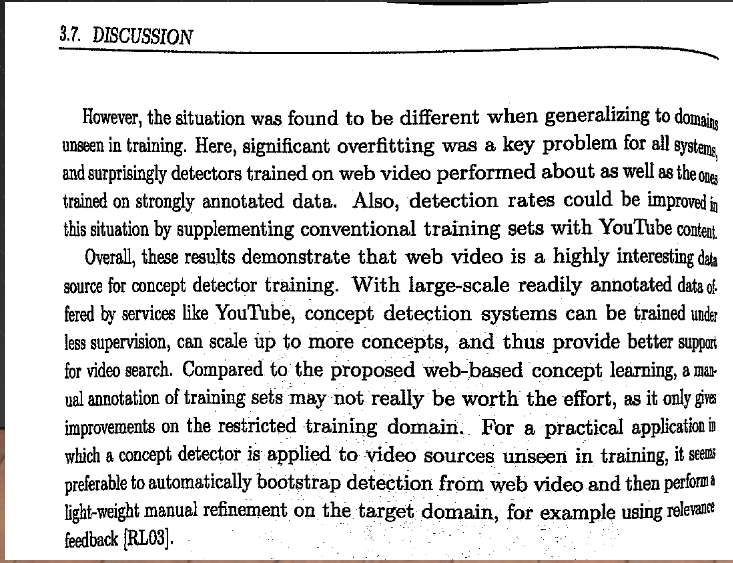


3.7. DISCUSSION

However, the situation was found to be different when generalizing to domains unseen in training. Here, significant overfitting was a key problem for all systems, and surprisingly detectors trained on web video performed about as well as the ones trained on strongly annotated data. Also, detection rates could be improved in this situation by supplementing conventional training sets with YouTube content.

Overall, these results demonstrate that web video is a highly interesting data source for concept detector training. With large-scale readily annotated data offered by services like YouTube, concept detection systems can be trained under less supervision, can scale up to more concepts, and thus provide better support for video search. Compared to the proposed web-based concept learning, a manual annotation of training sets may not really be worth the effort, as it only gives improvements on the restricted training domain. For a practical application in which a concept detector is applied to video sources unseen in training, it seems preferable to automatically bootstrap detection from web video and then perform a light-weight manual refinement on the target domain, for example using relevance feedback [RL03].

N = 3

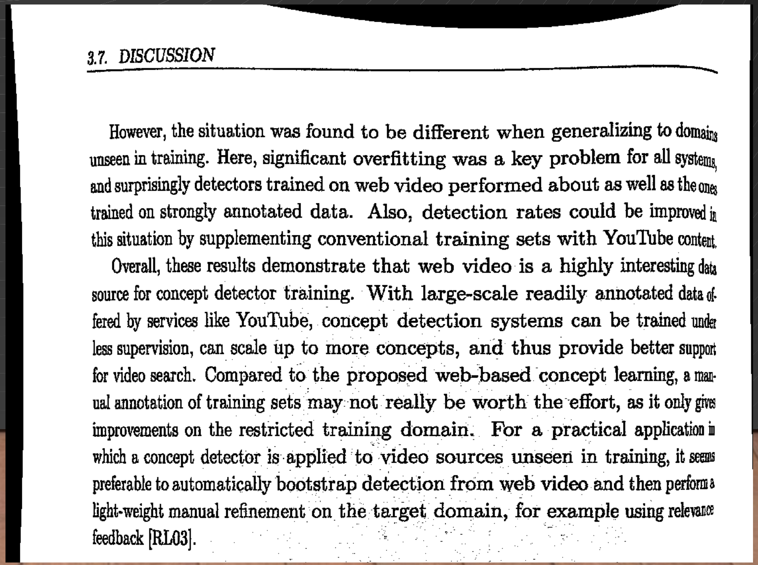


3.7. DISCUSSION

However, the situation was found to be different when generalizing to domains unseen in training. Here, significant overfitting was a key problem for all systems, and surprisingly detectors trained on web video performed about as well as the ones trained on strongly annotated data. Also, detection rates could be improved in this situation by supplementing conventional training sets with YouTube content.

Overall, these results demonstrate that web video is a highly interesting data source for concept detector training. With large-scale readily annotated data offered by services like YouTube, concept detection systems can be trained under less supervision, can scale up to more concepts, and thus provide better support for video search. Compared to the proposed web-based concept learning, a manual annotation of training sets may not really be worth the effort, as it only gives improvements on the restricted training domain. For a practical application in which a concept detector is applied to video sources unseen in training, it seems preferable to automatically bootstrap detection from web video and then perform a light-weight manual refinement on the target domain, for example using relevance feedback [RL03].

N = 4



3.7. DISCUSSION

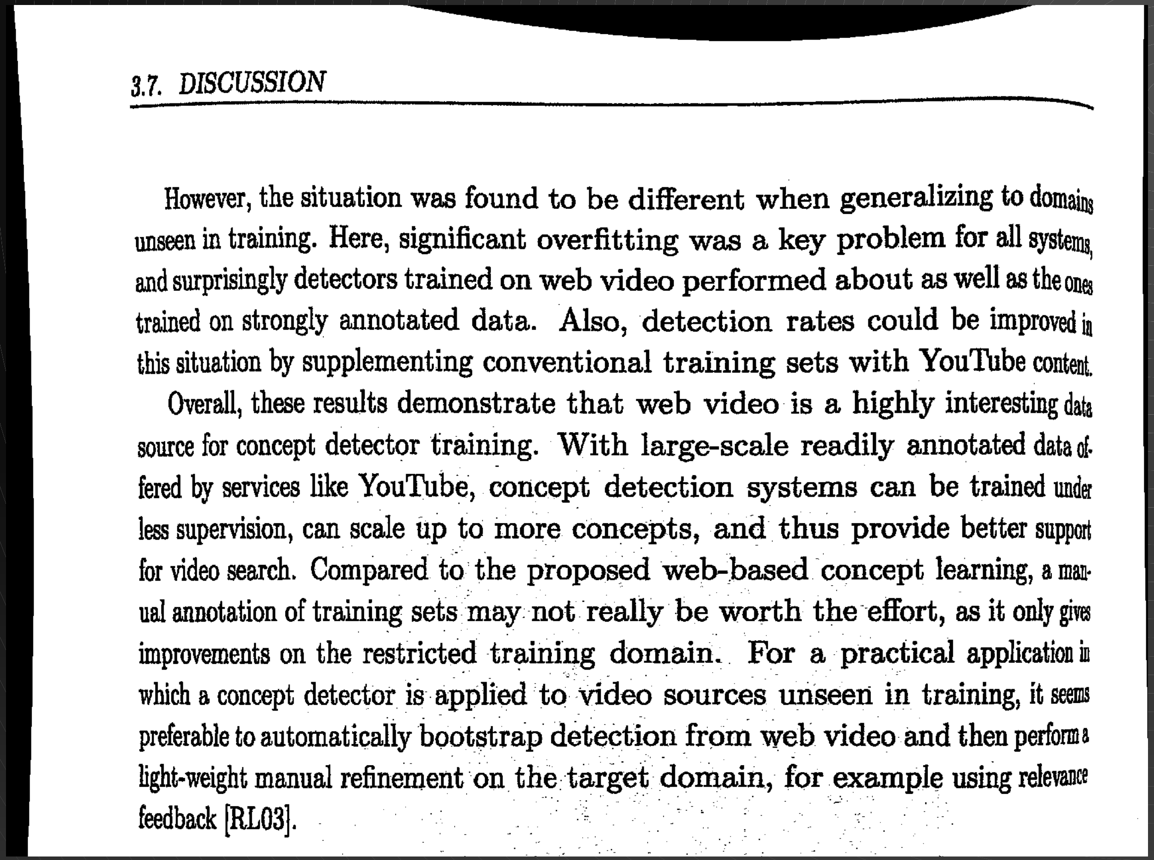
However, the situation was found to be different when generalizing to domains unseen in training. Here, significant overfitting was a key problem for all systems, and surprisingly detectors trained on web video performed about as well as the ones trained on strongly annotated data. Also, detection rates could be improved in this situation by supplementing conventional training sets with YouTube content.

Overall, these results demonstrate that web video is a highly interesting data source for concept detector training. With large-scale readily annotated data offered by services like YouTube, concept detection systems can be trained under less supervision, can scale up to more concepts, and thus provide better support for video search. Compared to the proposed web-based concept learning, a manual annotation of training sets may not really be worth the effort, as it only gives improvements on the restricted training domain. For a practical application in which a concept detector is applied to video sources unseen in training, it seems preferable to automatically bootstrap detection from web video and then perform a light-weight manual refinement on the target domain, for example using relevance feedback [RL03].

RESULTS (2): OCR PROCESSING OF DEWARPED IMAGE

18

POLYNOMIAL TRANSFORM (N = 4)



3.7. DISCUSSION

However, the situation was found to be different when generalizing to domains unseen in training. Here, significant overfitting was a key problem for all systems, and surprisingly detectors trained on web video performed about as well as the ones trained on strongly annotated data. Also, detection rates could be improved in this situation by supplementing conventional training sets with YouTube content.

Overall, these results demonstrate that web video is a highly interesting data source for concept detector training. With large-scale readily annotated data offered by services like YouTube, concept detection systems can be trained under less supervision, can scale up to more concepts, and thus provide better support for video search. Compared to the proposed web-based concept learning, a manual annotation of training sets may not really be worth the effort, as it only gives improvements on the restricted training domain. For a practical application in which a concept detector is applied to video sources unseen in training, it seems preferable to automatically bootstrap detection from web video and then perform a light-weight manual refinement on the target domain, for example using relevance feedback [RL03].

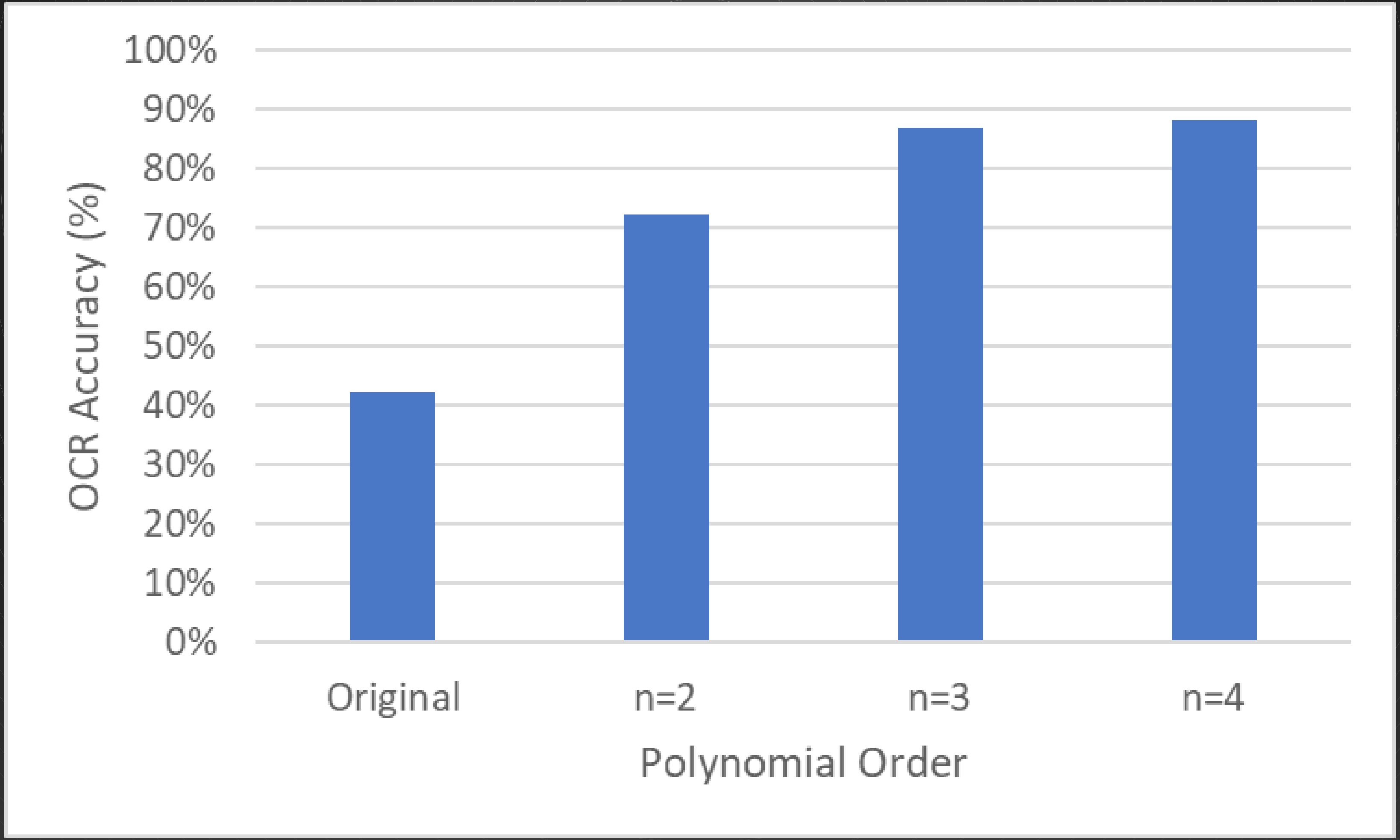
OCR RECOGNIZED TEXT

3.7. DISCUSSION

However, the situation was found to be different when generalizing to domains
unseen in training. Here, significant overfitting was a key problem for all system,
and surprisingly detectors trained on web video performed about as well as the ones
trained on strongly annotated data. Also, detection rates could be improvedin
this situation by supplementing conventional training sets with _You'l‘ube content.

Overall, these results demonstrate that web videos is a highly interesting data
source for concept detector training. eWith large-scale readily annotated data of-
fered by services like YouTube', tconceplit detection systems can be trained under
less supervision, can scale up to more concepts, 5and{ thus provide better support
for video search. Compared the proposed 'vve___b-l__lb_ased.ii”concept learning, a man-
ual annotation of training ysetslfrnay:-notl.'réallyS be lvvorth-' the “effort, as it only gives
improvements on the restricted —trainin.g domain;. For pracfltical applicationin
which a concept detector' is‘-applied eto:i{vfide¢ sources': yl11l_r1.'sLe'ely1 training, it seems
preferable to automatically "bootstrapl detection pfrolm vveb videofand then perform
lightweight manual refinementon,.the;*tvargetf for example using relevance
feedback [RL03l- s r _ g v = =

RESULTS (3): OCR ACCURACY VS POLYNOMIAL ORDER



DISCUSSION

- Run time: ~10 seconds per image
- Top OCR Accuracy: 88%, (n = 4)
- Best reported OCR Accuracy: 95-100% [REF]
- Requires manual adjustments to connected component analysis
- Dewarping method is least effective near left/right borders
- No line/paragraph spacing adjustment

Y. Tian and S. G. Narasimhan, “Rectification and 3D reconstruction of curved document images,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 377–384, 2011.

CONCLUSION

- Most basic dewarping method is implemented from scratch using Matlab, binary image processing, vector calculus
- Better open source algorithms available in C, Matlab, Python, and possibly many others
- Chi Zhang, chi.zhang@ucalgary.ca