DEWARPING CURVED DOCUMENT IMAGES WITH MATLAB, COURSE PROJECT

CPSC 635 - IMAGE ANALYSIS AND COMPUTER VISION - FALL 2017

CHI ZHANG



INTRODUCTION: DIGITIZING DOCUMENTS WITH CELL PHONE CAMERA



Digitizing books



Foreign language translation



Archiving rare documents

DISTORTED IMAGE

3.7. DISCUSSION

However, the situation was found to be different when generalizing to domains the situation was found to be different when generalizing to domains a line of the situation was found to performed about as well as the constant surprisingly detectors trained on web video performed about as well as the constant and surprisingly detectors trained on web video performed about as well as the constant of the situation of the situation by supplementing conventional training sets with You'llube content this situation by supplementing conventional training sets with You'llube content that web video is a highly interesting data.

Overall, these results demonstrate that web video is a nignly interesting data of Overall, these results demonstrate that web video is a nignly interesting data of overall, these results demonstrate that web video search concept detector training. With large-scale readily annotated data of source for concept detection systems can be trained under fered by services like YouTube, concept detection systems can be trained under fered by services like YouTube, concept, 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 performs 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

1t..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 c0l1C€3Pl'{Sa and thus Provide better Support for video search. Compared to the proposed Web-based concept learning, a manual annotation of training sets rnayrnot 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 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

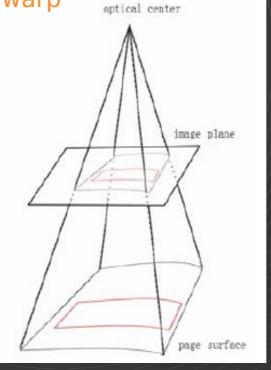
form

Dewarp: reverse process of warp

Intersive environments emphasize spiritualing realistic intersive environments environments emphasize spiritualing realistic intersive environments environments emphasize spiritualing realistic intersive environments e

Stroke analysis

Inmersive environments entire as possible through technique action and perception as closely as possible through technique such as stereo imagery delivered separately to each eye have depth perception, and full six degree of freedom have



The SB-tree of P. E. O'Neil [Acta Inf. 19 1994] And Inference of P. E. O'Neil [Acta Inf. 19 1994] And Inference of P. E. O'Neil [Acta Inf. 19 1994] And Inference of P. Ferragina and R. Grand Constructions where many const

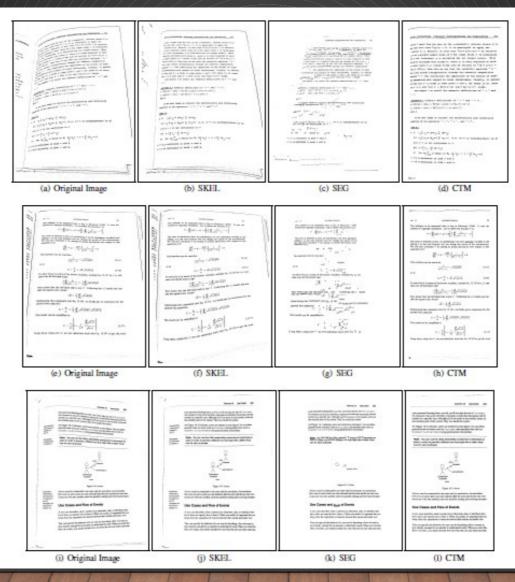
Baseline detection

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.



Best reported results: OCR accuracy 95-100%

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 2rd 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

AFTER

However, the situation was found to be different when generalizing to togate answer in the ring. Here, significant overfitting was as key problem for all synth, and significant ever fitting was as key problem for all synth, and significant could be the significant countries. Also, detection rates could be to the significant composited data. Also, detection rates could be to the significant composited data.

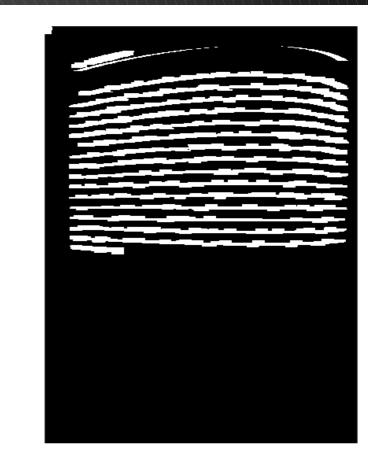
and significated detections trained on the Also, detection rates could be improved and significant could be improved in and carefully equotisted date. Also, detection rates could be improved in an experience of the structure of

The country area is a serious in the different when generalizing to some the area is given a serious flow, significant evenithing was a key problem for all even agents the flow, significant restricting was a key problem for all even agents in the country with the country of t

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

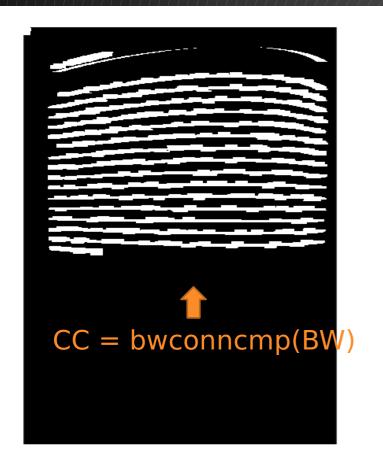
9 BEFORE AFTER

The control of which was found to be different when generalizing to some about a was found to be different when generalizing to some agent modern (dere, significant repetition) was a key problem for all week and applied to the control operator of strong and applied to the control operator of strong agranted costs. Also, discertion rates with You'll be considered and applied to the control training sets with You'll be considered to the strong department of the control operator of the control operator of the You'll be, concept distriction systems can be training as a greater, the You'll be more concept, and thus provide outers Act with a set of some of training department of the proposed were based comment features to the proposed with based comment features and a control feature gets may not really not worth the effect, as it contained to the proposed with large of the proposed application of a control decay to the control decay of the control of the co



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

10 BEFORE AFTER



Hosein the density was tend to be different when generalizing However, the detection of the problem for all the control of the problem for the p care is tracked fore sured or web video performed about so well entering and support of well-entering tracked data. Also, detection rates could be accompanied data. Also, detection rates could be and ampair the property accounted clubs. Also, detection rates could be improved. his straight appropriate conventional training sets with You labor on the the strains by suppler water that web victor is a highly interesting to over the same tribing. With organic readily abboute court mel by services like YenTube, convept detection systems can be trained inno supervision, can see up to more correcpts, and time provide bester store havilar section. Competed to the proposed web-based concept learning as -ul contains of treesing cots may and really be worth the effort, as it only sizimprovements on the restricted statisting domain. For a practical applications which a concept desector is applied to video sources anneed in training it me - swimble to wrome treatly bootstrap detection from web video and then points. igh wight mount refinement on the target domests, for example using mount

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

$oldsymbol{11}$ before

AFTER

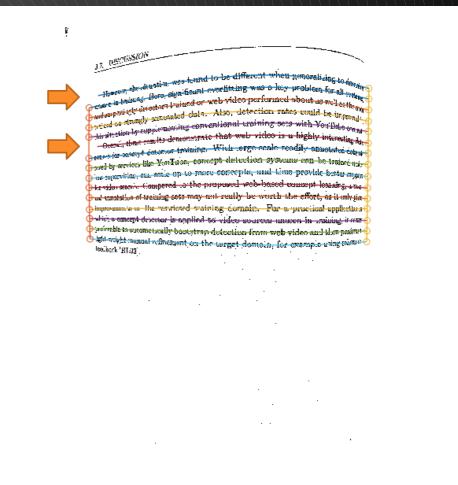
However, the details were tend to be different when generalizing to do not the control of the different eventiating was a key problem for all other case in this sea being problem for all other case in this sea the problem. Also, detection rates could be impossible or an adapted characteristic control training sets with Younger country that web video is a highly intending to open, there can be denoughed that web video is a highly intending to open, there can be training. With large could readily abnorated the many of the property control training. With large could readily abnorated that many intending to make up to more concept altertion systems can be trained as a specifical can are up to more concepts, and have provide bests may alter into a concept can be an account of the proposed web based concept learning as a the control control of the control control of the control control of the control control of the control of the proposed control. For a province and alter power which is proposed to be control of the control of the

The control of the co

STEP 3A. REFINING LEFT BORDER (REMOVE INDENTATION)

12 BEFORE AFTER

O Howard he district was toned to be different when generalizing to in-House, the department of the property of the problem for all many the control of the problem for all many the problem for the probl Contemporary Land of web video performed about so well the bridge performed about so w Control of structured duty. Also, deterritor rates could be in production of the internal could be in production. the straight supplementation of the supplementation of the straight straight supplementation of the supplementatio the stratum of support Decre the second energy training. With organizate readily abbother course Director Systems and be trained comept detection systems and be trained inin the state of the property of the state of he vilus see. Competed to the proposed web-based concept learning and but contains of trecining soon may and really be worth the effort, as it only size improvements on the restricted statisting domain. For a practical application (Фэлгиын to wnometically bootstrap detection from web video and the penture (dight-wight musual refinement on the target domein, for example using reason



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

13 BEFORE

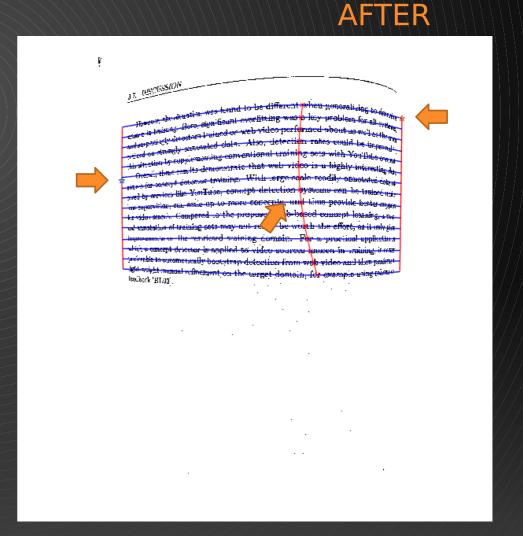
AFTER

However, the detection was bound to be different when generalizing to dominate the problem for all many contents of the problem of the problem for all many contents of the problem of the problem

Houses the situation was found to be different when generalizing to soon to be sufficient when the performed about us well entering and in training the entering on the performed about us well entering and an intering of the performed about us well entering and an intering to consiste data. Also, election rates could be imported and on a some intering to consiste the performed training sets with YouTuke course the interior by suppressing the some for consept election training. With large scale readily annotated data could be severed by severe like YouTuke, concept detection systems can be trained at a comparation, can seek up to more concepts, and thus provide better suppler in the search Compared to the proposed who based concept learning the material and training sets when really be worth the effort, as it only give in provening of training sets was really be consistent in the residual application; which a suppresent on the confident configuration from web video and then provide light weight manual relinement on the target dornain, for example using course feelback Thesi

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

Discussion was found to be different when generalizing to the service the situation was found to be different when generalizing to the service performed about us well as the manager is truthed. Here, significant everifiting was a key problem for all grants and service the service of the ser



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

15 BEFORE AFTER

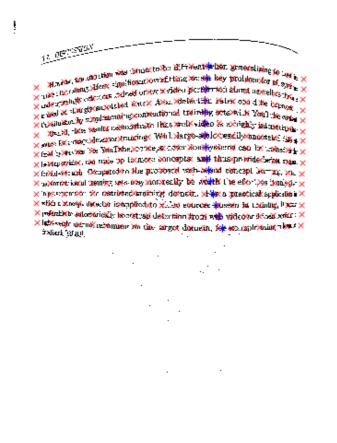
of contains of training cots may and really be worth the effort, as it only sign improvements on the restricted training domain. For a practical applications which concept desertes is applied to video sources between in training it me-

Analysis of the second second

STEP 7. CALCULATE FIXED POINTS (MODIFIED GRID POINTS)

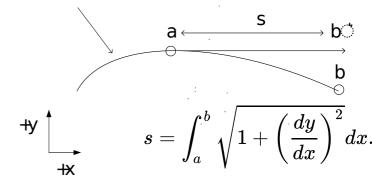
16 before

AFTER



places of the structure was found to be different when generalizing to industry the structure was found to be different when generalizing to industry the structure of the struc

y(x) =cubic polynomial fit to textline



RESULTS (1): DEWARPED IMAGES

tform = fitgeotrans(movingPoints, fixedPoints,

(polynomial', N)

Original

However, the situation was found to be different when generalizing to domain the situation was found to be different when generalizing to domain the situation was found to be different when generalizing to domain the situation was found to be different when generalizing to domain the situation was found to be different when generalizing to domain the situation was found to be different when generalizing to domain the situation was found to be different when generalizing to domain the situation was found to be different when generalizing to domain the situation was found to be different when generalizing to domain the situation was found to be different when generalizing to domain the situation was found to be different when generalizing to domain the situation was a key problem for all the situation was a situation to be domain the situation that the situation was a situation to be domain However, the situation was long to domeits unseen in training. Here, significant overfitting was a key problem for all system unseen in training. Here, significant overfitting was a key problem for all system. unseen in training. Here, significant on web video performed about as well as the one and surprisingly detectors trained on web video performed about as well as the one and surprisingly detectors trained on web video performed about as well as the one and surprisingly detectors. and surprisingly detections at the one and surprisingly annotated data. Also, detection rates could be improved in trained on strongly annotated data. trained on strongly amount of improved in this situation by supplementing conventional training sets with YouTube content is situation by Supplementation of Supplementation

Overall, these source for concept detector training. With large-scale readily annotated data of source for concept detection systems can be trained under fered by services like YouTube, concept detection systems can be trained under less supervision, can scale up to more concepts, and thus provide better silnness for video search. Compared to the proposed web-based concept learning a mannal annotation of training sets may not really be worth the effort, as it only enter 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 performs light-weight manual refinement on the target domain, for example using relevance feedback [RL03]

3.7. DISCUSSION

However, the situation was found to be different when generalizing to domain 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 one 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 of fered 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 give improvements on the restricted training domain. For a practical application which a concept detector is applied to video sources unseen in training, it seems preferable to automatically bootstrap detection from web video and then performs light-weight manual refinement on the target domain, for example using relevance feedback [RL03].

3.7. DISCUSSION

However, the situation was found to be different when generalizing to domain However, the statement of the However, unseen in training and surprisingly detectors trained on web video performed about as well as theorem 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 date source for concept detector training. With large-scale readily annotated data of fered 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 is 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 relevant feedback [RL03].

3.7. DISCUSSION

However, the situation was found to be different when generalizing to domain 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 of fered 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 give 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 performs light-weight manual refinement on the target domain, for example using relevance feedback [RL03].

N = 2

 $N \neq 4$

N = 3

RESULTS (2): OCR PROCESSING OF DEWARPED IMAGE

18 POLYNOMIAL TRANSFORM (N = 4)

OCR RECOGNIZED TEXT

3.7. DISCUSSION

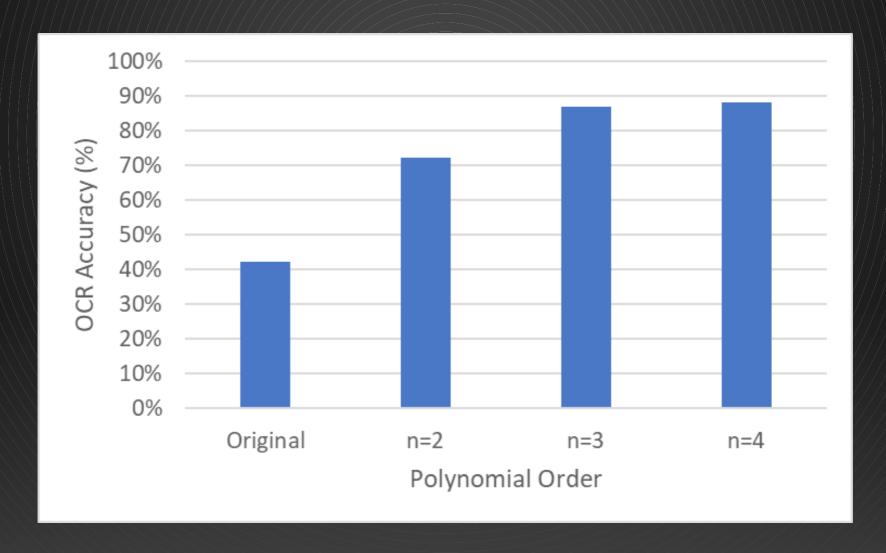
However, the situation was found to be different when generalizing to domains unseen in training. Here, significant overfitting was a key problem for all systems, and surprisingly detectors trained on web video performed about as well as the ones trained on strongly annotated data. Also, detection rates 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 of fered 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 performs light-weight manual refinement on the target domain, for example using relevance feedback [RLO3].

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 improved in this situation by supplementing conventional training sets with _You'I'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 offered 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 manual annotation of training ysetsIfrnay:-notl.'réallyS be lvvorth-' the "effort, as it only gives improvements on the restricted —trainin.g domain;. For pracfitical 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 =



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

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