Automatic classification of defect page content in scanned document collections

Reinhold Huber-Mörk, Alexander Schindler
Intelligent Vision Systems, Safety & Security Department, AIT Austrian Institute of Technology GmbH
Vienna, Austria
Email: reinhold.huber-moerk@ait.ac.at

***Abstract*—We describe a method for defect detection and classification for collections of digital images of historical book documents. Undistorted text images from various books characterized by strong variation of language, font and layout properties are discriminated from typical errors in digitization processes such as occlusion by an operator's hand, visible book edge or image warping artifacts. A bag of local features approach is compared to a global characterization of location, size and orientation properties of detected keypoints. Machine learning is used to discriminate between those classes. Results for different features are compared for the task of discrimination between undistorted text and the major distortion class which is presence of the operator's hand, where features based on the bag of local features derived histograms achieved a cross-validation accuracy better than 99 percent on a representative data set. Taking into account up to three classes of distortions still resulted in cross-validation accuracies beyond 90 percent using bag of local features derived visual histograms for classifier input.**

I. Introduction

Large scale automated and semi-automated scanning projects of books [1] and newspapers [2] face the issue of quality assurance. In this respect, the detection of duplicated pages, missing pages or page scans of limited quality are main tasks in quality assurance.

Image based approaches can be used for detection and classification of visual content and are commonly based on local image feature descriptors. One of the most prominent local keypoint detection and description methods is the Scale Invariant Feature Transform (SIFT) [3]. SIFT operates on a scale space representation for feature detection from local gradient distributions. The SIFT descriptor is an invariant representation of local image content which is further used in image matching, recognition or comparison frameworks. Visual ranking for large-scale image search based on SIFT descriptor comparison among images was described by Jing and Baluja [4]. The bag of features (BoF) [5] derived from local descriptors such as SIFT was described as an efficient approach to content based retrieval and detection from image data. The BoF approach is inspired by the bag of words approach based on term frequency weighting and comparison in text retrieval [6]. The expressive power of local descriptors was demonstrated by Weinzaepfel et al. [7] where it was shown that it is possible to reconstruct an image from the information contained in its local descriptors.

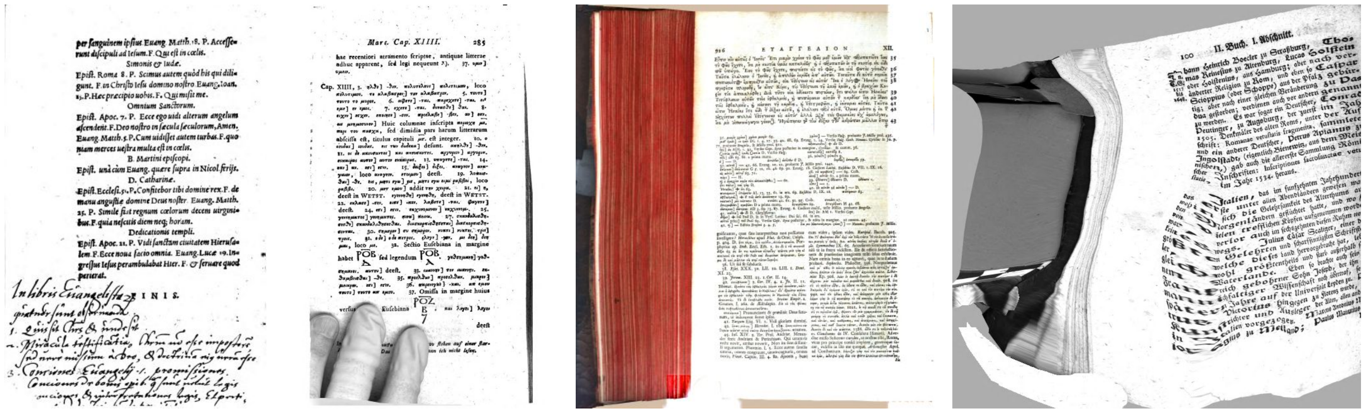
Our previous work concentrated on the quality assurance issues for images of Chinese handwritten documents from the International Dunhuang Project (IDP) archived at the British Library. Automatic detection of image quality was based on SIFT matching and perceptual difference estimation [8]. Detection of duplicated pages in an automatic document scanning workflow was investigated for documents from the late 19th century archived at the Austrian National Library. A BoF approach based on SIFT descriptors was chosen and results were compared to manually ground truth obtained data [9]. This paper demonstrates the classification of scanned content into classes related to possible defects occurring in automatic or semi-automatic book scan procedures.

In more detail, we address the problem of discriminating images of text pages without any corruption, see Fig. 1(a) for an example, and various classes of page scan corruption. One such class is the one where the hand of the scan machine operator is visible in an image, see Fig. 1(b) for an example. The task of hand and finger recognition was exhaustively addressed in the field of gesture recognition by different approaches, e.g. model-based approaches [10] or feature based approaches [11]. Our approach is a feature-based approach, where the presence of hand or finger defines on class of distortion. We also investigated two additional classes of distortion. The additional classes are the book edge class, i.e. in the un-hinged outer side of the book is visible in the image, see Fig. 1(c) for an example, and image warping artifacts, see Fig 1(d) for an example. The latter class of artifacts occurs due to erroneous postprocessing of the book, typically caused by page segmentation errors.

The remainder of this paper is organized as follows. Section II reviews related work in document image processing and page classification applied to historical documents. Our approach is described in Section III. Results are presented in Section IV and Setion V summarizes the paper.

II. Related work

The problem of duplication in large document archiving and information extraction systems was discussed by Doermann et. al where a method for duplicate detection in scanned documents based on shape descriptions for single characters showed advantages with respect to robustness and speed when compared to OCR [12]. Beusekom et. al [13] address the analysis of different versions of scanned historical documents and the difference is highlighted and not further quantified. Several authors mention that the use of optical character recognition, which is an obvious approach to extract relevant information from text documents, is quite limited with respect to accuracy and flexibility [14], [12], [15]. An approach combining page segmentation and Optical Character Recognition (OCR) for newspaper digitization, indexing and



(a) Normal text page

(b) Scan with finger

(c) Book edge in scan

(d) Warping defect

Fig. 1. Sampled of scanned book pages without and with different types of defects

search was described recently [2], where a moderate overall OCR accuracy on the order of magnitude of 80 percent was reported. Page segmentation is a prerequisite for the document image retrieval approach suggested in [14] where document matching is based on the earth mover's distance measured between layout blocks.

Baluja and Covell [16] describe an approach to differentiate between text and image content, especially line drawings, in scanned document pages. SIFT features are combined with AdaBoost learning to obtain relevant or descriptive information, e.g. preview pages, in automatic large-scale book-scanning systems. Discrimination of main body text and layout elements having a decorative meaning in historical manuscripts using SIFT and a support vector machine (SVM) classifier is described by Garz et. al [17]. Detection of specific graphical elements in digitized documents such as logos was discussed by Zhu and Doerman [18] as well as by Li et al. [19].

III. Image feature extraction and page classification

We used local features derived from interest regions which are quantized by the BoF approach as well as global features derived from location, size and orientation properties of all interest points in an image. To detect and describe interest regions in document images we used the SIFT keypoint extraction and description approach. Subpixel image location, scale and orientation are associated with each SIFT keypoint. The associated SIFT descriptor consists of a 4×4 location grid containing 8 gradient orientation bins in each grid cell. The descriptor vectors of length 128 will be used to learn a visual dictionary, i.e. the BoF.

A. *Visual histogram features*

Learning of the visual dictionary is performed using a clustering method applied to all SIFT descriptors of all images, which could become computationally very demanding. As a single scanned book page already contains a large number of local descriptors, we applied preclustering of descriptors to each image. In contrast to a similar procedure, where all descriptors for all images of the same category are clustered independently and subsequently appended to the BoF [20], we construct a list of clustered descriptors for each page and cluster this list in a second step in order to obtain a dictionary for the whole book. We used k-means for preclustering and final clustering of the BoF. Similar approaches include approximate and hierarchical k-means schemes [21].

Individual visual words, also called terms i occur on each page with varying frequency t_i. The histogram of visual word frequencies t_i for an individual book page is derived from the BoF representation by counting the indices of the closest descriptors. The term frequencies t_i are represented in its normalized form, i.e. $\sum_{i=1...|V|} t_i = 1$, where V is the set of visual words contained in the visual vocabulary for an individual book. The visual histogram corresponding to a scanned book pages is used as a input feature vector for the classifier.

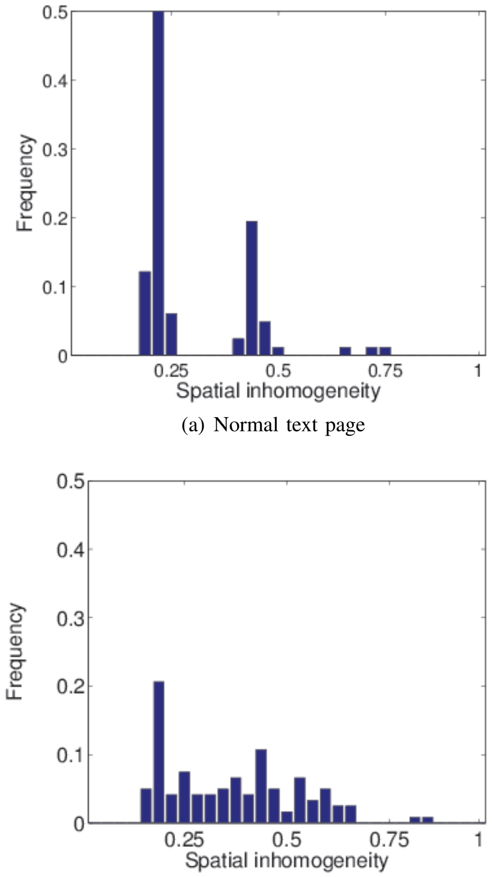
Note, that the histogram of visual term frequencies for an individual page does not take into account absolute location of detected features or their relative placement. Therefore, we introduced global features expressing spatial keypoint properties.

B. *Global keypoint property statistics*

Each SIFT keypoint is characterized by a sub-pixel and sub-scale location derived from its scale space extremum and an orientation derived from the gradient distribution in the vicinity of each keypoint. In the following we describe how to use these properties for classification of image content.

Relative spatial information could be introduced to the BoF approach through descriptive visual words counting the concurrent appearance of visual words in spatial neighborhoods [22], which could become a computationally demanding operation. Global statistics of keypoint location, orientation and detected scale delivers a global view of spatial keypoint distribution, which is obtained at moderate computational cost.

We use a measure of inhomogeneity to characterize the spatial distribution of keypoints [23]. The image is subdivided into a sequence $s^2, s = 1, 2, 4, \ldots$ of rectangular regions of equal size and the number of keypoints m_i falling into region



(a) Normal text page

(b) Scan with finger

Fig. 2. Distribution of spatial inhomogeneity of keypoint locations

$i, i = 1, \dots, s^2$ is obtained

$$
\begin{aligned}
h &= \sum_{j=1}^{\log_2 s} w^{1-j} h(2^j), \\
h(s) &= \frac{1}{2n} \sum_{i=1}^{n^2} |m_i - \frac{n}{s^2}|, \qquad (1)
\end{aligned}
$$

where $w = 4.79129$ was derived in [23]. Images having spatially uniformly distributed keypoints obtain values of $h \rightarrow 0$ and whereas for spatially concentrated keypoints we get $h \rightarrow 1$. We constructed a histogram r describing the distribution of inhomogeneity h evaluated over rectangular image subregions. Image subregions, so called tiles, were chosen to overlap by 50 percent with horizontally and vertically adjacent tiles.

The orientation information delivered for each keypoint is represented by a histogram o cointaining all orientations for all keypoints in the image. Although, each single keypoint is rotationally invariant, in order to be invariant with respect to rotation of the whole image page the keypoint orientation histogram is shifted with respect to the circular mean direction θ. The circular mean direction θ is evaluated over all keypoint orientations of all keypoints present in the page [24].

Finally, a size distribution histogram s is obtained from the distribution of SIFT keypoint size estimations. The keypoint size estimation is delivered from extremum search in position and scale in the Difference of Gaussian (DoG) scale space in SIFT. The sub-scale size estimation for each keypoint contributes to the size distribution histogram s.

TABLE I. CLASSIFICATION ACCURACY ESTIMATES BASED ON 5-FOLD CROSS-VALIDATION

	2 classes (text,finger)	3 classes (text,finger,edge)	4 classes (text,finger,edge,warping)
BoF features	99.09%	94.52%	91.15%
Global features	92.11%	86.50%	78.14%

The full feature vector g is obtained from concatenation of the individual histograms $g = (r, o, s)$. Each individual histogram r, o, s was normalized to sum to 1 and binned to b equally spaced bins.

C. Classification into document page classes

In all our experiments we used support vector machine (SVM) classifiers, namely the libSVM implementation[25], with radial basis function (RBF) kernels. The features describe in the previous section are input to the classifier. The histograms fed to the SVMs were normalized to sum to 1, but as feature vector elements are dependent on each other, no scaling of individual feature vector elements was done.

The task of hand and finger recognition was exhaustively addressed in the field of gesture recognition by different approaches, e.g. model-based approaches [10] or feature based approaches [11]. Our approach is a feature-based approach, where only hand or finger detection is aspired and two additional classes are discriminated. The additional classes are the book edge class, i.e. in the un-hinged outer side of the book is visible in the image, and image warping artifacts, e.g. those artifacts occur due to erroneous postprocessing of the scanned page, typically caused by page segmentation errors.

IV. RESULTS

For the three classes of distortions a limited number of examples were available, e.g. 300 images with visible book edge, 252 images with visible fingers or whole hands and 300 images with warping artifacts were available. A total number of 470 undistorted images were randomly sampled from scans of 59 different historical books. Visual histogram features were derived from a BoF of size $|V| = 1500$, a number which turned out to be suitable in our previous work on duplicate detection. For the global features based on location, size and orientation we used $b = 32$ which results in a 96-dimensional feature vector. Detailed classification results are based on training and test sets derived from partitioning the whole data set into 80% training and 20% test images. The parameters of the SVMs were found by exhaustively searching the parameter space. Tab. III summarizes the parameter of the cost function C and the RBF-kernel parameter γ found by exhaustive search (over a parameter grid sampled at powers of 2) and different features an number of classes.

We compared the accuracy based on 5-fold cross-validation for BoF and global approaches in Tab. I. Considering two classes the achieved accuracy is the BoF approach achieves more than 99%, even the global features based approach

TABLE II. CONFUSION MATRICES FOR VARYING NUMBER OF CLASSES BASED ON TEST SET BOF FEATURES.

Ground truth	Prediction text	Prediction finger
text	93	1
finger	3	57

(a) Two class discrimination

Ground truth	Prediction text	Prediction finger	Prediction edge
text	94	0	0
finger	4	40	6
edge	1	1	58

(b) Three class discrimination

Ground truth	Prediction text	Prediction finger	Prediction edge	Prediction warping
text	91	3	0	0
finger	1	48	1	0
edge	3	2	51	4
warping	0	2	3	55

(c) Four class discrimination

TABLE III. PARAMETERS C AND γ FOR SVM CLASSIFIERS WITH RBF KERNEL AND DIFFERENT FEATURE INPUT AND NUMBER OF CLASSES.

		2 classes (text,finger)	3 classes (text,finger,edge)	4 classes (text,finger,edge,warping)
BoF	C	128	16	32
features	γ	64	256	256
Global	C	8	8	4
features	γ	4	16	16

TABLE IV. AVERAGE PER IMAGE RUNTIME MEASUREMENT RESULTS FOR INDIVIDUAL COMPUTATION STEPS ON A XEON 3.6 GHZ COMPUTER.

Operation	Computation time in sec.
SIFT extraction	4.02
BoF learning	1.34
SVM training BoF	0.67
SVM classification BoF	0.48
Global feature extraction	0.21
SVM training global features	0.05
SVM classification global features	0.07

obtains more than 90% accuracy and might be useful in cases were computation is limited. In detail, based on the test set four false decisions were made in the BoF approach, see also Tab. II (a). Fig. 3(a) shows the text image falsely identified as an image containing an operator finger. Due to the structure made by the graphical element and the stamp, which reminds of hand structures, this image is supposed to be wrongly classified. One of the images showing a finger but falsely classified as text is shown in Fig. 3(b), where the finger covers only a small and saturated portion of the image.

For more than two classes we concentrate on the accuracy of the BoF based classification which slightly decreases, while the global features based approach becomes impractical, see Tab. II. Fig. 4(a) shows an image showing a saturated book edge which resulted in misclassification as text. Fig. 4(a) shows an image with mixed content which was categorized as a warping artifact in the ground truth, but the classifier decides to label it as an edge distortion.

The computational effort is summarized in Tab. IV, where average numbers for per image computation for the two-class task, i.e. discrimination between undistorted images and images distorted by a finger/hand, were considered. Each image was of the size of 1 million pixels on the average. Some images of the collection used color while others were greyscale images. To unify processing and due to limited relevancy of color in the data all image processing was performed on greyscale images.The computer used for the experiments was a Xeon 3.6 GHz machine using a MATLAB/C implementation of the algorithms. The most cost intense part was the SIFT feature extraction, which is mainly due to the content of the images, i.e. a large number of SIFT features are found in text images. The more discriminative BoF visual histogram features require approximately one order of magnitude more computation when compared to global features. Nevertheless, the BoF learning, when compared to SIFT features extraction, takes less computation time per image.

V. CONCLUSION

We have presented an approach for page content classification suitable for automatic quality assurance and assessment in document scanning workflows. The page distortion classes are learned from training data by a SVM and are not modeled explicitly. Although being inherently "location-blind" features based on visual words taken from a BoF of SIFT descriptors performed sufficiently good in order to reduce human interaction in quality control. Global features derived from keypoint location, size and orientation statistics are suited for simple or limited discrimination tasks, e.g. distinction of pages without and with an operator's finger or hand visible. Although being computationally much more efficient, the accuracy of those global features drops when attempting to discriminate more than two classes.

ACKNOWLEDGMENT

The authors would like to thank Sven Schlarb from the Austrian National Library (ONB) for providing data and expertise on library workflows.

This work is part of the SCAlable Preservation Environments (SCAPE) project which aims at developing scalable services for planning and execution of institutional digital preservation strategies. The SCAPE project is co-funded by the European Union under FP7 ICT-2009.4.1 (Grant Agreement number 270137).

REFERENCES

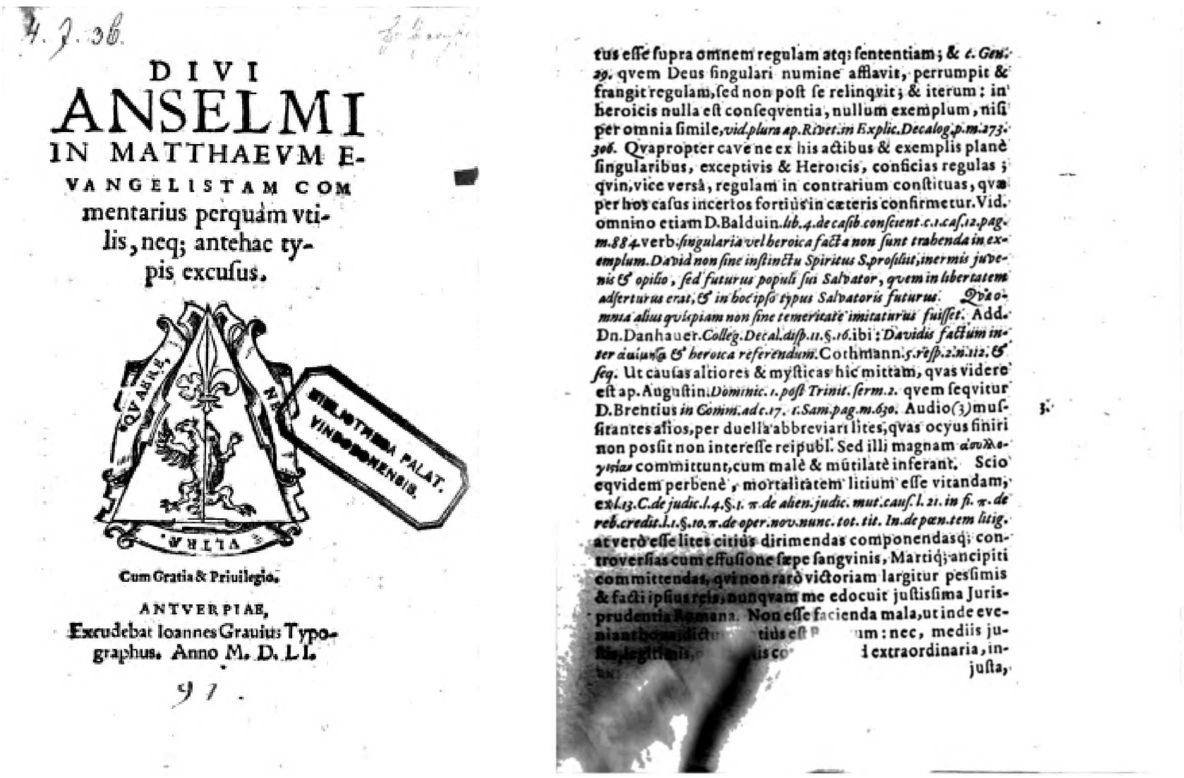
[1] A. Langley and D. S. Bloomberg, "Google books: making the public domain universally accessible," in *Proc. of SPIE, Document Recognition and Retrieval XIV*, vol. 6500, San Jose, CA, Jan 2007, pp. 65 000H–65 000H–10.

[2] K. Chaudhury, A. Jain, S. Thirthala, V. Sahasranaman, S. Saxena, and S. Mahalingam, "Google newspaper search - image processing and analysis pipeline," in *Proc. of Intl. Conf. on Document Analysis and Recognition*, Jul 2009, pp. 621 –625.

[3] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. of Comput. Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[4] Y. Jing and S. Baluja, "Visualrank: Applying pagerank to large-scale image search," *IEEE Trans. Pat. Anal. Mach. Intel.*, vol. 30, no. 11, pp. 1877–1890, Nov.

[5] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in *Workshop on Statistical Learning in Computer Vision, ECCV 2004*, 2004, pp. 1–22.



4. J. 36.

DIVI
ANSELMI
IN MATTHAEVM EVANGELISTAM COMmentarius perquàm vtilis, neq; antehac typis excuſus.

Cum Gratia & Priuilegio.

ANTVERPIAE,
Excudebat Ioannes Grauius Typographus. Anno M. D. LI.

97 .

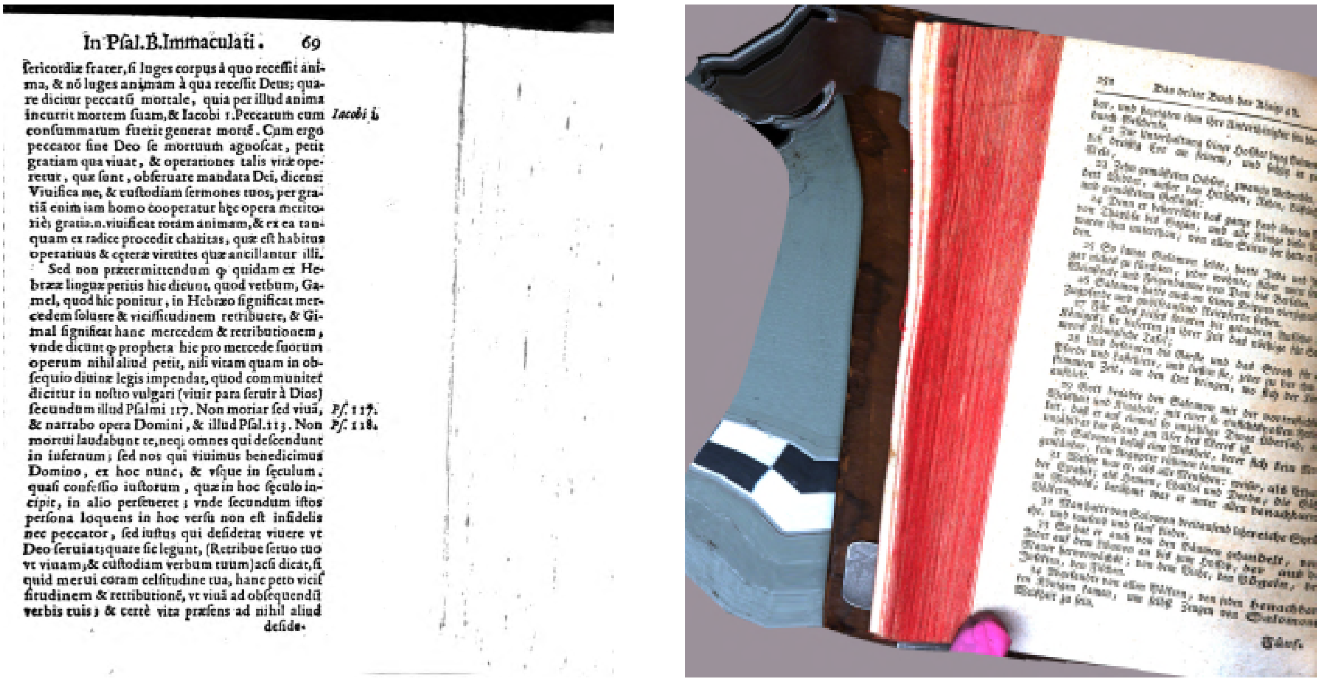
tus eſſe ſupra omnem regulam atq; ſententiam; & *c. Gen. 29.* qvem Deus ſingulari numine afflavit, perrumpit & frangit regulam, ſed non poſt ſe relinqvit; & iterum: in heroicis nulla eſt conſeqventia, nullum exemplum, niſi per omnia ſimile, *vid. plura ap. Rivet. in Explic. Decalog. p.m. 273. 306.* Qvapropter cave ne ex his actibus & exemplis planè ſingularibus, exceptivis & Heroicis, conficias regulas; qvin, vice verſà, regulam in contrarium conſtituas, qvæ per hos caſus incertos fortiùs in cæteris confirmetur. Vid. omnino etiam D. Balduin. *lib. 4. de caſib. conſcient. c. 1. caſ. 12. pag. m. 884.* verb. *ſingularia vel heroica facta non ſunt trahenda in exemplum. David non ſine inſtinctu Spiritus S. proſiliit, inermis juvenis & opilio, ſed futurus populi ſui Salvator, qvem in libertatem adſerturus erat, & in hoc ipſo typus Salvatoris futurus. Qvæ omnia alius qvispiam non ſine temeritate imitaturus fuiſſet.* Add. Dn. Danhauer. *Colleg. Decal. diſp. 11. §. 16.* ibi: *Davidis factum inter* [illegible] *& heroica referendum.* Cothmann. *5. reſp. 2. n. 112. & ſeq.* Ut cauſas altiores & myſticas hic mittam, qvas videre eſt ap. Auguſtin. *Dominic. 1. poſt Trinit. ſerm. 2.* qvem ſeqvitur D. Brentius *in Comm. ad c. 17. 1. Sam. pag. m. 630.* Audio (3) muſſitantes alios, per duella abbreviari lites, qvas ocyus finiri non poſſit non intereſſe reipubl. Sed illi magnam *ἀσυλλογισίας* committunt, cum malè & mutilatè inferant. Scio eqvidem perbenè, mortalitatem litium eſſe vitandam, ex *l. 13. C. de judic. l. 4. §. 1. π. de alien. judic. mut. cauſ. l. 21. in fi. π. de reb. credit. l. 1. §. 10. π. de oper. nov. nunc. tot. tit. In. de pœn. tem litig.* at verò eſſe lites citiùs dirimendas componendasq; controverſias cum effuſione ſæpe ſangvinis, Martiq; ancipiti committendas, qvi non rarò victoriam largitur peſſimis & facti ipſius reis, nunqvam me edocuit juſtiſſima Jurisprudentia Romana. Non eſſe facienda mala, ut inde eveniant bona [illegible] tius eſt [illegible] um: nec, mediis juſtis, legitimis, [illegible] extraordinaria, injuſta,

3.

(a) False classification of finger

(b) False classification of text

Fig. 3. Misclassifications in discrimination between undistorted text and text distorted by a finger.



In Pſal. B. Immaculati. 69

ſericordiæ frater, ſi luges corpus à quo receſſit anima, & nõ luges animam à qua receſſit Deus; quare dicitur peccatũ mortale, quia per illud anima incurrit mortem ſuam, & Iacobi 1. Peccatum cum conſummatum fuerit generat mortẽ. Cum ergo peccator ſine Deo ſe mortuum agnoſcat, petit gratiam qua viuat, & operationes talis vitæ operetur, quæ ſunt, obſeruare mandata Dei, dicens: Viuifica me, & cuſtodiam ſermones tuos, per gratiã enim iam homo cooperatur hęc opera meritoriè; gratia.n. viuificat totam animam, & ex ea tanquam ex radice procedit charitas, quæ eſt habitus operatiuus & cęteræ virtutes quæ ancillantur illi.

Iacobi 1.

Sed non prætermittendum ꝙ quidam ex Hebrææ linguæ peritis hic dicunt, quod verbum, Gamel, quod hic ponitur, in Hebræo ſignificat mercedem ſoluere & viciſſitudinem retribuere, & Gimal ſignificat hanc mercedem & retributionem, vnde dicunt ꝙ propheta hic pro mercede ſuorum operum nihil aliud petit, niſi vitam quam in obſequio diuinæ legis impendat, quod communiter dicitur in noſtro vulgari (viuir para ſeruir à Dios) ſecundum illud Pſalmi 117. Non moriar ſed viuã, & narrabo opera Domini, & illud Pſal. 113. Non mortui laudabunt te, neq; omnes qui deſcendunt in infernum; ſed nos qui viuimus benedicimus Domino, ex hoc nunc, & vſque in ſęculum. quaſi confeſſio iuſtorum, quæ in hoc ſęculo incipit, in alio perſeueret; vnde ſecundum iſtos perſona loquens in hoc verſu non eſt infidelis nec peccator, ſed iuſtus qui deſiderat viuere vt Deo ſeruiat; quare ſic legunt, (Retribue ſeruo tuo vt viuam, & cuſtodiam verbum tuum) acſi dicat, ſi quid merui coram celſitudine tua, hanc peto viciſſitudinem & retributionẽ, vt viuã ad obſequendũ verbis tuis; & certè vita præſens ad nihil aliud deſide-

Pſ. 117.
Pſ. 118.

(a) False classification of edge as text

(b) False classification of warping as edge

Fig. 4. Misclassifications in discrimination between undistorted text and text distorted by a finger, book edge or warping artifacts.

[6] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*, 7th ed. Cambridge University Press, 2008.

[7] P. Weinzaepfel, H. Jégou, and P. Pérez, "Reconstructing an image from its local descriptors," in *Computer Vision and Pattern Recognition*, Colorado Springs, USA, Jun. 2011.

[8] R. Huber-Mörk and A. Schindler, "Quality assurance for document image collections in digital preservation," in *Proc. of Advanced Concepts for Intelligent Vision Systems ACIVS 2012*, ser. Springer LNCS, vol.

7517, Brno, CZ, Sep 2012, pp. 108–119.

[9] R. Huber-Mörk, A. Schindler, and S. Schlarb, “Duplicate detection for quality assurance of document image collections,” in *Proc. of Conf. on Digital Preservation iPres 2012*, Toronto, CA, Oct 2012, pp. 136–143.

[10] J. Rehg and T. Kanade, “Digiteyes: vision-based hand tracking for human-computer interaction,” in *Proc.Workshop on Motion of Non-Rigid and Articulated Objects*, Nov, pp. 16–22.

[11] W. T. Freeman and M. Roth, “Orientation histograms for hand gesture recognition,” in *Proc. Workshop on Automatic Face- and Gesture-Recognition*, Jun. 1995, pp. 296–301.

[12] D. Doermann, H. Li, and O. Kia, “The detection of duplicates in document image databases,” *Image and Vision Computing*, vol. 16, no. 12-13, pp. 907 – 920, 1998.

[13] J. van Beusekom, F. Shafait, and T. Breuel, “Image-matching for revision detection in printed historical documents,” in *Proc. Symposium of the German Association for Pattern Recognition*, ser. LNCS, vol. 4713, Sep 2007, pp. 507–516.

[14] J. van Beusekom, D. Keysers, F. Shafait, and T. Breuel, “Distance measures for layout-based document image retrieval,” in *Second International Conference on Document Image Analysis for Libraries, 2006. DIAL ’06*, April 2006, pp. 231–242.

[15] S. Ramachandrula, G. Joshi, S. Noushath, P. Parikh, and V. Gupta, “Paperdiff: A script independent automatic method for finding the text differences between two document images,” in *The Eighth IAPR International Workshop on Document Analysis Systems, 2008. DAS ’08*, Sep 2008, pp. 585 –590.

[16] S. Baluja and M. Covell, “Finding images and line drawings in document-scanning systems,” in *Proc. Intl. Conf. on Document Analysis and Retrieval ICDAR 2009*, 2009.

[17] A. Garz, R. Sablatnig, and M. Diem, “Layout analysis for historic manuscripts using SIFT features,” in *Proc. of Intl. Conf. on Document Analysis and Recognition ICDAR 2011*, 2011, pp. 508–512.

[18] G. Zhu and D. Doermann, “Automatic document logo detection,” in *Proceedings of the International Conference on Document Analysis and Recognition (ICDAR)*, vol. 2, Washington, DC, USA, 2007, pp. 864–868.

[19] Z. Li, M. Schulte-Austum, and M. Neschen, “Fast logo detection and recognition in document images,” in *Proc. of International Conference on Pattern Recognition (ICPR)*, Aug., pp. 2716–2719.

[20] L. Hazelhoff, I. Creusen, D. van de Wouw, and P. H. N. de With, “Large-scale classification of traffic signs under real-world conditions,” in *Proc. SPIE Electronic Imaging: Algorithms and Systems VI*, 2012.

[21] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, “Object retrieval with large vocabularies and fast spatial matching,” in *Proc. of Computer Vision and Pattern Recognition*, 2007.

[22] S. Zhang, Q. Tian, G. Hua, Q. Huang, and S. Li, “Descriptive visual words and visual phrases for image applications,” in *Proc. Intl. Conf. on Multimedia MM’09*, 2009, pp. 75–84.

[23] U. Schilcher, M. Gyarmati, C. Bettstetter, Y. W. Chung, and Y. H. Kim, “Measuring inhomogeneity in spatial distributions,” in *Proc. Vehicular Technology Conference, VTC 2008*, May 2008, pp. 2690 –2694.

[24] N. I. Fisher, *Statistical Analysis of Circular Data*. Cambridge University Press, 1995.

[25] C.-C. Chang and C.-J. Lin, “Libsvm: A library for support vector machines,” *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, May 2011.